

JOURNAL OF  
Personality  
AND  
Social Psychology

Copyright © 1977 by the American Psychological Association, Inc.

Volume 35

September 1977

Number 9

Verbal Reports About Causal Influences on Social Judgments:  
Private Access Versus Public Theories

Richard E. Nisbett and Nancy Bellows  
University of Michigan

Female subjects were asked to make four judgments about a young woman after reading her "job application portfolio." Five characteristics of the young woman were manipulated orthogonally (e.g., physical appearance, academic credentials). Subjects were asked to report how each of the five manipulated factors had influenced each of their judgments. "Observer subjects," who had access only to very impoverished descriptions of each of the five factors, were asked to predict how each of the factors would influence each of the judgments. Results showed that (a) subject reports about the effects of the factors on the judgments were in general highly inaccurate; (b) observer predictions were extremely similar to subject reports; (c) for the single judgment for which subjects showed substantial accuracy (a judgment about intelligence), observer predictions were as accurate as subject reports. The results indicate that whatever introspective access to cognitive processes may exist, it is not sufficient to produce generally accurate reports about such processes, nor even to produce reports that differ much from those of poorly informed outside observers. Reports about cognitive processes may be based less on introspection than on a priori causal theories about the effects of particular stimuli on particular responses.

Nisbett and Wilson (1977) have argued recently that people may have little or no direct introspective access to complex mental processes such as those involved in judgment,

decision making, and the initiation of social behavior. Evidence for this proposition comes from an examination of the literature on dissonance, attribution, and problem-solving research, and from several demonstration studies by Nisbett and Wilson. Subjects' ability to report accurately on the influence of a particular stimulus on their inferences or inference-produced behavior appears to be very poor. Sometimes subjects are unable to report the existence of the critical stimulus, and sometimes they are unable to report the existence of the response. Even when they are thoroughly cognizant of the existence of

---

The research reported here was supported by Grant BNS 75-23191 from the National Science Foundation. The authors are indebted to Timothy D. Wilson for advice, criticism, and assistance; to Keith Smith for statistical help; and to Sharon Gurwitz, Ronald Lemley, Hazel Markus, and Lee Ross for comments on the manuscript.

Requests for reprints should be sent to Richard E. Nisbett, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

both stimulus and response, they often deny that a genuinely influential stimulus affected their response and assert that a noninfluential stimulus did affect their response. For example, Nisbett and Wilson found that subjects wrongly denied a serial-position effect on their evaluation of consumer goods and wrongly asserted that their evaluation of a film had been affected by machinery noise outside the viewing room. Thus, whatever introspective capacity exists, it is in many cases not sufficient to produce accurate reports about the effects of particular stimuli on particular responses.

Nisbett and Wilson suggested that when people are asked to report on their cognitive processes, they do not interrogate directly a memory of the process involved, perhaps because they are unable to do so. Instead, reports about the effects of a stimulus on a particular response are based on a priori theories about the causal links between the stimulus and the response. Such a priori theories have a variety of origins:

1. There may be a culturally prescribed, *explicit* rule linking stimulus and response. ("I came to a stop because the light started to change.")

2. There may be an *implicit* causal theory supplied by the culture or a subculture. Such theories may be seldom articulated but may be shared tacitly by groups or entire cultures. As Abelson (1968) put it, a particular stimulus may "psychologically imply" a particular response. ("Jim gave flowers to me. That's why I've been acting pleased as punch today.")

3. There may be strong verbal associative links between the stimulus and the response. The stronger the connotative network linking stimulus description and response description, the more likely it is that a causal theory linking the two can be generated. Thus, if a response is labeled as "angry," stimuli labeled as "aggravating" are likely to be adduced as explanations of the response.

Evidence indicating that people rely on a priori theories comes from studies employing "observer subjects"—subjects who are not placed in the original experimental situations but who receive mere verbal descriptions of them and are asked to predict how particular

stimuli would affect particular responses. Predictions made by such observers are in general highly similar to the reports provided by subjects who actually experienced the situations (Nisbett & Wilson, 1977). As Bem (1967) argued in a somewhat different context, if the reports of observers are identical to those of the subjects themselves, then it is unnecessary to assume that the latter are drawing on "a fount of privileged knowledge" (p. 186). Instead, it seems likely that both subjects and observers are drawing on the same source for their reports—namely their a priori causal theories.

One implication of the Nisbett and Wilson analysis is that verbal reports about mental processes should sometimes be correct, even in the absence of any introspective ability. Verbal reports should be correct wherever there exists a correct causal theory, correctly applied in the particular instance. The psychological literature contains evidence that fairly good accuracy of verbal reports in fact is found in at least one domain—the research on complex judgments reviewed by Slovic and Lichtenstein (1971). Typically, in such research, half a dozen stimulus factors are manipulated orthogonally, and subjects are asked to make a judgment about each of the resulting stimulus configurations. For example, financial indices describing corporations are manipulated, and stockbroker subjects are asked to make a judgment about the wisdom of investing in the corporation. Or Minnesota Multiphasic Personality Inventory scale configurations are manipulated, and clinical-psychologist subjects are asked to make a judgment about the degree of pathology represented by the configurations. Then subjects are asked to report about the extent to which they relied on each of the stimulus factors in making their judgments. These reports are compared to the true reliance on each of the stimulus factors as calculated from the regression equation describing the subject's actual weighting of the stimulus factors. Though Slovic and Lichtenstein (1971) concluded that all studies of this type find "serious discrepancies between subjective and objective relative weights" (p. 684), it is the case that almost all of the studies they reviewed showed at

least some degree of correspondence between subjective and objective weights. Correlations between reported weights and actual weights vary from .3 to .8.

The traditional assumption would be that such correspondence between reported and actual reliance on particular factors should be taken as evidence of at least partially effective introspection. From the standpoint of the present analysis, however, such a concession would be premature. Even a relatively high degree of accuracy might be found in the absence of any true introspective ability. Stockbrokers and clinicians are trained formally to make their judgments, they have had repeated experience in making them, and they have received feedback about the extent to which they are following rules that are stated in highly explicit form. Thus their accuracy need not be construed as *prima facie* evidence of introspective ability. The accuracy may reflect nothing more than an ability to follow, and to describe verbally, the formal rules they were taught.

This analysis is readily testable. If the accurate reports in the complex judgment studies are due to nothing more than an ability to follow the rules and accurately describe them, then well-trained observers should be able to match the accuracy of the subjects. Stockbroker A's subjective weights might match Stockbroker B's objective weights as well as Stockbroker B's own subjective weights. If so, then Stockbroker B's reports are based on nothing more privileged than access to the explicitly stated rules of a subculture.

If the present analysis is correct, it should be possible to show both accuracy and inaccuracy of subject reports in an experiment requiring subjects to make different types of judgments. Reports about influences on a particular judgment should be relatively accurate when there exist explicit rules for making the judgment and when there is substantial opportunity for feedback about the extent to which the rules are being followed. Reports should be relatively inaccurate when there exist few or no explicit rules and little opportunity for feedback. Reports should be especially inaccurate when there are powerful influences on the judgment that are not con-

tained in a priori causal theories or where there exist a priori theories that hold, incorrectly, that a particular factor exerts a powerful influence on the judgment.

The present analysis also leads to the expectation that observer predictions should be as accurate as subject reports. This expectation follows from the assumption that subjects gain little or nothing through private access to the operations of their own judgmental processes and base their reports exclusively on causal theories that are shared by observers.

In the present experiment, female college students were asked to read what they believed was the application folder of an employee at a local crisis-intervention center. Subjects made judgments about the employee's intelligence, likability, sympathy toward the feelings of others, and flexibility in solving problems. Five factors were manipulated orthogonally:

1. Half the subjects were told that the applicant was quite physically attractive, and half were told nothing about her appearance.
2. Half the subjects were told that the applicant had an excellent academic record, and half were told nothing about her academic credentials.
3. Half the subjects were told that the applicant spilled a cup of coffee over the interviewer's desk, and half were told nothing about such an incident.
4. Half the subjects were told that the applicant had been in a serious auto accident, the aftereffects of which still caused her considerable pain, and half were told nothing about an accident.
5. Half the subjects were told that they would later meet the applicant whose folder they had read, and half were told they would meet some other applicant.

It was anticipated that reports about influences on the intelligence judgment would be generally accurate. The culture provides many explicit, normative rules about what factors should and should not influence such a judgment, and people receive feedback about the extent to which they are correctly following such rules (e.g., a judgment that a *summa cum laude* graduate of a prestigious university was unintelligent would be met by

incredulity). In addition, one of the manipulated factors (academic credentials) was chosen because it was expected to be both a highly potent and a highly plausible influence on the intelligence judgment, and the other factors were chosen because they were expected to be both weak and implausible influences on the judgment.

Accuracy was expected to be poorer for the other, more subjective judgmental dimensions, because there are fewer explicitly articulated rules concerning what factors should or should not influence such judgments, and feedback about the "correctness" of such judgments is infrequent and carries fewer cultural sanctions (e.g., one can say that one likes or dislikes a particular individual for a host of reasons and with little fear of contradiction). In addition, several of the manipulated factors were selected because previous research has shown that they have implausible, that is, culturally unacknowledged, effects on some of the judgments. Aronson, Willerman, and Floyd (1966) have shown that the effect of a pratfall such as the coffee-spilling incident is to increase the likability of a generally competent-seeming and superior person. Lerner and his colleagues (e.g., Lerner, 1971; Lerner & Matthews, 1967; Lerner & Simmons, 1966) have shown that the victims of unavoidable misfortunes are derogated. Darley and Berscheid (1967) have shown that the mere anticipation of personal contact with an individual is sufficient to result in higher ratings of the attractiveness of the individual's personality. Since these effects are probably not represented in people's implicit causal theories, it was anticipated that subjects would fail to report them and observers would fail to predict them.

Though there probably exist relatively few explicitly articulated theories about what sorts of factors influence such a judgment as "sympathy toward the feelings of others," implicit theories about such influences probably do exist or could be generated readily via Abelson's (1968) "implicational principles." Thus, most people would probably assume that their judgment about another person's sympathy was favorably affected by the knowledge that the person had been in a serious accident. The implicational principle

drawn upon here would be something like "suffering makes you sensitive to the sufferings of others." In contrast, most people would probably not assume that their judgment about a person's sympathy would be affected by knowledge that they would soon meet the person. Such an influence would violate even the loose and generous deductive principles of psychology. Thus, although accuracy of subject reports about the influences on the more subjective judgments was expected to be low, it was anticipated these reports would closely resemble the predictions made by observers, because subjects and observers share the same implicit a priori theories and these theories serve as the basis for both subject reports and observer predictions.

## Method

### *Subjects*

Subjects were 128 female volunteers enrolled in introductory psychology classes at the University of Michigan. Thirty-four female students drawn from the same population served as observer controls. Females were employed exclusively because of greater availability in the subject pool.

### *Rationale Provided to Subjects*

Upon arrival at the laboratory, subjects were taken to individual rooms and asked to read a cover sheet describing the purpose of the study. The cover sheet alleged that the (fictitious) Washtenaw County Crisis Center had contacted the investigators and asked for help in screening applicants for positions on the staff of the center. The center had experienced difficulty in selecting effective staff members, in that staff members often had turned out to be unreliable or ineffective in dealing with clients. The investigators had proposed that college students might be able to provide valid predictions about the likely effectiveness of candidates, since students are closer in age and status to both candidates and clients than the senior staff of the center involved in selection. Therefore, the subject was being asked to read the application folder of a recently hired employee (a young woman like the subject) and make several judgments about the employee. The subject's judgments would be compared to evaluations made of the employee's performance at a later point. Then, if college students' judgments turned out to be significant predictors of job success, such judgments would be included in the selection process for future applicants.

### *The Application Folder*

All subjects read three double spaced pages of information concerning the employee—Jill. This information was allegedly drawn from a background questionnaire, an interview with the center director, and a letter of recommendation written by the head of a camp where Jill had worked for two summers as a counselor. The information was intended to be interesting and vivid and to convey an impression of a fairly well-adjusted and competent young woman who was neither particularly likable nor unlikable.

Subjects read that Jill had been brought up in a small town in Michigan and that she had not travelled outside the state. Her father was an accountant and her mother managed a clothing store. Jill had an identical twin sister and a sister and two brothers, who were all older than she. Her social life in high school had centered around a tightly knit group of girls involved in athletics, and though she had dated sporadically she had never had a steady boyfriend. She had played the flute and belonged to a folk-singing club, and her hobbies included knitting, listening to music, and photography. Jill had never been involved with drugs, though her attitude toward drugs was not moralizing.

Jill had been fluent and responsive in the interview with the center director. She seemed to talk a lot with her hands and to avoid eye contact, especially when she made a statement that called attention to herself. The director had presented her with several hypothetical examples of clients coming to the center with problems. Her responses to these hypothetical cases were intended to make her seem sensible though not highly concerned or sympathetic. When asked why she wanted to work at the center, Jill had answered that she hoped to use the experience as the basis for a decision about whether to pursue a career in social service.

Jill was described in the letter of recommendation as having been an effective, energetic, and reliable counselor who got along well with the campers, but was regarded as somewhat aloof by some of the other counselors.

### *Manipulation of Stimulus Factors*

The material in the application folder contained four of the five manipulated variables:

1. *Attractiveness.* Subjects who received the attractiveness manipulation read, in the first paragraph, that Jill was of "average height and slender, had short brown hair, large brown eyes, and a smooth, clear complexion" and was "quite attractive." Other subjects were told nothing about Jill's appearance.

2. *Academic credentials.* Subjects who received the academic credentials manipulation were told that she had "graduated as salutatorian of her high school class" and that she was now "an honors student at the University of Michigan, majoring in

political science." Other subjects were told only that she had not yet decided to go to college, but that if she did her major might be political science.

3. *Coffee-spilling incident.* Subjects who received the coffee-spilling manipulation were told that "at one point during the interview she became so engrossed in conversation that she accidentally knocked a cup of coffee over some papers on the desk." For other subjects there was no mention of such an incident.

4. *Auto accident.* Subjects who received the auto accident manipulation were told that, when asked if she had ever personally experienced any "crisis" situation, Jill explained that during her senior year of high school she had accepted a ride home from a party with someone who had been drinking a lot. The car had gone out of control and hit a pole. She received several injuries including a crushed pelvis. She had already had several operations but had recurring pain which might necessitate further operations. Other subjects read that, though Jill had never experienced a crisis situation, she had sometimes been distressed enough to contact a good friend or family member for support.

5. *Anticipated personal contact* (the cover sheet contained the manipulation of this variable). Subjects who received the personal contact manipulation were told that after evaluating the employee "on the basis of the written information given to you, you will have an opportunity to meet that person. We will ask you simply to get acquainted with the person in a brief conversation and later to write down your impression of the person. In this way we will be able to assess whether or not impressions of a person based on written information are similar to impressions formed by people who actually meet that person." Other subjects were told that after "evaluating one person on the basis of written information given to you, you will have an opportunity to meet another person who is on the staff." A similar rationale was provided for this meeting. The manipulation was reinforced and made more vivid by the experimenter, who explained a little more about when and where the conversation would take place and what some useful topics might be and repeated that the person the subject would meet was either the same person described in the folder or someone else entirely.

The orthogonal manipulation of the five factors resulted in a 32-condition design, with four subjects in each condition.

### *Judgments*

Immediately after reading the application folder, subjects were asked to make four judgments about Jill: "How much do you think you would like this person?" "How sympathetic would this person be to the feelings of others?" "How intelligent do you think the person is?" and "How flexible would the person be in solving problems?" Judgments were made on 17-point scales, with labels on alternate points (e.g., "like extremely," "like very much,"

“like,” “like a little,” “neither like nor dislike,” . . . “dislike extremely.”

*Reports About Effects of Stimulus Factors on Judgments*

In the next section of the questionnaire, subjects were told that the investigators wanted to know “specifically how certain factors influenced your judgments about the person. Each question below is broken into four parts, corresponding to each of the four judgments you made about the person. In each question, we name a specific factor and ask you how that factor influenced your judgments.” The format for these questions was as follows.

How did the person’s physical appearance influence:

a. how much you like the person?

+3	+2	+1	0
increased	increased	increased	did not
my liking	my liking	my liking	influence
a great		a little	my liking
deal			

-1	-2	-3
decreased	decreased	decreased
my liking	my liking	my liking
a little		a great
		deal

- b. how sympathetic the person seems? (followed by scale similar to that above)
- c. how intelligent the person seems?
- d. how flexible the person seems in solving problems?

All subjects were asked first about the effects of three nonmanipulated factors common to all conditions. Then each subject was asked about the effects of each of the manipulated factors she had been exposed to. This ranged from none, in one cell of the design, to five in another cell. The questions about the influence of the other factors were worded as follows:

- How did the person’s academic credentials influence . . .
- How did the coffee-spilling incident influence . . .
- How did the fact that the person had been in a serious accident influence . . .
- How did the fact that you are going to meet this person influence . . .

*Observer Subjects*

Thirty-four observer subjects were told that the investigators were interested in how people make judgments about others from particular kinds of information. Observers received none of the information in the three-page portfolio. Instead, observers were asked simply to “imagine that you have access to information about a person of your own age and sex that was obtained from a job interview. . . . We

would like to know how this information would affect four types of judgments that you might make about this person.”

Observers were then asked about the effects of each of the five factors on each of the four judgments. Factors were described in a very abbreviated and impoverished fashion:

Suppose you knew the person was quite physically attractive. How would that influence:

a. how much you would like the person?

+3	+2	+1	0
would	would	would	would not
increase	increase	increase	influence
my liking	my liking	my liking	my liking
a great		a little	
deal			

-1	-2	-3
would	would	would
decrease	decrease	decrease
my liking	my liking	my liking
a little		a great
		deal

- b. how sympathetic the person would seem? (followed by scale similar to that above)
- c. how intelligent the person would seem?
- d. how flexible the person would seem to be in solving problems?

The other four factors were described as follows: Suppose you knew . . .

- the person had a very good academic record.
- the person had spilled a cup of coffee during the interview.
- the person had been in a serious accident.
- that you were going to meet the person in the near future.

**Results**

There were three major expectations:

1. Subject reports should be inaccurate, especially for the relatively subjective judgments for which there exist few explicitly stated cultural rules for making the judgments and for which there is relatively little feedback about “accuracy.” Subject reports should be more accurate for the intelligence judgment, for which there do exist explicit cultural rules and feedback opportunities.
2. The accuracy of subject reports should not exceed the accuracy of observer predictions, because actors gain little or nothing from privileged access to their own judgmental processes.
3. Observer predictions should be highly similar to actor reports, because both actors

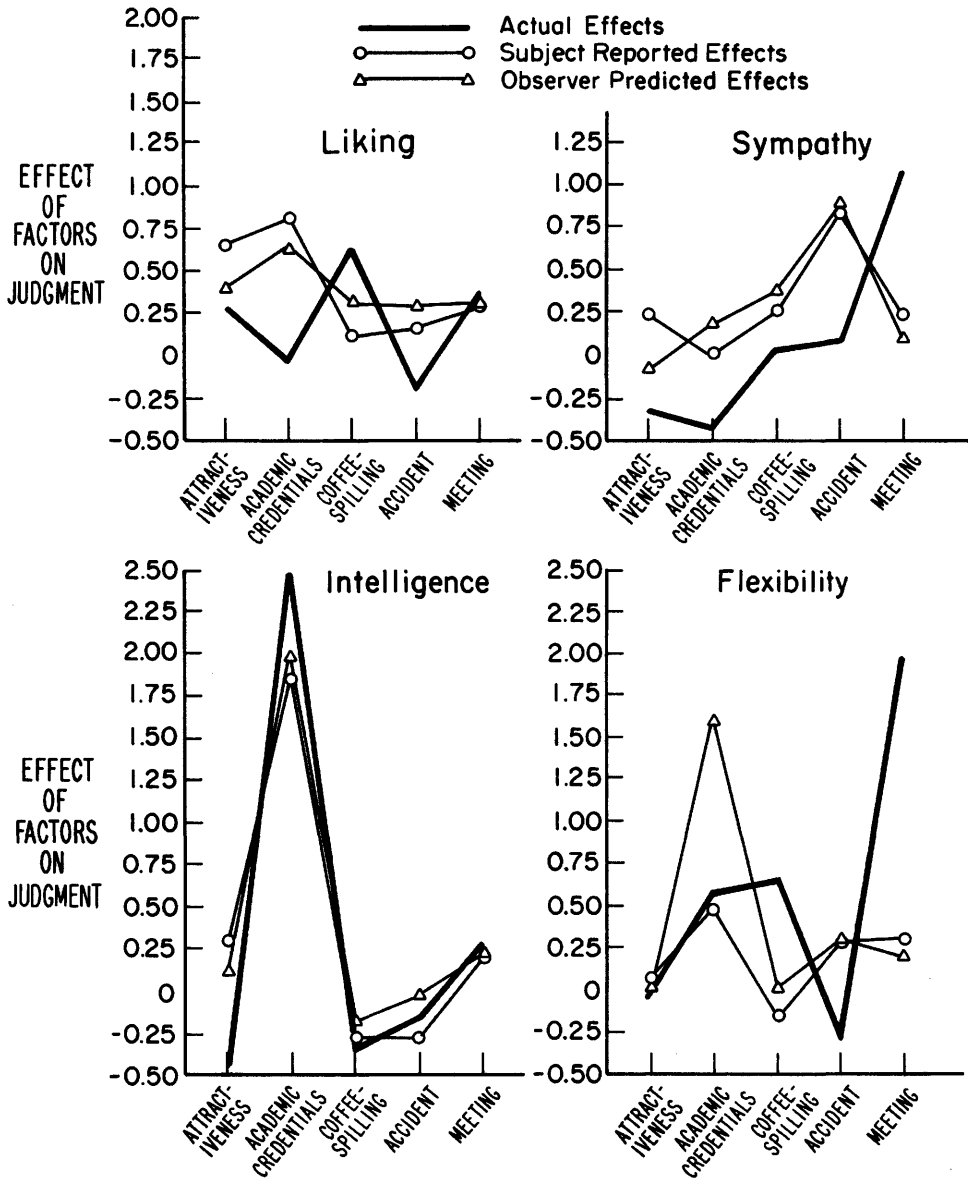


Figure 1. Effects of manipulated factors on judgments, reports by subjects about the effects of the factors, and predictions by observers about the effects of the factors.

and observers rely on the same a priori theories—explicit, normative theories in the case of the intelligence judgment, and implicit, psychological theories in the case of the more subjective judgments.

Figure 1 presents (a) the mean reported effects of the factors on each of the judgments (based on the answers of the 64 subjects who were exposed to each of the

factors); (b) the mean predicted effects (based on the answers of the 34 observers); and (c) the mean actual effects of the factors (based on the mean judgment for the 64 subjects exposed to a particular factor minus the mean judgment for the 64 subjects who were not). The first two measures are based on 7-point scale means. The last is based on the subtraction of two 17-point scale means.

Table 1

*Correlations Among Mean Actual Effects of Factors on Judgments, Mean Reports About the Effects of the Factors by Subjects, and Mean Predictions About the Effects of the Factors by Observers*

Correlation	Judgment			
	Liking	Sympathy	Intelligence	Flexibility
Actual effects with subject reports	-.31	.14	.94***	.11
Actual effects with observer predictions	-.32	-.02	.98***	-.02
Subject reports with observer predictions	.89**	.84*	.99***	.77

Note.  $n$  for each correlation = 5.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

It is therefore accidental, though convenient for purposes of presentation, that the former two values and the latter value could be plotted on the same numerical scale on the y-axis.

#### *Accuracy of Subject Reports About Effects of Factors on Judgments*

##### *Liking, Sympathy, and Flexibility Judgments*

It is visually obvious, in Figure 1, that subject reports about the effects of stimuli on the more subjective judgments bear little relation to the actual effects of the stimuli. In fact, subjects were wrong in 8 of 15 instances about the *directional* effect of the factors on these judgments. That is, subjects reported a positive influence when there was actually no net influence (one instance) or a negative influence (five instances), or they reported a negative influence when there was a positive influence (one instance), or they reported no net influence when there was a negative influence (one instance).

Only one of the actual main effects on each of the judgments was substantial. In all three cases, subjects failed to report these as relatively strong effects. For the liking judgment, the effect of the coffee-spilling incident was to marginally increase liking ( $.10 < p < .15$ ).<sup>1</sup> Subjects reported, however, that this was the *least* influential of the five factors. For the sympathy judgment, there was a significant effect of anticipated contact ( $p =$

.025), yet subjects reported that three other factors had as great or greater effect on this judgment. For the flexibility judgment, the effect of anticipated contact was very great ( $p < .001$ ), yet subjects reported that it was slight and reported that one other factor (academic credentials) was more influential.

Thus relatively large actual effects were not detected as being relatively large by subjects. The converse was also true—that is, effects that subjects reported to be relatively large in fact were not. For example, subjects reported that the factor having the largest effect on the liking judgment was the academic credentials factor ( $p$  contrasting mean reported effect of academic credentials factor with theoretical mean of zero  $< .001$ ). In fact, the effect of the academic credentials manipulation on the liking judgment was nil. Similarly, subjects reported that the factor having the greatest effect on the sympathy judgment was the accident factor ( $p < .001$ ). Again, the actual effect of the accident manipulation was nil. Finally, subjects reported that the factor having the greatest effect on the flexibility judgment was the academic credentials factor. In fact, two effects on that judgment were larger, and one of them was very much larger.

One way of quantitatively assessing the

<sup>1</sup> All  $p$  levels reported here are based on two-tailed tests.



accuracy of subject reports is simply to examine the correlation between the mean reported effects of the factors and the mean actual effects of the factors, as displayed in each of the four panels of Figure 1. These correlations are presented in Table 1, together with the correlations between observer-predicted effects and actual effects and the correlations between observer-predicted effects and subject-reported effects. Although correlation of means is a somewhat unconventional technique of data analysis, and the  $n$  of 5 for each judgment is low, the correlations provide an instructive way of describing what is visually apparent in Figure 1. It may be seen that the correlations between subject reports and actual effects are near zero for the liking, sympathy, and flexibility judgments. Thus, reports about the effects of the factors on these judgments bear literally no relationship to the actual effect of the factors.

#### *Intelligence Judgment*

Subject reports about the effects of the factors on the intelligence judgment stand in marked contrast to subject reports about the effects on other judgments. As anticipated, the effect of the academic credentials manipulation on the intelligence judgment was very pronounced ( $p < .001$ ), while none of the other factors had a significant effect on this judgment. Subject reports almost perfectly tracked these effects. The correlation between actual effects and reported effects was .94.

#### *Interaction Effects and Subject Accuracy*

There are two ways in which subjects might possibly have shown some accuracy on the more subjective judgments that could have gone undetected by the analyses presented in Figure 1 and Table 1. If the factors had interacted in their effects on some of the judgments, subjects might have been cognizant of these complicated patterns, but the simple main effect analysis reported above would obscure this potential sensitivity. Thus, for example, subjects who received the physical attractiveness manipulation might have liked the target more when they also received the academic credentials manipulation and

less when they did not receive the academic credentials manipulation, and subject reports might have reflected cognizance of this pattern. In fact, however, there were no more interactions than would be expected by chance. The  $2 \times 2 \times 2 \times 2 \times 2$  factorial design yielded 26 interactions for each of the 4 dependent variables. Only 4 of these 104 interactions were significant at the .05 level.

A second, somewhat more complicated possibility for accuracy also exists. Because of the factorial design of the experiment, subjects in some conditions were exposed to few of the manipulated factors and subjects in other conditions were exposed to many. If some of the factors had had an effect when they appeared in isolation or near isolation, but a weaker effect when they appeared in conjunction with several other factors, subjects who received few factors might have accurately reported a strong effect of a particular factor and those who received many factors might have accurately reported a weak effect or no effect of the same factor. The simple main effect analysis would have been insensitive to such differential reports. In order to examine this possibility, the effect of each of the factors was contrasted for conditions where it appeared with one or no other factor, with two other factors, and with three or four other factors. Two-way analyses of variance were performed with presence versus absence of the factor as one way of the design and the number of other factors present as the other way. Only 3 of the resulting 20 interactions were significant at the .05 level, and 1 of these was due to the greater effect of one of the factors when it appeared with more than when it appeared with fewer other factors. In none of the 3 instances did the pattern of subject reports accurately reflect the pattern of the actual interaction effects.

#### *Comparison of Subject and Observer Accuracy*

Subject accuracy did not exceed observer accuracy for any of the judgments. For the more subjective judgments, the accuracy of subject reports was nil and so was the accuracy of observer predictions. For the

intelligence judgment, subject accuracy was virtually perfect. This accuracy cannot be attributed to introspective capacity on the part of subjects, however, because the accuracy of observer predictions equalled the accuracy of subject reports. The correlation between observer predictions and actual effects (.98) was trivially higher than the correlation between subject reports and actual effects (.94). Thus subject reports were accurate if and only if observer predictions were also accurate.

#### *Similarity of Subject Reports and Observer Predictions*

The similarity of subject reports about the effects of the factors and observer predictions about their effects was so great as to indicate that they must have been generated by similar processes, that is, by reliance on similar a priori theories. The bottom row of Table 1 presents correlations between subject reports about the effects of the factors and observer predictions about their effects. The correlations range from a low of .77 to a high of .99. It is visually obvious, in Figure 1, that except for the effect of academic credentials on the flexibility judgment, subject reports and observer predictions are virtually identical. And in fact, with that single exception, none of the subject report means are significantly different from the observer prediction means. In 16 of 20 instances, moreover, the subject reports are closer to the observer predictions than to the actual effects. (This is true even when all three types of scores are standardized in order to remove any scaling differences.) The fraction 16/20 is significantly greater than would be expected even by the extremely conservative prior of 50%—which holds that subject reports would be expected to be no closer to the true effects than to observer predictions.

#### Discussion

The present results are consistent with Nisbett and Wilson's (1977) analysis of the basis of verbal reports about higher mental processes.

1. Subject reports about the influence on

their judgments were in general highly inaccurate, suggesting that whatever introspective capacity exists, it is not sufficient to produce universally accurate reports about the influences on social judgments.

2. Subject reports about the influences on one of the judgments—the intelligence judgment—were highly accurate. This judgment differs from the others in that there exist quite explicit rules about the factors that should influence such a judgment, and the culture provides frequent feedback to the individual about the extent to which he is following those rules.

3. Subjects were probably correct in their reports about the intelligence judgment not by virtue of any ability to introspect accurately about this judgment, but because their a priori causal theories about the influences on this judgment happened to be correct. Evidence for this comes from the fact that observers' predictions about the effects of the various factors on the intelligence judgment were fully as accurate as subjects' reports about the effects of the factors.

4. The similarity between observer predictions about the effects of the factors on their judgments and subject reports about the effects was so great as to suggest that introspection played little part in subject reports. Instead, such reports, whether accurate or inaccurate, appear to be based on a priori theories that are shared by observers and that serve to generate nearly identical predictions from observers.

#### *Were Subjects Aware of Their Unawareness?*

There is another, more informal result that should be noted. It might be asked whether subjects are aware of their unawareness, that is, did they answer the questions about the influences of the factors on their judgments in a spirit of uncertainty and confusion, or did they do so with confidence? In an attempt to learn about this, the experimenter probed subjects' questionnaire responses in a postexperimental interview, asking them whether they were confident about their answers to particular questions or were merely guessing about the nature of the influ-

ences. The best summary of these interviews is that most subjects answered some questions in a quite confident mood and others in a spirit of real uncertainty. When the experimenter noted to the subject that she had indicated that a particular factor had influenced a particular judgment and asked the subject how confident she was that this was the case, subjects were sometimes quite firm in their answers ("Yes, I'm sure that her academic record influenced my liking for her. Intelligence is very important to me in a person. All of the people I like best are very smart"). In other instances, however, subjects did appear to be aware of their unawareness ("I really don't know, I said it increased my liking, but that was just a guess, really"). The same pattern held for cases where subjects indicated that a particular factor had not influenced a particular judgment. Subjects sometimes allowed that they might well be wrong but sometimes were firm to the point of belligerence ("Of course, I couldn't have been affected by the fact that I was going to meet her. How could that possibly affect what I thought about her?"). Thus subjects sometimes answered questions about causal influences in a spirit of humility that, from the present viewpoint, would be appropriate. But they sometimes were quite insistent and confident about their reports. We would suggest that this confidence, where found, rested on their certainty about their a priori theories and not on their certainty about their memories of the judgment process.

#### *The Appropriate Operational Definition of Awareness*

One of the chief implications of the present study is that the traditional operational definition of *awareness* is misleading and should be modified substantially. Previously, *awareness* of the effect of a stimulus on a response has been defined as "ability to report accurately" on the effect of the stimulus on the response. But the present data suggest that, at least under some circumstances, observers lacking access to any but the most fragmentary information about the stimulus configuration can be as accurate in their reports about stimulus effects as subjects. Such

observers are not "introspecting" about the effects of a particular stimulus among a host of others, but are simply making a prediction based on their implicit theories. Thus the term *awareness* should be reserved for accuracy in excess of that achieved by observers having knowledge only of a verbal description of the stimulus in question."

Both the line of reasoning proposed here and the indicated methodology have analogies in Bem's (1967) "interpersonal replication" work. Bem argued that if dissonance-free observers estimate the same attitudes for subjects that dissonance-experiment subjects themselves report, then it is unnecessary to assume that subjects' attitudes are based on information any more privileged than that available to observers. We argue analogously that it is unnecessary to assume that subjects' reports about the influences on their attitudes and behavior are based on information any more privileged than that available to observers, if the accuracy of subjects' reports about such influences does not exceed the accuracy of observers' predictions.

#### *Subject Accuracy Versus Observer Accuracy*

We do not wish to imply that the accuracy of subjects' reports should never be expected to exceed that of observers. Without conceding that subjects have any ability to introspect about process, it is possible to admit that they may sometimes have privileged access to knowledge that might permit them to be more accurate in their reports about the effects of stimuli than observers. In fact, there is some reason to suspect that the present data include an example of just such an instance of subject superiority based on access to knowledge lacking to observers.

It will be recalled that observers predicted that their judgments about an individual's flexibility in solving problems would be greatly increased if they knew that the individual had a very good academic record. Subjects, however, reported that the candidate's academic record had had a relatively slight influence on their judgments about flexibility, and they were correct in this. We are inclined to suspect that this is an instance

of true subject superiority and not a chance occurrence. The question about flexibility immediately followed the question about intelligence. It seems likely that observers, just having said that a good academic record would make them think that a person was intelligent, then naturally went on to say it would also make them think that a person would be more flexible in solving problems. Subjects, however, had read about the candidate's responses to two hypothetical problems. Thus it might have seemed plausible to subjects that their judgments about the candidate's flexibility in solving problems were influenced more by her response to the hypothetical problems than by her academic record. Such a conclusion does not require the assumption that subjects had any access to the process by which they integrated all available information into a judgment about flexibility in solving problems. It merely assumes that they had knowledge of a potential causal influence that was lacking to observers and that their implicit theories about this factor happened to be correct in this instance.

If the present experiment is any guide, however, such instances of subject superiority over observers may turn out to be relatively rare. Access to privately held knowledge, moreover, is probably a two-edged sword. It would probably not be difficult to show that such access can sometimes lead individuals to give reports that are more in error than observer predictions. This might be the case if, for example, the existence of a noninfluential stimulus were known to subjects and not to observers. If the stimulus were highly salient and a highly plausible cause of a particular response, it might be regarded as more influential than genuinely influential stimuli whose existence was known to both subjects and observers. The genuinely influential stimuli might then be slighted in subject reports in favor of the salient,

plausible, but noninfluential stimulus.

Finally, it should be clear that the present findings raise serious methodological questions concerning the validity of subject reports as a tool of social science investigation. Nisbett and Wilson (1977) have given many examples of the use by social scientists of the essentially introspective technique of asking subjects to report causal influences on their behavior or evaluations. The present results suggest that such reports may have utility only for the study of verbal explanations and may have little or no value as a guide to the true influences on behavior.

### References

- Abelson, R. P. Psychological implication. In R. P. Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand-McNally, 1968.
- Aronson, E., Willerman, B., & Floyd, J. The effect of a pratfall on increasing personal attractiveness. *Psychonomic Science*, 1966, 4, 227-228.
- Bem, D. J. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 1967, 74, 183-200.
- Darley, J. M., & Berscheid, E. Increased liking as a result of the anticipation of personal contact. *Human Relations*, 1967, 20, 29-40.
- Lerner, M. J. Justice, guilt, and veridical perception. *Journal of Personality and Social Psychology*, 1971, 20, 127-135.
- Lerner, M. J., & Matthews, P. Reactions to suffering of others under conditions of indirect responsibility. *Journal of Personality and Social Psychology*, 1967, 5, 319-325.
- Lerner, M. J., & Simmons, C. Observer's reaction to the innocent victim: Comparison or rejection? *Journal of Personality and Social Psychology*, 1966, 4, 203-210.
- Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, 84, 231-259.
- 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.

Received March 31, 1977 ■