Psychological Review

VOLUME 84 NUMBER 3 MAY 1977

Telling More Than We Can Know: Verbal Reports on Mental Processes

Richard E. Nisbett and Timothy DeCamp Wilson
University of Michigan

Evidence is reviewed which suggests that there may be little or no direct introspective access to higher order cognitive processes. Subjects are sometimes (a) unaware of the existence of a stimulus that importantly influenced a response, (b) unaware of the existence of the response, and (c) unaware that the stimulus has affected the response. It is proposed that when people attempt to report on their cognitive processes, that is, on the processes mediating the effects of a stimulus on a response, they do not do so on the basis of any true introspection. Instead, their reports are based on a priori, implicit causal theories, or judgments about the extent to which a particular stimulus is a plausible cause of a given response. This suggests that though people may not be able to observe directly their cognitive processes, they will sometimes be able to report accurately about them. Accurate reports will occur when influential stimuli are salient and are plausible causes of the responses they produce, and will not occur when stimuli are not salient or are not plausible causes.

"Why do you like him?" "How did you solve this problem?" "Why did you take that job?"

In our daily lives we answer many such questions about the cognitive processes underlying our choices, evaluations, judgments, and behavior. Sometimes such questions are asked by social scientists. For example, investigators have asked people why they like particular political candidates (Gaudet, 1955) or detergents (Kornhauser & Lazarsfeld, 1935), why they chose a particular occupation (Lazarsfeld, 1931), to go to graduate school (Davis, 1964) or to become a juvenile delinquent (Burt, 1925), why they got married or divorced (Goode, 1956) or joined a voluntary organization (Sills, 1957) or moved to a new home (Rossi, 1955) or sought out a psychoanalyst (Kadushin, 1958), or failed to use a contraceptive technique (Sills, 1961). Social psychologists routinely ask the subjects in their experiments why they behaved, chose, or evaluated as they did. Indeed, some social psychologists have advocated the abandonment of the social psychology experiment and its deceptive practices and have urged that subjects simply be asked how their cognitive processes *would* work if they were to be confronted with particular stimulus situations (Brown, 1962; Kelman, 1966).

Recently, however, several cognitive psychologists (Mandler, 1975a, 1975b; Miller,

The writing of this paper, and some of the research described, was supported by grants GS-40085 and BNS75-23191 from the National Science Foundation. The authors are greatly indebted to Eugene Borgida, Michael Kruger, Lee Ross, Lydia Temoshok, and Amos Tversky for innumerable ideas and generous and constructive criticism. John W. Atkinson, Nancy Bellows, Dorwin Cartwright, Alvin Goldman, Sharon Gurwitz, Ronald Lemley, Harvey London, Hazel Markus, William R. Wilson, and Robert Zajonc provided valuable critiques of earlier drafts of the paper.

Requests for reprints should be sent to Richard E. Nisbett, Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48109.

1962; Neisser, 1967) have proposed that we may have no direct access to higher order mental processes such as those involved in evaluation, judgment, problem solving, and the initiation of behavior. The following quotations will serve to indicate the extent to which these investigators doubt people's ability to observe directly the workings of their own minds. "It is the *result* of thinking, not the process of thinking, that appears spontaneously in consciousness" (Miller, 1962, p. 56). "The constructive processes [of encoding perceptual sensations] themselves never appear in consciousness, their products do" (Neisser, 1967, p. 301). And in Neisser's next paragraph: "This general description of the fate of sensory information seems to fit the higher mental processes as well" (p. 301). Mandler's (1975a) suggestions are still more sweeping: "The analysis of situations and appraisal of the environment . . . goes on mainly at the nonconscious level" (p. 241). "There are many systems that cannot be brought into consciousness, and probably most systems that analyze the environment in the first place have that characteristic. In most of these cases, only the products of cognitive and mental activities are available to consciousness" (p. 245). And finally: "unconscious processes . . . include those that are not available to conscious experience, be they feature analyzers, deep syntactic structures, affective appraisals, computational processes, language production systems, action systems of many kinds" (p. 230).

It is important to note that none of these writers cites data in support of the view that people have no direct access to higher order mental processes. In fact, when the above quotations are read in context, it is clear that the source of the speculations is not research on higher order processes such as "thinking," "affective appraisal," and "action systems," but rather research on more basic processes of perception and memory. Recent research has made it increasingly clear that there is almost no conscious awareness of perceptual and memorial processes. It would be absurd, for example, to ask a subject about the extent to which he relied on parallel line convergence when making a judgment of depth or whether he stored the meanings of animal names in a hierarchical tree fashion or in some other manner. Miller (1962) has provided an excellent example of our lack of awareness of the operation of memorial processes. If a person is asked, "What is your mother's maiden name?", the answer appears swiftly in consciousness. Then if the person is asked "How did you come up with that?", he is usually reduced to the inarticulate answer, "I don't know, it just came to me."

It is a substantial leap, however, from research and anecdotal examples concerning perception and memory to blanket assertions about higher order processes. In the absence of evidence indicating that people cannot correctly report on the cognitive processes underlying complex behaviors such as judgment, choice, inference, and problem solving, social scientists are not likely to abandon their practice of quizzing their subjects about such processes. The layman is even less likely to abandon his habit of asking and answering such questions.

A second problem with the new anti-introspectivist view is that it fails to account for the fact, obvious to anyone who has ever questioned a subject about the reasons for his behavior or evaluations, that people readily answer such questions. Thus while people usually appear stumped when asked about perceptual or memorial processes, they are quite fluent when asked why they behaved as they did in some social situation or why they like or dislike an object or another person. It would seem to be incumbent on one who takes a position that denies the possibility of introspective access to higher order processes to account for these reports by specifying their source. If it is not direct introspective access to a memory of the processes involved, what is the source of such verbal reports?

Finally, a third problem with the anti-introspectivist view is that it does not allow for the possibility that people are ever correct in their reports about their higher order mental processes. It seems intuitively unlikely that such reports are always inaccurate. But if people are sometimes accurate, several questions arise. (a) What is the basis of these accurate reports? (b) Are accurate reports fundamentally different in kind from inaccurate ones? (c) Is it possible to specify what sorts of reports will be accurate and what sorts will be inaccurate?

The first part of this article is concerned with

a review of the evidence bearing on the accuracy of subjective reports about higher mental processes. The second part of the paper presents an account of the basis of such reports. We shall argue for three major conclusions.

1. 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. The accuracy of subjective reports is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports.

2. 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. If the stimulus psychologically implies the response in some way (Abelson, 1968) or seems "representative" of the sorts of stimuli that influence the response in question (Tversky & Kahneman, 1974), the stimulus is reported to have influenced the response. If the stimulus does not seem to be a plausible cause of the response, it is reported to be noninfluential.

3. Subjective reports about higher mental processes are sometimes correct, but even the instances of correct report are not due to direct introspective awareness. Instead, they are due to the incidentally correct employment of a priori causal theories.

Verbal Reports on Cognitive Processes in Dissonance and Attribution Studies

Much of the evidence that casts doubt on the ability of people to report on their cognitive processes comes from a study of the literature that deals with cognitive dissonance and self-perception attribution processes. Or rather, the evidence comes from a consideration of what was *not* published in that literature. A review of the nonpublic, sub rosa aspects of these investigations leads to three conclusions: (a) Subjects frequently cannot report on the existence of the chief response that was produced by the manipulations; (b) even when they are able to report the existence of the responses, subjects do not report that a *change process* occurred, that is, that an evaluational or attitudinal response underwent any alterations; and (c) subjects cannot correctly identify the stimuli that produced the response.

Awareness of the Existence of the Response

The central idea of insufficient justification or dissonance research is that behavior that is intrinsically undesirable will, when performed for inadequate extrinsic reasons, be seen as more attractive than when performed for adequate extrinsic reasons. In the view of Festinger (1957) and other dissonance theorists, attitude change occurs because the cognition "I have done something unpleasant without adequate justification" is dissonant and therefore painful; and the person revises his opinion about the behavior in order to avoid the psychic discomfort.

The central idea of attribution theory is that people strive to discover the causes of attitudinal, emotional, and behavioral responses (their own and others), and that the resulting causal attributions are a chief determinant of a host of additional attitudinal and behavioral effects. Thus, for example, if someone tells us that a particular Western movie is a fine film, our acceptance of that opinion, and possibly our subsequent behavior, will be determined by our causal analysis of the person's reasons for the evaluation: Does he like all movies? All Westerns? All John Wayne movies? Do other people like the movie? Does this person tend to like movies that other people do not like?

Many insufficient-justification studies and many attribution studies where the subject makes inferences about himself have employed behavioral dependent variables. Substantial effects have been shown on behavior of inherent interest and with significant social implications, including pain, hunger and thirst tolerance, psychopathology, task perseverance, and aggressive behavior. Two examples will serve to illustrate research with behavioral consequences.

Zimbardo, Cohen, Weisenberg, Dworkin, and Firestone (1969) asked subjects to accept

a series of painful electric shocks while performing a learning task. When the task was completed, subjects were asked to repeat it. Some subjects were given adequate justification for performing the task a second time and accepting another series of shocks (the research was very important, nothing could be learned unless the shocks were given again), while other subjects were given insufficient justification (the experimenter wanted to satisfy his more or less idle curiosity about what would happen). Subjects with insufficient justification for accepting the shocks showed lower GSR responses and better learning performance on the second task than subjects with sufficient justification. The explanation offered for this finding is that insufficient-justification subjects sought to justify taking the shocks, which they did by deciding that the shocks were not all that painful. Thus the evaluation of the painfulness of the shocks was lowered, and physiological and behavioral indicators reflected this evaluation.

A study by Valins and Ray (1967) will illustrate the attribution paradigm. These investigators asked snake-phobic subjects to watch slides while receiving occasional electric shocks. Subjects were wired for what they believed were recordings of heart rate. They were allowed to hear a rhythmic pattern of sounds which, they were told, was the amplified sound of their own heart beats. Subjects were shown a series of slides of snakes interspersed with slides of the word "SHOCK." Following each presentation of the shock slide, subjects were given an electric shock. After a few such pairings, the appearance of the shock slide was accompanied by an increased rate of "heartbeats." Snake slides were never accompanied by any change in apparent heart rate. Following this procedure, subjects were requested to approach, and if possible, to touch, a 30-inch (76.2-cm) boa constrictor. Such subjects approached the snake more closely than subjects who had gone through the identical procedure but who believed that the "heartbeats" were simply "extraneous sounds" (which, of course, they actually were). The finding is explained as follows. Subjects in the heart rate condition learned that their "heart rate" indicated they were appropriately frightened when they saw the shock slide, because of the electric shock it portended, but that they were not frightened by the snake slides. If they were not frightened by the snake slides, perhaps they were not as afraid of live snakes as they had thought. Armed with this new self-attribution of snake fearlessness, they were more willing to approach the boa.

The two experiments just described share a common formal model. Verbal stimuli in the form of instructions from the experimenter, together with the subject's appraisal of the stimulus situation, are the inputs into a fairly complicated cognitive process which results in a changed evaluation of the relevant stimuli and an altered motivational state. These motivational changes are reflected in subsequent physiological and behavioral events. Thus: stimuli → cognitive process → evaluative and motivational state change → behavior change. Following traditional assumptions about higher mental processes, it has been tacitly assumed by investigators that the cognitive processes in question are for the most part verbal, conscious ones. Thus the subject consciously decides how he feels about an object, and this evaluation determines his behavior toward it. As several writers (Bem, 1972; Nisbett & Valins, 1972; Storms & Nisbett, 1970; Weick, 1966) have pointed out, there is a serious problem with this implicit assumption: Typically, behavioral and physiological differences are obtained in the absence of *verbally reported* differences in evaluations or motive states. For example, in the study by Zimbardo, Cohen, Weisenberg, Dworkin, and Firestone (1969), experimental subjects given inadequate justification for taking shock learned much more quickly and showed much less GSR reactivity to the shock than did control, adequate-justification subjects, but the former did not report the shock to be significantly less painful than did the latter. And subjects in the Valins and Ray (1967) experiment who had "inferred" that they were not very frightened of snakes, as indicated by their willingness to approach the boa constrictor, showed no evidence of any such inference when asked a direct question about how frightened they were of snakes.

We have reviewed all the insufficient-justification and attribution studies we have been able to find that meet the following criteria: (a) behavioral or physiological effects were ex-

amined, and (b) at approximately the same time, verbal reports of evaluations and motivational states were obtained. Studies that did not permit a clear, uncontroversial comparison of the strength of behavioral and self-report indicators were not included (e.g., Brehm, Back, & Bogdonoff, 1969; Schachter & Singer, 1962), nor were studies that employed controversial, poorly understood techniques such as hypnosis (e.g., Brock & Grant, 1969).

Three striking generalizations can be made about these studies:

1. In the majority of studies, no significant verbal report differences were found at all. This applies to studies by Cohen and Zimbardo (1969), Cottrell and Wack (1967), Davison and Valins (1969), Ferdinand (1964), Freedman (1965), Grinker (1969), Pallak (1970), five experiments by Pallak, Brock, and Kiesler (1967), Experiment 1 in Pallak and Pittman (1972), Schachter and Wheeler (1962), Snyder, Schultz, and Jones (1974), Storms and Nisbett (1970), Valins and Ray (1967), Waterman (1969), Weick and Penner (1969), Weick and Prestholdt (1968), and Zimbardo, Cohen, Weisenberg, Dworkin, and Firestone (1969).

2. In the remainder of studies, the behavioral effects were in most cases stronger (i.e., more statistically reliable) than the verbal report effects (Berkowitz & Turner, 1974; Kruglanski, Friedman, & Zeevi, 1971; Schlachet, 1969; Nisbett & Schachter, 1966; Experiment 2 of Pallak and Pittman, 1972; and Weick, 1964). Exceptions to this are reports by Brehm (1969), Freedman (1963), Mansson (1969), and Zimbardo, Weisenberg, Firestone, and Levy (1969).

3. In two studies where it was reported, the correlation between verbal report about motive state and behavioral measures of motive state was found to be nil (Storms & Nisbett, 1970; Zimbardo, Cohen, Weisenberg, Dworkin, & Firestone, 1969). The rest of the literature in this area is strangely silent concerning the correlations between verbal report and behavior. Since positive correlations would have constituted support for investigators' hypotheses, while zero or negative correlations would have been difficult to understand or interpret in terms of prevailing assumptions about the nature of the cognitive processes involved, the failure to report the correlations constitutes presumptive evidence that they were not positive. In order to check on this possibility, we wrote to the principal investigators of the studies described above, asking for the correlations between verbal report and behavior. Only three investigators replied by saying that they still had the data and could provide the correlations. In all three instances, the correlations were in fact nonsignificant and close to zero (Davison & Valins, 1969; Freedman, 1965; Snyder, Schultz, & Jones, 1974).

The overall results thus confound any assumption that conscious, verbal cognitive processes result in conscious, verbalizable changes in evaluations or motive states which then mediate changed behavior. In studies where the data are available, no association is found between degree of verbal report change and degree of behavior change in experimental groups. And in most studies no evidence is found that experimental subjects differ from control subjects in their verbal reports on evaluations and motivational states.

What of the studies that do find differences in the verbal reports of experimental and control subjects? (It should be noted that this includes many studies not reviewed here where the *only* dependent measure was a verbal one and where differences between experimental and control groups were obtained.) Should these studies be taken as evidence that the traditional model sometimes works, that subjects are sometimes aware of the cognitive processes that occur in these experiments? Evidence to be discussed below casts doubt on such a conclusion.

Awareness of the Existence of a Change Process

There is an important difference between awareness of the *existence* of an evaluation or motive state and awareness of a *changed* evaluation or motive state. The former sort of awareness does not imply true recognition of the process induced by insufficient justification and attribution manipulations—which in fact always involves a change in evaluations. Thus if it could be shown that subjects cannot report on the fact that a change has taken place as a

consequence of such manipulations, this would suggest that they are not aware of the occurrence of a process.

Bem and McConnell (1970) contrived an experiment to demonstrate that in fact, subjects do not experience a subjective change in their evaluations in response to insufficient-justification manipulations. A stock-in-trade of the dissonance tradition is the counterattitudinal advocacy experiment. In this type of experiment, subjects are asked to write an essay opposing their own views on some topic and are then asked what their attitudes are toward the topic. Subjects who are coerced (or heavily bribed) into writing the essays show no change in evaluation of the topic. Subjects who are given insufficient justification for writing the essay, or who are manipulated into believing that they had free choice in the matter, typically shift their evaluations in the direction of the position advocated in the essay. On the face of it, this would seem to indicate that subjects are aware of the existence of a change process since the means employed for assessing the response is a verbal report, and this report changes from premanipulation measures to postmanipulation measures.

Bem and McConnell contested this assumption by the simple expedient of asking the subjects, at the time of the postmanipulation measure, what their attitude *had been* 1 week earlier, at the time of the premanipulation measure. Control subjects had no difficulty reporting accurately on their previous opinions. In contrast, though the postmanipulation attitudes of experimental subjects were substantially different from their premanipulation attitudes, they reported that their current attitudes were the same as their premanipulation attitudes. Thus subjects apparently changed their attitudes in the absence of any subjective experience of change. This suggests that though subjects can sometimes report on the existence of the new evaluation, they may still be unaware of the fact that the evaluation has changed. If so, then they cannot be aware of the nature of the cognitive process that has occurred, because they are not even aware of the fact that a process has occurred at all.

Such a conclusion gains credence in view of a truly stunning demonstration of the same phenomenon by Goethals and Reckman (1973). These investigators assessed the opinions of high school students on 30 social issues, including attitudes toward busing of school children to achieve racial integration. One to two weeks later, students were called and asked to participate in a group discussion of the busing issue. Each group was composed of three subjects whose pretest opinions indicated that they were all pro-busing or all anti-busing, plus one high school student confederate who was armed with a number of persuasive opinions and whose job it was to argue persistently against the opinion held by all other group members. He was highly successful in this task. Following the discussion, subjects indicated their opinions on the busing issue—on a scale different in form from the original measure. The original anti-busing subjects had their opinions sharply moderated in a pro-direction. Most of the pro-busing subjects were actually converted to an anti-busing position. Then Goethals and Reckman asked their subjects to recall, as best they could, what their original opinions on the busing question had been. Subjects were reminded that the experimenters were in possession of the original opinion scale and would check the accuracy of the subjects' recall. Control subjects were able to recall their original opinions with high accuracy. In contrast, among experimental subjects, the original anti-busing subjects "recalled" their opinions as having been much more pro-busing than they actually were, while the original pro-busing subjects actually recalled their original opinions as having been, on the average, anti-busing! In fact, the original pro-busing subjects recalled that they had been more anti-busing than the original anti-busing subjects recalled that they had been.

It would appear that subjects in the Goethals and Reckman (1973) study did not actually experience these enormous shifts as opinion change:

> Some subjects listened carefully to the course of the discussion and began to nod their heads in agreement with the confederate's arguments. They seemed to come to agree with him without any awareness of their earlier attitude. In the debriefing they gave every indication that the position they adopted after the discussion was the position they had basically always held. . . .Most commented that the discussion had served to broaden their awareness of the issues involved or had provided

support for their original position. No subject reported that the discussion had had any effect in changing or modifying his position. (p. 499)

Thus research in the insufficient-justification and attribution traditions seems to indicate that (a) subjects sometimes do not report the evaluational and motivational states produced in these experiments; and (b) even when they can report on such states, they may not report that a change has taken place in these states.

It may have occurred to the reader that the most direct approach to the question of accuracy of subjects' reports in these experiments would be simply to ask subjects why they behaved as they did and listen to what they have to say about their own cognitive processes. This would indeed be a fruitful approach, and it is discussed below.

Reports About Cognitive Processes

A literal reading of the literature would give the impression that researchers working in the areas of insufficient-justification and attribution have not bothered to ask their subjects about their thought processes. We have been able to find only a single report of the results of such questioning. This is the terse and intriguing report by Ross, Rodin, and Zimbardo (1969) in their experiment on reattribution of arousal symptoms that the subjects "never explicitly mentioned any conflict about, or searching for, the 'explanation' for their arousal. This suggests that attribution may never have been consciously debated by these subjects" (p. 287). Fortunately, additional unpublished data, collected from subjects following their participation in attribution experiments by Nisbett and Schachter (1966) and Storms and Nisbett (1970), are available. These data are consistent with the description supplied by Ross et al.

In the experiment by Nisbett and Schachter (1966), subjects were requested to take a series of electric shocks of steadily increasing intensity. Prior to exposure to the shock, some of the subjects were given a placebo pill which, they were told, would produce heart palpitations, breathing irregularities, hand tremor, and butterflies in the stomach. These are the physical symptoms most often reported by subjects as accompanying the experience of electric shock. It was anticipated that when subjects with these instructions were exposed to the shock, they would attribute their arousal symptoms to the pill, and would therefore be willing to tolerate more shock than subjects who could only attribute these aversive symptoms to the shock. And, in fact, the pill attribution subjects took four times as much amperage as shock attribution subjects.

Following his participation in the experiment, each subject in the pill attribution group was interviewed following a Spielberger-type (1962) graded debriefing procedure. (a) Question: "I notice that you took more shock than average. Why do you suppose you did?" Typical answer: "Gee, I don't really know. . . . Well, I used to build radios and stuff when I was 13 or 14, and maybe I got used to electric shock." (b) Question: "While you were taking the shock, did you think about the pill at all?" Typical answer: "No, I was too worried about the shock." (c) Question: "Did it occur to you at all that the pill was causing some physical effects?" Typical answer: "No, like I said, I was too busy worrying about the shock." In all, only 3 of 12 subjects reported having made the postulated attribution of arousal to the pill. (d) Finally, the experimenter described the hypothesis of the study in detail, including the postulated process of attribution of symptoms to the pill. He concluded by asking the subject if he might have had any thoughts like those described. Subjects typically said that the hypothesis was very interesting and that many people probably would go through the process that the experimenter described, but so far as they could tell, they themselves had not.

A similar blank wall was discovered by Storms and Nisbett (1970) in their experiment on the reattribution of insomnia symptoms. In that experiment, insomniac subjects were asked to report, for 2 consecutive nights, on the time they had gone to bed and the time they had finally gotten to sleep. Arousal condition subjects were then given a placebo pill to take 15 minutes before going to bed for the next 2 nights. These subjects were told that the pill would produce rapid heart rate, breathing irregularities, bodily warmth, and alertness—the physical and emotional symptoms, in other words, of insomnia. Relaxation subjects were

told that their pills would produce the opposite symptoms—lowered heart rate, breathing rate, body temperature, and a reduction in alertness. It was anticipated that subjects in the arousal condition would get to sleep more quickly on the nights they took the pills because they would attribute their arousal symptoms to the pills rather than to emotionally laden cognitions concerning work or social life. Relaxation subjects were expected to take longer to get to sleep since they would infer that their emotional cognitions must be particularly intense because they were as fully aroused as usual even though they had taken a pill intended to lower arousal. These were in fact the results. Arousal subjects reported getting to sleep 28% quicker on the nights with the pills, and relaxation subjects reported taking 42% longer to get to sleep. Sleep onset was unaffected for control subjects.

In the interview following completion of the experiment, it was pointed out to subjects in experimental conditions that they had reported getting to sleep more quickly (or more slowly) on experimental nights than on the previous nights, and they were asked why. Arousal subjects typically replied that they usually found it easier to get to sleep later in the week, or that they had taken an exam that had worried them but had done well on it and could now relax, or that problems with a roommate or girlfriend seemed on their way to a resolution. Relaxation subjects were able to find similar sorts of reasons to explain their increased sleeplessness. When subjects were asked if they had thought about the pills at all before getting to sleep, they almost uniformly insisted that after taking the pills they had completely forgotten about them. When asked if it had occurred to them that the pill might be producing (or counteracting) their arousal symptoms, they reiterated their insistence that they had not thought about the pills at all after taking them. Finally, the experimental hypothesis and the postulated attribution processes were described in detail. Subjects showed no recognition of the hypothesized processes and (unlike subjects in the Nisbett and Schachter study) made little pretense of believing that *any* subjects could have gone through such processes.

Since many skilled and thorough investigators have worked in the dissonance tradition, it seemed highly unlikely that the silence in that literature concerning subjects' reports on their mental processes was due to simple failure to ask subjects the relevant questions. Instead, it seemed more likely that subjects had been asked, and asked often, but that their answers had failed to reflect any ability to report the inferences that investigators believed to have occurred. If so, those investigators, like Nisbett and his colleagues, might have failed to report the answers because they made little sense in terms of the traditional assumptions about the conscious, verbalizable nature of cognitive processes. Accordingly, we contacted two of the most prolific and innovative researchers in that tradition—E. Aronson and P. Zimbardo—and asked them if they had ever quizzed their subjects about their mental processes. They had indeed, with results similar to those described above.

Aronson (Note 1) responded as follows:

> We occasionally asked our subjects why they had responded as they did. The results were very disappointing. For example, in the initiation experiment (Aronson & Mills, 1959), subjects did a lot of denying when asked if the punishment had affected their attitudes toward the group or had entered into their thinking at all. When I explained the theory to the subjects, they typically said it was very plausible and that many subjects had probably reasoned just the way I said, but not they themselves.

Zimbardo (Note 2) gave a similar account:

> Pretest subjects were routinely asked why they had behaved as they did. I don't remember any subject who ever described anything like the process of dissonance reduction that we knew to have occurred. For example, in the shock experiment (Zimbardo, Cohen, Weisenberg, Dworkin, & Firestone, 1969), we pointed out to experimental subjects that they had learned more quickly the second time. A typical response would have been, "I guess maybe you turned the shock down." Or, in the grasshopper experiment (Zimbardo, Weisenberg, Firestone, & Levy, 1969), we asked subjects why they had been willing to eat a grasshopper. A typical response would have been, "Well, it was just no big deal whether I ate a grasshopper or not."

Thus the explanations that subjects offer for their behavior in insufficient-justification and attribution experiments are so removed from the processes that investigators presume to have occurred as to give grounds for considerable doubt that there is direct access to these processes. This doubt would remain, it should be noted, even if it were eventually to be shown

that processes other than those posited by investigators were responsible for the results of these experiments. Whatever the inferential process, the experimental method makes it clear that something about the manipulated stimuli produces the differential results. Yet subjects do not refer to these critical stimuli in any way in their reports on their cognitive processes.

As a final point, we note Kelley's (1967) observation that the results of insufficient-justification experiments could never be obtained if subjects were aware of the critical role played by the social pressure from the experimenter. If subjects realized that their behavior was produced by this social pressure, they would not change their attitudes so as to move them into line with their behavior, because they would realize that their behavior was governed by the social pressure and not by their attitudes. We concur with Kelley's view that this fundamental unawareness of the critical role of the experimenter's behavior is essential to the erroneous attitude inferences obtained in these experiments.

Other Research on Verbal Reports About Cognitive Processes

There are at least five other literatures bearing on the question of the ability of subjects to report accurately about the effects of stimuli on complex, inferential responses: (a) The learning-without-awareness literature, (b) the literature on subject ability to report accurately on the weights they assign to particular stimulus factors in complex judgment tasks (reviewed by Slovic & Lichtenstein, 1971), (c) some of the literature on subliminal perception, (d) the classic Maier (1931) work on awareness of stimuli influencing problem solving, and (e) work by Latané and Darley (1970) on awareness of the effect of the presence of other people on helping behavior.

We shall discuss the first two areas of research in a later context, and it would take us far afield to review the subliminal perception literature in its entirety. Brief mention of the current status of the subliminal perception question is in order, however, since it bears directly on the issue of subject ability to report accurately on the effects of stimuli. If, as some writers claim, stimuli can be responded to in the literal absence of awareness of their existence, then it logically follows that they could not possibly report on the influence of those stimuli on their responses.

Subliminal Perception

The subliminal perception question has had a stormy, controversial history, chronicled by Dixon (1971). It is fair to say, however, that the basic question of whether people can respond to a stimulus in the absence of the ability to verbally report on its existence would today be answered in the affirmative by many more investigators than would have been the case a decade ago. The reasons for this have been reviewed by Dixon (1971) and Erdelyi (1974). The new acceptance rests on (a) methodological innovations in the form chiefly of signal detection techniques and dichotic listening procedures and (b) persuasive theoretical arguments by Erdelyi (1974) and others that have succeeded in deriving the subliminal perception phenomenon from the notion of selective attention and filtering (Broadbent, 1958; Moray, 1969; Treisman, 1969).

An example of recent research employing signal detection and dichotic listening procedures is provided by W. R. Wilson (1975). Wilson played tone sequences into the unattended auditory channel while subjects tracked a human voice in the attended channel. Subjects subsequently reported having heard no tones, in fact, nothing at all, in the unattended channel. Moreover, in a signal detection task, subjects were presented (binaurally) with tone sequences which were either new or which had previously been presented up to five times in the unattended channel. Subjects were unable to discriminate new from old stimuli at a level exceeding chance. Despite this fact, subjects showed the traditional familiarity effect on liking of the tone stimuli (Zajonc, 1968). "Familiar" tone sequences, that is, tone sequences previously presented to the unattended channel, were preferred to novel stimuli. Wilson argued that the experiment provides evidence that affective processes are triggered by information that is too weak to produce subsequent verbal recognition.

Results such as those provided by Wilson are

well understood in terms of recent theoretical developments in the field of attention and memory. It is now generally recognized (Erdelyi, 1974; Mandler, 1975b) that many more stimuli are apprehended than can be stored in short-term memory or transferred to long-term memory. Thus, subliminal perception, once widely regarded as a logical paradox ("How can we perceive without perceiving"), may be derived as a logical consequence of the principle of selective filtering. We cannot perceive without perceiving, but we can perceive without remembering. The subliminal perception hypothesis then becomes theoretically quite innocuous: Some stimuli may affect ongoing mental processes, including higher order processes of evaluation, judgment, and the initiation of behavior, without being registered in short-term memory, or at any rate without being transferred to long-term memory.

Thus if recent data and theory are correct in their implications, it follows that subjects sometimes cannot report on the existence of influential stimuli. It therefore would be quite impossible for them to describe accurately the role played by these stimuli in influencing their responses; and any subsequent verbal report by subjects about the cause of their responses would be at least partially in error.

Reports on Problem-Solving Processes

There is a striking uniformity in the way creative people—artists, writers, mathematicians, scientists, and philosophers—speak about the process of production and problem solving. Ghiselin (1952) has collected into one volume a number of essays on the creative process by a variety of creative workers from Poincaré to Picasso. As Ghiselin accurately described the general conclusion of these workers, "Production by a process of purely conscious calculation seems never to occur" (p. 15). Instead, creative workers describe themselves almost universally as bystanders, differing from other observers only in that they are the first to witness the fruits of a problem-solving process that is almost completely hidden from conscious view. The reports of these workers are characterized by an insistence that (a) the influential stimuli are usually completely obscure —the individual has no idea what factors prompted the solution; and (b) even the fact that a process is taking place is sometimes unknown to the individual prior to the point that a solution appears in consciousness.

Some quotations from Ghiselin's (1952) collection will serve to illustrate both these points. The mathematician Jacques Hadamard reports that "on being very abruptly awakened by an external noise, a solution long searched for appeared to me at once without the slightest instant of reflection on my part . . . and in a quite different direction from any of those which I previously tried to follow" (p. 15). Poincaré records that "the changes of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry" (p. 37).

Whitehead writes of "the state of imaginative muddled suspense which precedes successful inductive generalization" (Ghiselin, 1952, p. 15), and Stephen Spender describes "a dim cloud of an idea which I feel must be condensed into a shower of words" (p. 15). Henry James speaks of his deliberate consignment of an idea to the realm of the unconscious where it can be worked upon and realized: "I was charmed with my idea, which would take, however, much working out; and because it had so much to give, I think, must I have dropped it for the time into the deep well of unconscious cerebration: not without the hope, doubtless, that it might eventually emerge from that reservoir, as one had already known the buried treasure to come to light, with a firm iridescent surface and a notable increase of weight" (p. 26).

That mundane problem-solving in everyday life differs little, in its degree of consciousness, from the problem-solving of creative geniuses, is indicated by the very elegant work of Maier, done some 45 years ago. In Maier's (1931) classic experiment, two cords were hung from the ceiling of a laboratory strewn with many objects such as poles, ringstands, clamps, pliers, and extension cords. The subject was told that his task was to tie the two ends of the cords

together. The problem in doing so was that the cords were placed far enough apart that the subject could not, while holding onto one cord, reach the other. Three of the possible solutions, such as tying an extension cord to one of the ceiling cords, came easily to Maier's subjects. After each solution, Maier told his subjects "Now do it a different way." One of the solutions was much more difficult than the others, and most subjects could not discover it on their own. After the subject had been stumped for several minutes, Maier, who had been wandering around the room, casually put one of the cords in motion, Then, typically within 45 seconds of this cue, the subject picked up a weight, tied it to the end of one of the cords, set it to swinging like a pendulum, ran to the other cord, grabbed it, and waited for the first cord to swing close enough that it could be seized. Immediately thereafter, Maier asked the subject to tell about his experience of getting the idea of a pendulum. This question elicited such answers as "It just dawned on me." "It was the only thing left." "I just realized the cord would swing if I fastened a weight to it." A psychology professor subject was more inventive: "Having exhausted everything else, the next thing was to swing it. I thought of the situation of swinging across a river. I had imagery of monkeys swinging from trees. This imagery appeared simultaneously with the solution. The idea appeared complete."

Persistent probing after the free report succeeded in eliciting reports of Maier's hint and its utilization in the solution of the problem from slightly less than a third of the subjects. This fact should be quickly qualified, however, by another of Maier's findings. Maier was able to establish that one particular cue—twirling a weight on a cord—was a useless hint, that is, subjects were not aided in solving the problem by exposure to this cue. For some of the subjects, this useless cue was presented prior to the genuinely helpful cue. All of these subjects reported that the useless cue had been helpful and denied that the critical cue had played any role in their solution. These inaccurate reports cast doubt on any presumption that even the third of Maier's subjects who accurately reported that they used the helpful cue were reporting such use on the basis of genuinely insightful introspection, since when they were offered a false "decoy" cue they preferred it as an explanation for their solution.

Reports on the Effects of the Presence of Others on Helping Behavior

Latané and Darley (1970) have shown, in a large number of experiments in a wide variety of settings, that people are increasingly less likely to help others in distress as the number of witnesses or bystanders increases. Thus, for example, the more people who overhear an individual in another room having what sounds like an epileptic seizure, the lower the probability that any given individual will rush to help. Latané and Darley early became intrigued by the fact that their subjects seemed utterly unaware of the influence of the presence of other people on their behavior. Accordingly, they systematically asked the subjects in each of their experiments whether they thought they had been influenced the the presence of other people. "We asked this question every way we knew how: subtly, directly, tactfully, bluntly. Always we got the same answer. Subjects persistently claimed that their behavior was not influenced by the other people present. This denial occurred in the face of results showing that the presence of others did inhibit helping" (p. 124). It should also be noted that when Latané and Darley described their experiments in detail to other subjects and asked these subjects to predict how others, and they themselves, would behave when alone or with other people present, these observer subjects uniformly agreed that the presence of other people would have no effect on their own or other people's behavior.

Thus the literature contains evidence from domains other than insufficient-justification and attribution research suggesting that people may have little ability to report accurately about their cognitive processes. The subliminal perception literature suggests that people may sometimes be unable to report even the existence of influential stimuli, and anecdotal reports of creative workers suggest that this may frequently be the case in problem-solving. In addition, these anecdotal reports suggest the most extreme form of inaccessibility to cognitive processes—literal lack of awareness that a

process of any kind is occurring until the moment that the result appears. The work of Maier and of Latané and Darley additionally suggests that even when subjects are thoroughly cognizant of the existence of the relevant stimuli, and of their responses, they may be unable to report accurately about the influence of the stimuli on the responses.

Demonstrations of Subject Inability to Report Accurately on the Effects of Stimuli on Responses

Though the evidence we have reviewed is consistent with the skepticism expressed by cognitive psychologists concerning people's ability to introspect about their cognitive processes, the evidence is limited in several respects. Dissonance and attribution processes may be unique in important ways. For example, deceptive practices are often employed in structuring the stimulus situations in such experiments, and these practices may result in people being misled in ways that do not normally occur in daily life. The subliminal perception literature is controversial, and though the new data and theoretical arguments have proved to be convincing to many investigators, the critical response to these new developments has not been formulated, and it may yet prove to be as devastating as the previous wave of criticism was to older evidence and formulations. The evidence on problem-solving processes is anecdotal, except for one series of experiments employing a single type of problem. That particular problem, moreover, was a spatial one, and subjects may find special difficulty in reporting verbally about spatial reasoning. The Latané and Darley findings are impressive, but they deal with awareness of only a single type of response. Moreover, subjects and even observers may be highly motivated to deny the role of such a trivial factor as the presence of others in such an important ethical domain as the rendering of help to another human in distress.

In order to fill in the gaps in the literature, we have performed a series of small studies investigating people's ability to report accurately on the effects of stimuli on their responses. They were designed with several criteria in mind:

1. The cognitive processes studied were of a routine sort that occur frequently in daily life. Deception was used minimally, and in only a few of the studies.
2. Studies were designed to sample a wide range of behavioral domains, including evaluations, judgments, choices, and predictions.
3. Care was taken to establish that subjects were thoroughly cognizant of the existence of both the critical stimulus and their own responses.
4. With two exceptions, the critical stimuli were verbal in nature, thus reducing the possibility that subjects could be cognizant of the role of the critical stimulus but simply unable to describe it verbally.
5. Most of the stimulus situations were designed to be as little ego-involving as possible so that subjects would not be motivated on grounds of social desirability or self-esteem maintenance to assert or deny the role of particular stimuli in influencing their responses.

The reader is entitled to know that the stimulus situations were chosen in large part because we felt that subjects would be wrong about the effects of the stimuli on their responses. We deliberately attempted to study situations where we felt that a particular stimulus would exert an influence on subjects' responses but that subjects would be unable to detect it, and situations where we felt a particular stimulus would be ineffective but subjects would believe it to have been influential. It is even more important to note, however, that we were highly unsuccessful in this attempted bias. In general, we were no more accurate in our predictions about stimulus effects than the subjects proved to be in their reports about stimulus effects. Most of the stimuli that we expected to influence subjects' responses turned out to have no effect, and many of the stimuli that we expected to have no effect turned out to be influential.

In all of the studies, some component of a complex stimulus situation was manipulated and the impact of this stimulus component on responses could thus be assessed. Subjects, as it turned out, were virtually never accurate in their reports. If the stimulus component had a

significant effect on responses, subjects typically reported that it was noninfluential; if the stimulus component had no significant effect, subjects typically reported that it had been influential.

Failure to Report the Influence of Effective Stimulus Factors

Erroneous Reports about Stimuli Influencing Associative Behavior

The phenomenon of verbal association seemed a fruitful one for illustrating an inability to report accurately about the role of influential stimuli. For example, it seems likely that simultaneous associative behavior—when two people speak the same thought or begin humming the same tune at the same time—may occur because of the presence of some stimulus which sets off identical associative processes in the two people. Then, because these associative processes are hidden from conscious view, both parties are mystified about the occurrence of the "coincidental" mutual behavior.

In order to test subject ability to report influences on their associative behavior, we had 81 male introductory psychology students memorize a list of word pairs. Some of these word pairs were intended to generate associative processes that would elicit certain target words in a word association task to be performed at a later point in the experiment. For example, subjects memorized the word pair "ocean–moon" with the expectation that when they were later asked to name a detergent, they would be more likely to give the target "Tide" than would subjects who had not previously been exposed to the word pairs. In all, eight word pair cues were employed, and all eight did in fact have the effect of increasing the probability of target responses in the word association task. The average effect of the semantic cuing was to double the frequency of target responses, from 10% to 20% ($p < .001$). Immediately following the word association task, subjects were asked in open-ended form why they thought they had given each of their responses in the word association task. Despite the fact that nearly all subjects could recall nearly all of the words pairs, subjects almost never mentioned a word pair cue as a reason for giving a particular target response. Instead, subjects focused on some distinctive feature of the target ("Tide is the best-known detergent"), some personal meaning of it ("My mother uses Tide"), or an affective reaction to it ("I like the Tide box"). When specifically asked about any possible effect of the word cues, approximately a third of the subjects did say that the words had probably had an effect, but there is reason to doubt that these reports were indications of any true awareness. An "awareness ratio" was calculated for each target word. This was the number of subjects who reported an influence of the cues divided by the number of subjects who were influenced by the cues. This latter number was an estimate, based on the number of cued subjects who gave the target response minus the number of uncued subjects who gave the target response. These awareness ratios for the eight target words ranged from 0 to 244%. This means that for some of the target words, none of the subjects reported any influence of the word cues, and for others, many more subjects reported an influence than were probably influenced.

Erroneous Reports about Position Effects on Appraisal and Choice

We conducted two studies that serendipitously showed a position effect on evaluation of an array of consumer goods. (We had attempted, unsuccessfully, to manipulate the smell of garments in the array.) In both studies, conducted in commercial establishments under the guise of a consumer survey, passersby were invited to evaluate articles of clothing—four different nightgowns in one study (378 subjects) and four identical pairs of nylon stockings in the other (52 subjects). Subjects were asked to say which article of clothing was the best quality and, when they announced a choice, were asked why they had chosen the article they had. There was a pronounced left-to-right position effect, such that the right-most object in the array was heavily overchosen. For the stockings, the effect was quite large, with the right-most stockings being preferred over the left-most by a factor of almost four to one. When asked about the reasons for their choices, no subject ever mentioned spontaneously the position of the article in the

array. And, when asked directly about a possible effect of the position of the article, virtually all subjects denied it, usually with a worried glance at the interviewer suggesting that they felt either that they had misunderstood the question or were dealing with a madman.

Precisely why the position effect occurs is not obvious. It is possible that subjects carried into the judgment task the consumer's habit of "shopping around," holding off on choice of early-seen garments on the left in favor of later-seen garments on the right.

Erroneous Reports about Anchoring Effects on Predictions

In an unpublished study by E. Borgida, R. Nisbett, and A. Tversky, subjects (60 introductory psychology students) were asked to guess what the average behavior of University of Michigan students would be in three different experiments. Some subjects were given an "anchor" in the form of knowledge about the behavior of a particular "randomly chosen subject." Of the anchor subjects, some were given only information about the individual's behavior, while others were also shown a brief videotaped interview with the individual. It had been anticipated that the videotape would increase the salience and vividness of the anchor and that subjects who were exposed to it would show a greater anchoring effect, that is, that their estimates of the average behavior of the sample would cluster more closely about the anchor value. Only weak support for the prediction was found, and anchoring effects across the experiments described to subjects ranged from huge and highly statistically significant ones down to actual "anti-anchoring" effects (i.e., somewhat greater variance of estimates of the sample average for anchor conditions than for the no-anchor condition). This range of effects, however, made possible a test of subjects' ability to report on their utilization of the anchor value. Immediately after making their estimates of average sample behavior, subjects were asked about the extent to which they had relied on knowledge about the particular individual's behavior in making these estimates. Subjects reported moderate utilization of the anchor value in all conditions. Thus they reported the same degree of utilization of the anchor value for experiments where it had not been used at all as they did for experiments where it had heavily influenced their estimates.

Erroneous Reports about the Influence of an Individual's Personality on Reactions to his Physical Characteristics

Perhaps the most remarkable of the demonstrations is one we have described in detail elsewhere (Nisbett & Wilson, in press). This study, an experimental demonstration of the halo effect, showed that the manipulated warmth or coldness of an individual's personality had a large effect on ratings of the attractiveness of his appearance, speech, and mannerisms, yet many subjects actually insisted that cause and effect ran in the opposite direction. They asserted that their feelings about the individual's appearance, speech, and mannerisms had influenced their liking of him.

Subjects were shown an interview with a college teacher who spoke English with a European accent. The interview dealt with teaching practices and philosophy of education. Half the subjects saw the teacher answering the questions in a pleasant, agreeable, and enthusiastic way (warm condition). The other half saw an autocratic martinet, rigid, intolerant, and distrustful of his students (cold condition). Subjects then rated the teacher's likability and rated also three attributes that were by their nature essentially invariant across the two experimental conditions: his physical appearance, his mannerisms, and his accent. Subjects who saw the warm version of the interview liked the teacher much better than subjects who saw the cold version of the interview, and there was a very marked halo effect. Most of the subjects who saw the warm version rated the teacher's appearance, mannerisms, and accent as attractive, while a majority of subjects who saw the cold version rated these qualities as irritating. Each of these differences was significant at the .001 level.

Some subjects in each condition were asked if their liking for the teacher had influenced their ratings of the three attributes, and some were asked if their liking for each of the three attributes had influenced their liking of the teacher. Subjects in both warm and cold condi-

tions strongly denied any effect of their overall liking for the teacher on ratings of his attributes. Subjects who saw the warm version also denied that their liking of his attributes had influenced their overall liking. But subjects who saw the cold version asserted that their disliking of each of the three attributes had lowered their overall liking for him. Thus it would appear that these subjects precisely inverted the true causal relationship. Their disliking of the teacher lowered their evaluation of his appearance, his mannerisms, and accent, but subjects denied such an influence and asserted instead that their dislike of these attributes had decreased their liking of him!

Reporting the Influence of Ineffective Stimulus Factors

Three of our demonstrations involved the manipulation of stimulus factors that turned out to have no effect on subjects' judgments. In each of these studies, subjects reported that at least some of these actually ineffective factors had been highly influential in their judgments.

Erroneous Reports about the Emotional Impact of Literary Passages

In the first of these studies, 152 subjects (introductory psychology students) read a selection from the novel *Rabbit, Run* by John Updike. The selection described an alcoholic housewife who has just been left by her husband and who is cleaning up her filthy home in preparation for a visit by her mother. While drunkenly washing her infant girl, she accidentially allows the child to drown. The selection is well written and has a substantial emotional impact even when read out of the context of the rest of the novel. There were four conditions of the experiment. In one condition, subjects read the selection as it was written. In a second condition, a passage graphically describing the messiness of the baby's crib was deleted. In a third condition, subjects read the selection minus a passage physically describing the baby girl. In the fourth condition, both passages were deleted.

After reading the selection, all subjects were asked what emotional impact it had had. Then the manipulated passages were presented, and subjects were asked how the presence of the passage had affected (or would have affected, for subjects for whom the passage was deleted), the emotional impact of the selection. As it turned out, there was no detectable effect on reported emotional impact due to inclusion versus deletion of either passage. (Both pairs of means differed by less than .10 on a 7-point scale.) Subjects reported, however, that the passages had increased the impact of the selection. Subjects exposed to the passage describing the messiness of the baby's crib were virtually unanimous in their opinion that the passage had increased the impact of the selection: 86% said the passage had increased the impact. Two thirds of the subjects exposed to the physical description of the baby reported that the passage had had an effect, and of those who reported it had an effect, two and a half times as many subjects said it had increased the impact of the selection as said it decreased the impact. The subjects who were not exposed to the passages on the initial reading predicted that both passages would have increased the impact of the selection had they been included. Predicted effects by these subjects were in fact extremely close to the pattern of (erroneous) reported effects by subjects who were exposed to the passages.

Erroneous Reports about the Effects of Distractions on Reactions to a Film

In another study, 90 subjects (introductory psychology students) were asked to view a brief documentary on the plight of the Jewish poor in large cities. Some subjects viewed the film while a distracting noise (produced by a power saw) occurred in the hall outside. Other subjects viewed the film while the focus was poorly adjusted on the projector. Control subjects viewed the film under conditions of no distraction. After viewing the film, subjects rated it on three dimensions—how interesting they thought it was, how much they thought other people would be affected by it, and how sympathetic they found the main character to be. Then, for experimental conditions, the experimenter apologized for the poor viewing conditions and asked subjects to indicate next to each rating whether he had been influenced by

the noise or poor focus. Neither the noise nor the poor focus actually had any detectable effect on any of the three ratings. (Ratings were in general trivially *higher* for distraction subjects.) In the first and only demonstration of reasonably good accuracy in subject report of stimulus effects we found, most of the subjects in the poor focus condition actually reported that the focus had not affected their ratings (although 27% of the subjects reported that the focus had lowered at least one rating, a proportion significantly different from zero). A majority of subjects in the noise condition, however, erroneously reported that the noise had affected their ratings. Fifty-five per cent of these subjects reported that the noise had lowered at least one of their ratings.

Erroneous Reports about the Effects of Reassurance on Willingness to Take Electric Shocks

In a third study, 75 subjects (male introductory psychology students) were asked to predict how much shock they would take in an experiment on the effects of intense electric shocks. One version of the procedural protocol for the experiment included a "reassurance" that the shocks would do "no permanent damage." The other version did not include this "reassurance." Subjects receiving the first version were asked if the phrase about permanent damage had affected their predictions about the amount of shock they would take, and subjects receiving the second version were asked if the phrase would have affected their predictions, had it been included. Inclusion of the phrase in fact had no effect on predicted shock taking, but a majority of subjects reported that it did. Of those reporting an effect, more than 80% reported it had increased their predictions. Subjects who had not received the phrase were similarly, and erroneously, inclined to say that it would have increased their willingness to take shock had it been included.

Taken together, these studies indicate that the accuracy of subject reports about higher order mental processes may be very low. We wish to acknowledge that there are methodological and interpretive problems with some of the individual studies, however. Although the magnitude of effects induced by effective critical stimuli ranged from a ratio of 2:1 over control values to a ratio of 4:1, the critical stimuli may often have been merely necessary and not sufficient causes of the responses in question. Therefore subjects may often have been correct in asserting that some other stimulus was a more important determinant of their responses. In studies where the manipulated stimuli were ineffective (e.g., the literary passage and distraction studies), it is conceivable that perceived experimenter demands could have contributed to the results. And finally, in some of the studies it could be argued that the subjects denied the role of the influential stimulus in order to avoid looking silly or foolish (e.g., the position effect study), and not because they were unaware of its causal role.

We also wish to acknowledge that the studies do not suffice to show that people *could never* be accurate about the processes involved. To do so would require ecologically meaningless but theoretically interesting procedures such as interrupting a process at the very moment it was occurring, alerting subjects to pay careful attention to their cognitive processes, coaching them in introspective procedures, and so on. What the studies do indicate is that such introspective access as may exist is not sufficient to produce accurate reports about the role of critical stimuli in response to questions asked a few minutes or seconds after the stimuli have been processed and a response produced.

The Origin of Verbal Reports About Cognitive Processes

The Fount That Never Was

In summary, it would appear that people may have little ability to report accurately on their cognitive processes:

1. Sometimes, as in many dissonance and attribution studies, people are unable to report correctly even about the existence of the evaluative and motivational responses produced by the manipulations.

2. Sometimes, as in dissonance and attribution studies, and in the reports of creative artists and scientists, people appear to be unable to report that a cognitive process has occurred.

3. Sometimes, as in the subliminal perception literature and the reports of creative workers, people may not be able to identify the existence of the critical stimulus.

4. Even when people are completely cognizant of the existence of both stimulus and response, they appear to be unable to report correctly about the effect of the stimulus on the response. This is true in dissonance and attribution studies, in the subliminal perception literature, in the reports of creative workers, and in the work by Maier (1931), Latané and Darley (1970), and in our own studies described above.

In addition, we might point out that at least some psychological phenomena probably would not occur in the first place if people were aware of the influence of certain critical stimuli. For example, if people were aware of the effects of the presence of other people on their tendency to offer help to a person in distress, they would surely strive to counteract that influence, and would therefore not show the typical effect. Similarly, if people were aware of position effects on their evaluations, they would attempt to overcome these effects. A number of other phenomena would seem to depend on lack of awareness of the role played by certain critical factors, for example, halo effects, contrast effects, and order effects. If people knew that their judgments were subject to influence from other judgments made about an object or from judgments just previously made about other objects, or from the order in which the object was examined, then they would correct for such influences and these effects would not exist.

Polanyi (1964) and others (e.g., Gross, 1974) have argued persuasively that "we can know more than we can tell," by which it is meant that people can perform skilled activities without being able to describe what they are doing and can make fine discriminations without being able to articulate their basis. The research described above suggests that the converse is also true—that we sometimes tell more than we can know. More formally, people sometimes make assertions about mental events to which they may have no access and these assertions may bear little resemblance to the actual events.

The evidence reviewed is then consistent with the most pessimistic view concerning people's ability to report accurately about their cognitive processes. Though methodological implications are not our chief concern, we should note that the evidence indicates it may be quite misleading for social scientists to ask their subjects about the influences on their evaluations, choices, or behavior. The relevant research indicates that such reports, as well as predictions, may have little value except for whatever utility they may have in the study of verbal explanations per se.

More importantly, the evidence suggests that people's erroneous reports about their cognitive processes are not capricious or haphazard, but instead are regular and systematic. Evidence for this comes from the fact that "observer" subjects, who did not participate in experiments but who simply read verbal descriptions of them, made predictions about the stimuli which were remarkably similar to the reports about the stimuli by subjects who had actually been exposed to them. In experiments by Latané and Darley (1970), and in several of our own studies, subjects were asked to predict how they themselves, or how other people, would react to the stimulus situations that had actually been presented to other subjects. The observer subjects made predictions that in every case were similar to the erroneous reports given by the actual subjects. Thus Latané and Darley's original subjects denied that the presence of other people had affected their behavior, and observer subjects also denied that the presence of others would affect either their own or other people's behavior. When our word association study was described to observer subjects, the judgments of the probability that particular word cues would affect particular target responses were positively correlated with the original subjects' "introspective reports" of the effects of the word cues on the target responses. (Both subject reports and observer predictions were slightly negatively correlated with true cuing effects.) In two of our other studies, subjects were asked to predict how they would have responded to stimuli that were actually presented to subjects in another condition. In both cases, predictions about behavior were very similar to the inaccurate reports of subjects who had actually been exposed to the conditions. Thus, whatever capacity for introspection exists, it

does not produce accurate reports about stimulus effects, nor does it even produce reports that differ from predictions of observers operating only with a verbal description of the stimulus situation. As Bem (1967) put it in a similar context, if the reports of subjects do not differ from the reports of observers, then it is unnecessary to assume that the former are drawing on "a fount of privileged knowledge" (p. 186). It seems equally clear that subjects and observers are drawing on a similar source for their verbal reports about stimulus effects. What might this be?

A Priori Causal Theories

We propose that when people are asked to report how a particular stimulus influenced a particular response, they do so not by consulting a memory of the mediating process, but by applying or generating causal theories about the effects of that type of stimulus on that type of response. They simply make judgments, in other words, about how plausible it is that the stimulus would have influenced the response. These plausibility judgments exist prior to, or at least independently of, any actual contact with the particular stimulus embedded in a particular complex stimulus configuration. Causal theories may have any of several origins.

1. The culture or a subculture may have explicit rules stating the relationship between a particular stimulus and a particular response ("I came to a stop because the light started to change." "I played a trump because I had no cards in the suit that was led").

2. The culture or a subculture may supply implicit theories about causal relations. In Abelson's (1968) terms, the presence of a particular stimulus may "psychologically imply" a particular response ("Jim gave flowers to Amy [me]; that's why she's [I'm] acting pleased as punch today"). In Kelley's (1972) terms, people growing up in a given culture learn certain "causal schemata," psychological rules governing likely stimulus–response relations ("The ballplayer [I] was paid to endorse Aqua-Velva, that's the only reason he [I] endorsed it").

3. An individual may hold a particular causal theory on the basis of empirical observation of covariation between stimuli of the general type and responses of the general type. ("I'm grouchy today. I'm always grouchy when I don't break 100 in golf.") There is reason to suspect, however, that actual covariation may play less of a role in perceived or reported covariation than do theories about covariation. The Chapmans (Chapman, 1967; Chapman & Chapman, 1967, 1969) have shown that powerful covariations may go undetected when the individual lacks a theory leading him to suspect covariation and, conversely, that the individual may perceive covariation where there is none if he has a theory leading him to expect it. The present position, of course, leads to the expectation that people would be as subject to theory-induced errors in self-perception as in the perception of convariation among purely external events.

4. In the absence of a culturally supplied rule, implicit causal theory, or assumption about covariation, people may be able to generate causal hypotheses linking even novel stimuli and novel responses. They may do so by searching their networks of connotative relations surrounding the stimulus description and the response description. If the stimulus is connotatively similar to the response, then it may be reported as having influenced the response. To the extent that people share similar connotative networks they would be expected to arrive at similar judgments about the likelihood of a causal link between stimulus and response.

We do not wish to imply that all or even most a priori causal theories are wrong. Verbal reports relying on such theories will typically be wrong not because the theories are in error in every case but merely because they are incorrectly applied in the particular instance.

The tools that people employ when asked to make judgments about causality are analogous to the "representativeness heuristic" described by Tversky and Kahneman (1973, 1974; Kahneman & Tversky, 1973). These writers have proposed that when making judgments about the probability that an individual is, say, a librarian, one does so by comparing his information about the individual with the contents of his stereotype concerning librarians. If the information is representative of the con-

tents of the stereotype concerning librarians, then it is deemed "probable" that the individual is a librarian. Information that is more pertinent to a true probability judgment, such as the proportion of librarians in the population, is ignored. We are proposing that a similar sort of representativeness heuristic is employed in assessing cause and effect relations in self-perception. Thus a particular stimulus will be deemed a representative cause if the stimulus and response are linked via a rule, an implicit theory, a presumed empirical covariation, or overlapping connotative networks.

In the experiments reviewed above, then, subjects may have been making simple representativeness judgments when asked to introspect about their cognitive processes. Worry and concern seem to be representative, plausible reasons for insomnia while thoughts about the physiological effects of pills do not. Seeing a weight tied to a string seems representative of the reasons for solving a problem that requires tying a weight to a cord, while simply seeing the cord put into motion does not. The plight of a victim and one's own ability to help him seem representative of reasons for intervening, while the sheer number of other people present does not. The familiarity of a detergent and one's experience with it seem representative of reasons for its coming to mind in a free association task, while word pairs memorized in a verbal learning experiment do not. The knit, sheerness, and weave of nylon stockings seem representative of reasons for liking them, while their position on a table does not. And a reassurance that electric shock will cause no permanent damage seems representative of reasons for accepting shock; reading about the behavior of a particular experimental subject (the "anchor" value) seems representative of the reasons for choosing a similar behavior as the average value for the subject population as a whole; a passage graphically describing the physical characteristics of a child seems representative of reasons for being emotionally affected by a literary selection ending with the death of a child, and a distracting noise seems representative of reasons for not liking a film.

When subjects were asked about their cognitive processes, therefore, they did something that may have felt like introspection but which in fact may have been only a simple judgment of the extent to which input was a representative or plausible cause of output. It seems likely, in fact, that the subjects in the present studies, and ordinary people in their daily lives, do not even attempt to interrogate their memories about their cognitive processes when they are asked questions about them. Rather, they may resort in the first instance to a pool of culturally supplied explanations for behavior of the sort in question or, failing in that, begin a search through a network of connotative relations until they find an explanation that may be adduced as psychologically implying the behavior. Thus if we ask another person why he enjoyed a particular party and he responds with "I liked the people at the party," we may be extremely dubious as to whether he has reached this conclusion as a result of anything that might be called introspection. We are justified in suspecting that he has instead asked himself Why People Enjoy Parties and has come up with the altogether plausible hypothesis that in general people will like parties if they like the people at the parties. Then, his only excursion into his storehouse of private information would be to make a quick check to verify that his six worst enemies were not at the party. If not, he confidently asserts that the people-liking was the basis of his party-liking. He is informationally superior to observers, in this account, only by virtue of being able to make this last-minute check of his enemies list, and not by virtue of any ability to examine directly the effects of the stimuli (the people) on his response (enjoyment).

The present view carries two important implications that go beyond a merely anti-introspectivist position: (a) People's reports will sometimes be correct, and it should be possible to predict when they will be likely to be correct. (b) People's reports about their higher mental processes should be neither more nor less accurate, in general, then the predictions about such processes made by observers. An experiment by Nisbett and Bellows (Note 3), reported below, tested both these implications.

Accuracy of Subject Reports and Observer Predictions

The above analysis implies that it should be possible to demonstrate accuracy and inaccu-

racy in verbal reports in the same experiment by simply asking subjects to make two sorts of judgments—those for which the influential factors are plausible and are included in a priori causal theories, and others which are influenced by implausible factors not included in such theories. In the former case, both subjects and observers should be accurate; in the latter, neither should be accurate.

Nisbett and Bellows (Note 3) asked female subjects to read a lengthy description of a woman who was applying for a job as a counselor in a crisis intervention center. Subjects read what they believed was the application portfolio, a lengthy document including a letter of recommendation and a detailed report of an interview with the center's director. Five stimulus factors were manipulated. (a) The woman's appearance was either described in such a way as to make it clear that she was quite physically attractive, or nothing was said about her appearance. (b) The woman was either described as having superb academic credentials, or nothing was said about her academic credentials. (c) The woman was described as having spilled a cup of coffee over the interviewer's desk, or nothing was said about any such incident. (d) The woman was described as having been in a serious auto accident, or nothing was said about an accident. (e) Subjects were either told that they would meet the woman whose folder they were reading, or they were told that they would meet some other applicant. These stimuli were manipulated factorially.

After reading the portfolio, subjects were asked to make four judgments about the woman: (a) how much they liked her, (b) how sympathetic they thought she would be toward clients' problems, (c) how intelligent they thought she was, and (d) how flexible they thought she would be in dealing with clients' problems. Then subjects were asked how each of the factors (ranging from 0 for some subjects up to 5 for others) had influenced each of the four judgments.

In addition, "observer" subjects were asked to state how each of the five factors would influence each of the four judgments. These subjects did not read any portfolio and, indeed, the factors were described only in summary form (e.g., "Suppose you knew that someone was quite physically attractive. How would that influence how much you would like the person?"). Both observers and subjects answered these questions on 7-point scales ranging from "increase(d) my liking a great deal" to "decrease(d) my liking a great deal."

Two predictions about the results of the study follow from the present analysis.

1. Subjects should be much more accurate in their reports about the effects of the stimulus factors on the intelligence judgment than in their reports about the effects of the factors on their other judgments. This is because the culture specifies more clearly what sorts of factors ought to influence a judgment of intelligence, and in what way they should do so, than it does for judgments such as liking, sympathy toward others, or flexibility. In fact, the other factor–judgment combinations were chosen with malice aforethought. Recent work by social psychologists has shown that several of the factors have implausible effects on several of the judgments, for example, people tend to give more favorable ratings on a number of dimensions to people whom they believe they are about to meet than to people whom they do not expect to meet (Darley & Berschied, 1967).

2. Whether subjects are generally accurate in reports about the effects of the factors on a given judgment or generally inaccurate, their accuracy will be equalled by observers working from impoverished descriptions of the factors.

The results gave the strongest possible support to both predictions. Mean subject reports about the effects of the factors, mean observer reports, and mean actual effects (experimental minus control means) were compared for each of the judgments. The most remarkable result was that subject and observer reports of factor utilization were so strongly correlated for each of the judgments that it seems highly unlikely that subjects and observers could possibly have arrived at these reports by different means. Mean subject and observer reports of factor utilization were correlated .89 for the liking judgment, .84 for the sympathy judgment, .99 for the intelligence judgment, and .77 for the flexibility judgment. Such strong correspondence between subject and observer reports suggests that both groups produced these reports

via the same route, namely by applying or generating similar causal theories.

As anticipated, subject accuracy was extremely high for the intelligence judgment. Subject reports about the effects of the factors were correlated .94 with true effects of the factors. Also as anticipated, however, observer predictions were fully as accurate as subject reports: Observer predictions were correlated .98 with true effects of the factors on the intelligence judgment.

For the other judgments, the accuracy of subject reports was literally nil. Subject reports were correlated −.31 with true effects on the liking judgment, .14 with true effects on the sympathy judgment, and .11 with true effects on the flexibility judgment. Once again, observers were neither more nor less accurate than subjects. Correlations of their predictions with true effects were highly similar to the correlations of subject reports with true effects.

It should be noted that the experiment provides good justification for requiring a change in the traditional empirical definition of awareness. "Awareness" has been equated with "correct verbal report." The Nisbett and Bellows experiment and the present analysis strongly suggest that this definition is misleading and overgenerous. The criterion for "awareness" should be instead "verbal report which exceeds in accuracy that obtained from observers provided with a general description of the stimulus and response in question." Even highly accurate reports, therefore, provide no evidence of introspective awareness of the effects of the stimuli on responses if observers can equal that level of accuracy.

Accuracy and Inaccuracy in Verbal Explanations

When Will We Be Wrong In Our Verbal Reports?

It is possible to speculate further about the circumstances that should promote accuracy in reports about higher mental processes and those that should impair accuracy. We will need to call on another Tversky and Kahneman (1973) concept to help describe these circumstances. These writers proposed that a chief determinant of judgments about the frequency and probability of events is the *availability* in memory of the events at the time of judgment. Events are judged as frequent in proportion to their availability, and their availability is determined by such factors as the salience of the events at the time they were encountered, the strength of the network of verbal associations that spontaneously call the events to mind, and instructional manipulations designed to make the events more salient at the time of judgment.

The representativeness and availability heuristics are undoubtedly intertwined in the appraisal of cause and effect relations. If a particular stimulus is not available, then it will not be adduced in explanation of a given effect, even though it might be highly representative or plausible once called to mind. Similarly, the representativeness heuristic may be a chief determinant of availability in cause–effect analysis: A particular stimulus may be available chiefly because it is a highly representative cause of the effect to be explained.

It is possible to describe many circumstances that would serve to reduce the availability of a given causal candidate that is in fact influential, or to enhance the availability of a causal candidate that is in fact noninfluential. Similarly, influential causes will sometimes be nonrepresentative of the effects they produce, and noninfluential factors nevertheless will be highly representative causes. Any of these circumstances should promote error in verbal reports.

Removal in time. Perhaps chief among the circumstances that should decrease accuracy in self-report is a separation in time between the report and the actual occurrence of the process. In almost all the research described above, subjects were asked about a cognitive process immediately after its occurrence, often within seconds of its occurrence. While the present viewpoint holds that there may be no direct access to process even under these circumstances, it is at least the case that subjects are often cognizant of the existence of the effective stimuli at this point. Thus subjects have some chance of accurately reporting that a particular stimulus was influential if it happens to seem to be a plausible cause of the outcome. At some later point, the existence of the stimulus may be forgotten, or become less available, and thus there would be little chance that it could be

correctly identified as influential. Similarly, the vagaries of memory may allow the invention of factors presumed to be present at the time the process occurred. It is likely that such invented factors would be generated by use of causal theories. Thus it would be expected that the more removed in time the report is from the process, the more stereotypical should be the reported explanation.

Mechanics of judgment.[1] There is a class of influential factors to which we should be particularly blind. That class may be described as the mechanics of judgment factors—for example, serial order effects, position effects, contrast effects, and many types of anchoring effects. Such factors should seem particularly implausible as reasons for liking or disliking an object, or for estimating its magnitude on some dimension as high or low. Indeed, it seems outrageous that such a judgment as one concerning the quality of a nightgown might be affected by its position in a series, or that the estimation of the size of an object should be affected by the size of a similar object examined just previously.

Context. Generally, it should be the case that we will be blind to contextual factors, or at any rate be particularly poor at disentangling the effects of the stimulus from the context in which it was encountered. Contextual cues are not likely to be spontaneously salient when we are asked, or ask ourselves, why we evaluated an object as we did. Any question about an object is likely to focus our attention on the properties of the object itself and to cause us to ignore contextual cues. When a question about context is asked directly, on the other hand, as when we questioned our subjects about the effects of noise on their reactions to a film, contextual factors might well be reported as influential even when they are not. Unlike mechanics of judgment factors, many context factors, once they are made available, should seem highly plausible causes.

Nonevents. Ross (in press) has pointed out that many judgments and evaluations probably are based at least in part on the nonoccurrence of certain events. Thus one person may correctly perceive that another person does not like him, and this perception may be based largely on the nonoccurrence of friendly behaviors rather than on the outright manifestation of hostility. There is good reason to suspect that nonbehavior will be generally less available and salient than behavior, and therefore should rarely be reported as influential. But the effect will still require explanation, and thus noninfluential events will often be invoked in preference to influential nonevents.

Nonverbal behavior. In evaluating other people, we probably rely heavily on nonverbal cues such as posture, distance, gaze, and the volume and tone of voice (Argyle, Solter, Nicholson, Williams, & Burgess, 1970). Yet it seems likely that such nonverbal cues would be less available than verbal behavior, if only because verbal labels for nonverbal behavior are few and impoverished. To the extent that we rely on verbal memory to explain our evaluations of other people there will be proportionately more verbal behaviors to serve as causal candidates than nonverbal behaviors. To the extent that nonverbal behaviors are important to evaluations, relative to verbal behaviors, they will be wrongly overlooked.

Discrepancy between the magnitudes of cause and effect. In general, we would expect that factors will be perceived as causal to the degree that their magnitudes resemble the magnitude of the effects they are adduced to explain. In the development of causal schemata, both the notion that large causes can produce large effects and the notion that small causes can produce small effects probably precede the development of the notion that large causes can produce small effects. The notion that small causes can produce large effects probably develops very late and never attains very great stability. It is likely that conspiracy theories often feed on the discrepancy between officially provided causal explanations and the large effects they are invoked to explain. It is outrageous that a single, pathetic, weak figure like Lee Harvey Oswald should alter world history. When confronted with large effects, it is to comparably large causes that we turn for explanations. Thus when Storms and Nisbett (1970) interviewed insomniacs and asked them why they slept so little, both on particular occasions and in general, they were inclined to explain their insomnia in terms of the stress of

[1] We are indebted to Amos Tversky for this idea.

their current life situation or even in terms of neurosis or chronic anxiety. Smaller causes, such as an overheated room, a tendency to work or exercise or smoke just before going to bed, or a tendency to keep irregular hours, were overlooked. Many judgments of the plausibility of cause and effect relations are probably based at least in part on the fittingness of cause and effect magnitudes. Thus both mechanics of judgment factors and nonevents should often be perceived as implausible causes simply because of their smallness and seeming inconsequentiality.

When Will We Be Correct In Our Verbal Reports?

The present analysis corresponds to common sense in that it allows that we will often be right about the causes of our judgments and behavior. If a stranger walks up to a person, strikes him, and walks away, and the person is later asked if he likes the stranger, he will reply that he does not and will accurately report the reason. The interaction he has had with the stranger will be highly salient and a highly plausible reason for disliking someone. And, in general, the conditions that promote accuracy in verbal report will be the opposite of those described previously. These conditions may be summarized briefly by saying that reports will be accurate when influential stimuli are (a) available and (b) plausible causes of the response, and when (c) few or no plausible but noninfluential factors are available.

There is, in fact, some evidence in the literature that people can sometimes accurately report on the stimuli that influenced particular cognitive processes. Ironically, both of the areas where this had been systematically demonstrated have been developed by investigators who were seeking to show lack of awareness for the cognitive processes concerned. These are the literatures on (a) learning without awareness and (b) awareness of factors influencing complex judgments.

Learning Without(?) Awareness

Most of the literature concerning people's ability to report on the factors that influence learning, or the increased emission of an operant, has focused on the Greenspoon (1955) or related phenomena. In this paradigm, subjects say the names of words that come to mind, or generate sentences employing particular words. After a baseline, no-reinforcement period, subjects are systematically reinforced (by "uh huhs" or "goods") for a particular class of responses (e.g., plural nouns, or sentences employing first-person pronouns). The reinforcement typically elicits an increased rate of response for the reinforced class. Early investigators reported that subjects were unaware of this influence on their behavior. Later investigators (see e.g., Dulany, 1962; Erikson, 1962; Spielberger, 1962) insisted that subjects had been inadequately questioned, and that extensive probing revealed that all subjects who showed learning were also aware of the experimenter's reinforcements and the link between these reinforcements and their own increased output of reinforced responses.

Many writers have proposed that the subjects' "awareness" is due to nothing more than a Heisenberg-type effect. That is, the measurement procedure itself may suggest to the subject a connection that was not apparent to him before. Be that as it may, the present analysis makes it clear that there is every reason to expect that subjects in these experiments *should* be able to accurately report about cause and effect. (a) The response possibilities allowed the subject are extremely constrained. He is permitted very little latitude in the sorts of behavior he may emit. (b) The stimulus situation is even more fixed and static. In fact, virtually the only stimulus that occurs is the experimenter's "uh huh" or "good." (c) Finally, the causal connection between this critical stimulus or reinforcement and the increased frequency of a particular response class should be a highly plausible one.

It is thus hardly surprising that subjects report, or at least can be induced to report, a connection between the experimenter's stimulus and their own responses. Devotees of learning-without-awareness could scarcely have designed a paradigm more likely to result in accurate verbal report if they had set out deliberately to do so. There is some evidence, in fact, that when even relatively minor steps are taken to disguise the connection between stimulus and response, subjects will fail to report

such a connection. This comes from the so-called "double agent" study by Rosenfeld and Baer (1969) in which the subject believed himself to be the experimenter and the response (for which he was reinforced by the "subject" confederate) was not so focal as for the traditional Greenspoon subject. Under these circumstances, subjects reported no awareness, even under extensive probing, of the connection between the confederate's behavior and their own.

The Correspondence Between Actual and Subjective Weights in Judgment Tasks

Slovic and Lichtenstein (1971) have reviewed the literature concerning the ability of subjects to report accurately on the weights they assign to various stimulus factors in making evaluations. Most of the investigations of this question have employed either clinical psychologists or stockbrokers as subjects, and the judgmental domain has been largely limited to clinical diagnoses and assessments of the financial soundness of stocks. Subjects are asked to diagnose patients using Minnesota Multiphasic Personality Inventory (MMPI) scores or to assess stocks using such indicators as growth potential and earnings ratio. Then subjects are asked to state the degree of their reliance on various factors. These subjective weights are then compared to the objective weights derived from regression of the subject's judgments on the various factors. Slovic and Lichtenstein (1971) concluded that self-insight was poor and that of the studies which allowed for a comparison of perceived and actual cue utilization, "all found serious discrepancies between subjective and objective relative weights" (p. 684). While this is a fair assessment of this literature, what strikes one as impressive from the present vantage point is that almost all the studies reviewed by Slovic and Lichtenstein (1971) found evidence of at least some correspondence between subjective and objective weights. This is almost the sole evidence we have been able to uncover, outside the learning-without-awareness literature, that people can be at all accurate in reporting about the effects of stimuli on their responses.

The present framework is useful in understanding this lonely outcropping of accuracy. Clinical psychologists and stockbrokers undertake a formal study of the decision processes they should employ. They are taught explicitly how various factors should be weighed in their evaluations. Thus, for example, elevation of the schizophrenia scale will seem to be a highly plausible reason for a judgment of severe pathology because this is an association that clinicians are formally taught. It seems likely, in fact, that clinicians and stockbrokers could assign accurate weights *prior* to making the series of judgments in these experiments simply by calling on the stored rules about what such judgments should reflect. If so, one would scarcely want to say they were engaging in prospective introspection, but merely that they remember well the formal rules of diagnosis or financial counseling they were taught.

And in general, we may say that people will be accurate in reports about the causes of their behavior and evaluations wherever the culture, or a subculture, specifies clearly what stimuli should produce which responses, and especially where there is continuing feedback from the culture or subculture concerning the extent to which the individual is following the prescribed rules for input and output. Thus university admissions officials will be reasonably accurate about the weights they assign to various types of information in admissions folders, and auto mechanics will be reasonably accurate about the weights they assign to various factors in deciding whether a car has ignition or carburetor troubles. But such accuracy cannot be regarded as evidence of direct access to processes of evaluation. It is evidence for nothing more than the ability to describe the formal rules of evaluation.

The implication of this analysis is that the judgment studies lack what might be called "causal theory controls." Subjects' reported weights should not be compared directly to their actual weights. Instead, investigators should examine the *increment* in prediction of actual weights that is obtained by asking subjects about their subjective weights over the prediction that is obtained by (a) asking subjects about their subjective weights *prior* to their examination of the data set in question; or (b) asking subjects about their beliefs concerning the weights employed by the average, or ideal, or some particular, clinician or stock-

broker; or even (c) by simply asking subjects about the weights they were *taught* to employ in such judgments.

In the present view, little or no "awareness" might be found in studies employing such controls. That is, the subject's actual weights might be as well predicted by his subjective weights for his neighboring stockbroker as by the subject's reports about the weights he himself used.

Why are We Unaware of Our Unawareness?

There is of course a problem with any characterization of "introspection" as nothing more than judgments of plausibility. It does not feel like that at all. While we may sometimes admit to confusion about how we solved a particular problem or why we evaluated a person in a given way, it is often the case that we feel as though we have direct access to cognitive processes. We could retreat behind our data and assert that there is by now enough evidence discrediting introspective reports to allow us to ignore any argument based on introspection. But there is more that can be said than this.

It seems likely that there are regularities concerning the conditions that give rise to introspective certainty about cognitive processes. Confidence should be high when the causal candidates are (a) few in number, (b) perceptually or memorially salient, (c) highly plausible causes of the given outcome (especially where the basis of plausibility is an explicit cultural rule), and (d) where the causes have been observed to be associated with the outcome in the past. In fact, we appeal to introspection to support this view. Does the reader feel there is anything beyond factors such as these that need be adduced to account for occasions of subjective certainty?

The above view, it should be noted, is eminently testable. It should be possible to show that subjective certainty is great when causal candidates are salient and highly plausible, but are in reality noninfluential. Subjective certainty should be lower when the causal candidates are actually influential, but are not salient, not plausible, or compete with more salient or plausible but noninfluential causal candidates.

There are several factors that may help to sustain the illusion of introspective awareness. These are sketched below.

Confusion Between Content and Process

An important source of our belief in introspective awareness is undoubtedly related to the fact that we do indeed have direct access to a great storehouse of private knowledge. Jones and Nisbett (1972) have enumerated a list of types of privately held knowledge that bears repeating in the present context. The individual knows a host of personal historical facts; he knows the focus of his attention at any given point in 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 and much additional "knowledge" that is at least superior to that of any observer, 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.

A related point is that we are often capable of describing intermediate results of a series of mental operations in such a way as to promote the feeling that we are describing the operations themselves. Thus, for example, it is undoubtedly true that Maier's (1931) psychology professor subject had "imagery of monkeys swinging from trees." It is even conceivable that that imagery preceded and even facilitated the final steps in the mental operations that resulted in the cord-swinging solution. But is it scarcely reasonable to propose that such imagery *was* the process by which the problem was solved. A second example of the confusion of intermediate output with process was provided by an acquaintance of the authors' who was asked to introspect about the process by which he had just retrieved from memory his mother's maiden name. "I know just what the process was," he said. "I first thought of my uncle's last name, and since that happens to be my

mother's maiden name, I had the solution." This only pushes the process question back a step further, of course, and our acquaintance's answer would appear to reflect a confusion of intermediate results with the process by which the final result was obtained.

It should be noted that the individual's private access to content will sometimes allow him to be more accurate in his reports about the causes of his behavior than an observer would be. The occasions when the individual is correct or at least not provably wrong, and the observer is manifestly wrong, should serve to sustain the individual's sense of privileged access to process. Even these instances of superiority to observers, we would argue, are based not on access to process but access to content. It is possible to describe three kinds of content access that will allow for the individual's superior accuracy on occasion.

Knowledge of prior idiosyncratic reactions to a stimulus category. We have argued that perceived covariation between stimuli and responses is determined more by causal theories than by actual covariation. There are probably some cases, however, where individuals have idiosyncratic reactions to a particular stimulus that only they have knowledge of. For example, a person may believe that he generally loathes strangers who slap him on the back, and this belief may make him superior to observers in explaining his feelings in such a situation. We would suggest that such cases may be rare, however, and that the vast majority of perceived covariations between stimuli and responses may be determined by causal theories shared by both actors and observers.

Differences in causal theories between subcultures. It should be obvious that the individual's reports will be superior to those of observers when the observer is from a subculture that holds different causal theories. Thus, after attending a party at which a lively, high decibel band was playing, an 18-year-old would probably accurately say that the music increased his liking of the party. If a 40-year-old were asked to predict how much he would enjoy such a party and why, he would be more apt to say that the music would decrease his enjoyment. Thus, actors' reports about stimulus effects will differ from observers' predictions about their own reactions to that stimulus whenever the actors and observers belong to subcultures that have different causal theories about the effects of that stimulus. These reports will be similar whenever the actor and observer share causal theories about the effects of the stimulus in question, or whenever observers are asked to predict the effects of the stimulus on a member of a subculture with which they are familiar. Thus, whereas the 40-year-old's predictions about the influence of the music on himself will not correspond to the self-reports of the 18-year-old, his predictions about the influence of the music on the 18-year-old would probably correspond well to the latter's own reports.

Attentional and intentional knowledge. An individual may know that he was or was not attending to a particular stimulus or that he was or was not pursuing a particular intention. An observer, lacking such private knowledge of content, might often be more prone to error in his assumptions about the causes of an individual's behavior than the individual himself. On the other hand, private access to such content may also serve to mislead the individual. Occasionally, noninfluential stimuli may be more vivid and available to the individual than to an outside observer, for example, and thus the observer might sometimes be more accurate by virtue of disregarding such salient but noninfluential stimuli.

Inadequate Feedback

It seems likely that another important reason for our belief in introspective awareness stems from lack of feedback. Disconfirmation of hypotheses about the workings of our minds is hard to come by. If an insomniac believes that he is unable to get to sleep because of the stress of his life situation, he will always be able to find evidence supporting the view that his life situation is currently stressful. Indeed, the insomnia should be proof enough of the stressfulness of his life situation! And should he, in the midst of the very most stressful episode of his life, get a good night's sleep, he scarcely need abandon his sensible hypothesis about the cause of his insomnia in general. He can simply infer that the unusual stress must have left him so exhausted that it conquered his insomnia momentarily.

Motivational Reasons

A final factor that may serve to sustain our belief in direct introspective awareness is motivational. It is naturally preferable, from the standpoint of prediction and subjective feelings of control, to believe that we have such access. It is frightening to believe that one has no more certain knowledge of the workings of one's own mind than would an outsider with intimate knowledge of one's history and of the stimuli present at the time the cognitive process occurred.

Reference Notes

1. Aronson, E. Personal communication, 1975.
2. Zimbardo, P. Personal communication, 1975.
3. Nisbett, R. E., & Bellows, N. *Accuracy and inaccuracy in verbal reports about influences on evaluations.* Unpublished manuscript, University of Michigan, 1976.

References

Abelson, R. P. Psychological implication. In R. P. Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

Argyle, M., Salter, U., Nicholson, H., Williams, M., & Burgess, P. The communication of inferior and superior attitudes by verbal and nonverbal signals. *British Journal of Social and Clinical Psychology*, 1970, *9*, 222–231.

Aronson, E., & Mills, J. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 1959, *59*, 177–181.

Bem, D. J. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 1967, *74*, 183–200.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Bem, D. J., & McConnell, H. K. Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, 1970, *14*, 23–31.

Berkowitz, L., & Turner, C. Perceived anger level, instigating agent, and aggression. In H. London & R. E. Nisbett (Eds.), *Thought and feeling: Cognitive alteration of feeling states.* Chicago: Aldine-Atherton, 1974.

Brehm, J. W. Modification of hunger by cognitive dissonance. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Brehm, M. L., Back, K. W., & Bogdonoff, M. D. A physiological effect of cognitive dissonance under food deprivation and stress. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Broadbent, D. E. *Perception and communication.* London, Pergamon Press, 1958.

Brock, T. C., & Grant, L. D. Dissonance, awareness, and thirst motivation. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Brown, R. Models of attitude change. In R. Brown, E. Galanter, E. H. Hess, & G. Mandler (Eds.), *New directions in psychology* (Vol. 1). New York: Holt, Rinehart & Winston, 1962.

Burt, C. L. *The young delinquent* (4th ed.). London: University of Toronto Press, 1925.

Chapman, L. J. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 151–155.

Chapman, L. J., & Chapman, J. P. Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology*, 1967, *72*, 193–204.

Chapman, L. J., & Chapman, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, *74*, 271–280.

Cohen, A. R., & Zimbardo, P. G. Dissonance and the need to avoid failure. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Cottrell, N. B., & Wack, D. L. The energizing effect of cognitive dissonance on dominant and subordinate responses. *Journal of Personality and Social Psychology*, 1967, *6*, 132–138.

Darley, J. M., & Berscheid, E. Increased liking as a result of the anticipation of personal contact. *Human Relations*, 1967, *20*, 29–40.

Davis, J. A. *Great aspirations: The graduate school plans of America's college students.* Chicago: Aldine, 1964.

Davison, G. C., & Valins, S. Maintenance of self-attributed and drug-attributed behavior change. *Journal of Personality and Social Psychology*, 1969, *11*, 25–33.

Dixon, N. F. *Subliminal perception: The nature of a controversy.* London: McGraw-Hill, 1971.

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

Erdelyi, M. H. A new look at the new look: Perceptual defense and vigilance. *Psychological Review*, 1974, *81*, 1–25.

Eriksen, C. W. Figments, fantasies, and foibles: A search for the unconscious mind. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N.C.: Duke University Press, 1962.

Ferdinand, P. R. *The effect of forced compliance on recognition.* Unpublished master's thesis, Purdue University, 1964.

Festinger, L. *Cognitive dissonance.* Stanford, Calif.: Stanford University Press, 1957.

Freedman, J. L. Attitudinal effects of inadequate justification. *Journal of Personality*, 1963, *31*, 371–385.

Freedman, J. L. Long-term behavioral effects of cognitive dissonance. *Journal of Experimental Social Psychology*, 1965, *1*, 145–155.

Gaudet, H. A model for assessing changes in voting intention. In. P. F. Lazarsfeld & M. Rosenberg (Eds.),

The language of social research. New York: Free Press of Glencoe, 1955.

Ghiselin, B. *The creative process.* New York: Mentor, 1952.

Goethals, G. R., & Reckman, R. F. The perception of consistency in attitudes. *Journal of Experimental Social Psychology,* 1973, *9,* 491–501.

Goode, W. J. *After divorce.* New York: Free Press of Glencoe, 1956.

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

Grinker, J. Cognitive control of classical eyelid conditioning. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Gross, L. Modes of communication and the acquisition of symbolic competence. *Seventy-third Yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press, 1974.

Jones, E. E., & Nisbett, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones et al. (Eds), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Kadushin, C. Individual decisions to undertake psychotherapy. *Administrative Science Quarterly,* 1958, *3,* 379–411.

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review,* 1973, *80,* 237–251.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. Attribution in social interaction. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Kelman, H. Deception in social research. *Transaction,* 1966, *3,* 20–24.

Kornhauser, A., & Lazarsfeld, P. F. The analysis of consumer actions. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The language of social research.* Glencoe, Ill.: Free Press, 1955.

Kruglanski, A. W., Friedman, I., & Zeevi, G. The effects of extrinsic incentive on some qualitative aspects of task performance. *Journal of Personality,* 1971, *39,* 606–617.

Latané, B., & Darley, J. M. *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts, 1970.

Lazarsfeld, P. F. (Ed.), *Jugend und Beruf.* Jena, Germany: Fischer, 1931.

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.

Mandler, G. Consciousness: Respectable, useful and probably necessary. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, N.J.: Erlbaum, 1975. (a)

Mandler, G. *Mind and emotion.* New York: Wiley, 1975. (b)

Mansson, H. H. The relation of dissonance reduction to cognitive, perceptual, consummatory, and learning measures of thirst. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Miller, G. A. *Psychology: The science of mental life.* New York: Harper & Row, 1962.

Moray, N. *Attention: Selective processes in vision and hearing.* London: Hutchinson Educational, 1969.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Nisbett, R. E., & Schachter, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology,* 1966, *2,* 227–236.

Nisbett, R. E., & Valins, S. Perceiving the causes of one's own behavior. New York: General Learning Press, 1972.

Nisbett, R. E., & Wilson, T. D. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology,* in press.

Pallak, M. S. The effects of unexpected shock and relevant or irrelevant dissonance on incidental retention. *Journal of Personality and Social Psychology,* 1970, *14,* 271.

Pallak, M. S., Brock, T. C., & Kiesler, C. A. Dissonance arousal and task performance in an incidental verbal learning paradigm. *Journal of Personality and Social Psychology,* 1967, *7,* 11–21.

Pallak, M. S., & Pittman, T. S. General motivational effects of dissonance arousal. *Journal of Personality and Social Psychology,* 1972, *21,* 349–358.

Polanyi, M. *Personal knowledge: Toward a post-critical philosophy.* New York: Harper, 1964.

Rosenfeld, H. M., & Baer, D. M. Unnoticed verbal conditioning of an aware experimenter by a more aware subject: The double-agent effect. *Psychological Review,* 1969, *76,* 425–432.

Ross, L. D. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, in press.

Ross, L., Rodin, J., & Zimbardo, P. G. Toward an attribution therapy: The reduction of fear through induced cognitive-emotional misattribution. *Journal of Personality and Social Psychology,* 1969, *12,* 279–288.

Rossi, P. H. *Why families move: A study in the social psychology of urban residential mobility.* New York: Free Press of Glencoe, 1955.

Schachter, S., & Singer, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, *69,* 379–399.

Schachter, S., & Wheeler, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology,* 1962, *65,* 121–128.

Schlachet, P. J. The motivation to succeed and the memory for failure. In P. G. Zimbardo, *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Sills, D. L. *The volunteers: Means and ends in a national organization.* Glencoe, Ill.: Free Press, 1957.

Sills, D. L. On the art of asking "Why not?" Some problems and procedures in studying acceptance of family planning. In All India Conference on Family Planning (Fourth, 1961) *Report of the Proceedings: 29th January-3rd February 1961, Hyderabad.* Bombay: Family Planning Association of India, 1961.

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, *6*, 649–744.

Snyder, M., Schulz, R., & Jones, E. E. Expectancy and apparent duration as determinants of fatigue. *Journal of Personality and Social Psychology*, 1974, *29*, 426–434.

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.

Storms, M. D., & Nisbett, R. E. Insomnia and the attribution process. *Journal of Personality and Social Psychology*, 1970, *2*, 319–328.

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, *76*, 282–299.

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

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *184*, 1124–1131.

Valins, S., & Ray, A. A. Effects of cognitive desensitization on avoidance behavior. *Journal of Personality and Social Psychology*, 1967, *1*, 345–350.

Waterman, C. K. The facilitating and interfering effects of cognitive dissonance on simple and complex paired-associate learning tasks. *Journal of Experimental Social Psychology*, 1969, *5*, 31–42.

Weick, K. E. Reduction of cognitive dissonance through task enhancement and effort expenditure. *Journal of Abnormal and Social Psychology*, 1964, *68*, 533–539.

Weick, K. E. Task acceptance dilemmas: A site for research on cognition. In S. Feldman (Ed.), *Cognitive consistencey*. New York: Academic Press, 1966.

Weick, K. E., & Penner, D. D. Discrepant membership as an occasion for effective cooperation. *Sociometry*, 1969, *32*, 413–424.

Weick, K. E., & Prestholdt, P. Realignment of discrepant reinforcement value. *Journal of Personality and Social Psychology*, 1968, *8*, 180–187.

Wilson, W. R. *Unobtrusive induction of positive attitudes*. Unpublished doctoral dissertation, University of Michigan, 1975.

Zajonc, R. B. The attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 1968, *8* (2, Pt. 2).

Zimbardo, P. G., Cohen, A., Weisenberg, M., Dworkin, L., & Firestone, I. The control of experimental pain. In P. G. Zimbardo, *The cognitive control of motivation*. Glenview, Ill.: Scott, Foresman, 1969.

Zimbardo, P. G., Weisenberg, M., Firestone, I., & Levy, B. Changing appetites for eating fried grasshoppers with cognitive dissonance. In P. G. Zimbardo, *The cognitive control of motivation*. Glenview, Ill.: Scott, Foresman, 1969.

Received December 6, 1976 ■