

The Accuracy of Verbal Reports About the Effects of Stimuli on Evaluations and Behavior



Timothy de Camp Wilson; Richard E. Nisbett

Social Psychology, Vol. 41, No. 2. (Jun., 1978), pp. 118-131.

Stable URL:

<http://links.jstor.org/sici?sici=0147-829X%28197806%2941%3A2%3C118%3ATAOVRA%3E2.0.CO%3B2-M>

Social Psychology is currently published by American Sociological Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/asa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

The Accuracy of Verbal Reports About the Effects of Stimuli on Evaluations and Behavior*

TIMOTHY DE CAMP WILSON
Duke University

RICHARD E. NISBETT
The University of Michigan

Four experiments are reported, each having the following design: (1) Components of a complex stimulus situation were manipulated; (2) subjects responded to the situation in some fashion—evaluational, judgmental, or behavioral; and (3) subjects were asked what effect, if any, the manipulated stimulus components had had upon their responses. The results indicate that people's reports are highly inaccurate. Stimulus components with substantial effects on responses were reported to have had no effect on responses, and stimulus components with little or no effect were reported to have had substantial effects. The results indicate that research which relies on people's introspective reports about the causes of their behavior may have little value as a guide to the true causal influences.

Nisbett and Wilson (1977) have argued recently that people have little or no introspective access to higher-order cognitive processes, such as those involved in judgments, decision making, and the initiation of social behavior. Stated in a less extreme yet more testable fashion, their hypothesis is that whatever capacity for direct introspection on cognitive processes may exist, it is insufficient to produce generally accurate verbal reports about the effects of stimuli on one's own behavior. Nisbett and Wilson (1977) reviewed in depth the evidence bearing on people's ability to report accurately about the effects of stimuli on their behavior, including research on dissonance and attribution processes, problem solving, and helping behavior.

Very few of the reviewed studies, however, were specifically designed to address the question of accuracy of verbal

reports about stimulus effects. Since social scientists often ask people to report the reasons for their judgments and choices—for example, why they like political candidates (Gaudet, 1955), why they failed to use a contraceptive technique (Sills, 1961), or why they sought out a psychoanalyst (Kadushin, 1958)—the validity of such verbal reports needs to be more directly ascertained. If subjects are found to be often incorrect in their assessments of what factors influenced their responses, research of this type would be called into question. In addition, some social psychologists have suggested that laboratory experiments involving deception be abandoned in favor of simply asking subjects how their cognitive processes would work if they were exposed to various stimulus situations (Brown, 1962; Kelman, 1966; cf. also a symposium in *Personality and Social Psychology Bulletin*, 1977). Research that investigates the ability of subjects to describe accurately their cognitive processes would also speak to this research practice.

The present article presents the detailed procedures and results of a number of studies reported in summary fashion by Nisbett and Wilson (1977). Each of these studies investigated people's ability to report accurately on the effects of stimuli on their judgmental or behavioral responses.

* This research was supported by Grant GS 40085 from the National Science Foundation. We thank Eugene Borgida for assistance and criticism in every stage of the research, Michael Kruger, Linda Temoshok, and Amos Tversky for many useful comments and suggestions, John McKenzie for assistance in data analysis, and Tim Van of Meijer's Thrifty Acres for aiding us in one of the studies. Address communications to Timothy D. Wilson, Department of Psychology, Duke University, Durham, NC 27706.

Some of the earlier studies showing inaccurate verbal reports were based on situations that were quite ego-involving, and self-esteem or social desirability motives could have accounted for the inaccuracy. To avoid arousing such motives, the present studies were designed to be as little ego-involving as possible. The studies examined routine cognitive processes, of a sort that occur frequently in daily life, and little deception was involved. A range of behavioral domains was examined, including evaluations, choices, and predictions.

The reader is entitled to know that the stimulus situations were not sampled in some random fashion; indeed it was impossible to do so because it is not possible even to describe the domain of "all stimulus situations." Instead, the 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 in which we felt that a particular stimulus would exert an influence on subjects' responses but that subjects would be unable to detect it, and situations in which we felt that 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 would influence subjects' responses turned out to have no effect, and many of the stimuli that we expected would 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 was 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.

FAILURES TO REPORT THE INFLUENCE OF EFFECTIVE STIMULUS FACTORS

Experiment 1: Erroneous Reports About Stimuli Influencing Associative Behavior

It is not uncommon for two people to begin humming the same song at the same time, to think suddenly of an old, mutual friend, or in some other way simultaneously to have identical thoughts and associations. Such simultaneous associations are used by some as evidence for the existence of ESP, and probably are regarded universally by lovers as evidence of a mystical bond of union. Hard-bitten scientists are wont to scoff at such occurrences, dismissing them as mere coincidences bound to occur in a world of limited popular songs and acquaintances. A third alternative between the extremes of the spiritual and the merely statistical is possible, however. If, as we have claimed, people are rather poor at identifying the source of influences on their thought processes, then it is possible that these simultaneous associations often occur as a result of some common external stimulus, such as a sign in a shop window that reminds two individuals of a particular popular song. When they both start humming this song, however, neither is able to examine his associative processes adequately enough to recognize their true source, and the simultaneous behavior remains unexplained.

Overview

The general procedure of the experiment was to have subjects memorize a paired-associate word list and view a series of color slides. The contents of both tasks were designed to influence their responses on a subsequent word association task. The subjects were then questioned as to their awareness of the influence of the words and pictures on their responses. The word-association task was a modified one where the experimenter said the name of a category, such as, "U.S. City," and the subjects were to say the first thing in that category that came to their minds, such as, "New York." There were eight such categories, given to all subjects.

Table 1. Word Cues, Picture Cues, Word-Association Categories, and Target Responses.

Word Pair	Picture	Word-Association Category	Target Response on Word-Association Task
jelly-purple	fruit stand	fruit	grapes
ocean-moon	beach scene	brand of detergent	Tide
mold-yellow	fondue dish	dairy product	cheese
empire-subway	statue of liberty	U.S. City	New York
fork-plate	hold-up scene	weapon	knife
pizza-olive	Pope Pius	foreign country	Italy
slave-free	Lincoln automobile	U.S. President	Lincoln
snow-holiday	snow scene	month	December or January

Eight pairs of words and eight color slides were used, with one word pair and one slide designed to influence each of the eight categories. The word-association categories, word pairs, and pictures employed, plus the responses that the words and pictures were intended to elicit (target responses), are listed in Table 1. The setting was a small, well appointed conference room in the Institute for Social Research at The University of Michigan. The subjects, 81 male students from introductory psychology classes, participated in groups of 3-11.¹

Method

Procedure. A Latin square design (cf. Plan 6 in Winer, 1962:543) was used, where four groups of subjects each gave associations to all eight of the categories, yet the word and picture cues given differed between the groups. Each group received both the word and picture cues for two of the categories, the word cues only for two of the categories, the picture cues only for two of the categories, and no cues for the final two categories. A set of four filler word pairs and four filler pictures was also presented in every condition.

Subjects were told that the experiment was concerned with "various cognitive and perceptual processes," and would be divided into three parts: A verbal learning task, a perceptual aesthetics task, and a word-association task.

Word cues. For the "verbal learning task," the subjects were told that they would be asked to memorize a list of eight pairs of words, in forward and reverse order, and that at the end of the experiment they would be asked to recall as many of the words as possible. A printed list of the words was given face down to each subject. At a signal from the experimenter, subjects turned the lists over and examined them for thirty seconds, after which time the lists were collected. Next, two identical trials were run sequentially as follows: The experimenter held up cards with the first word of each pair printed on them. The subjects wrote down on an answer sheet the word each stimulus word had been paired with. After holding up each card for approximately five seconds, the experimenter flipped the card over, displaying the correct response word. Subjects were instructed to leave their answer sheet blank if they did not get the correct answer in the specified time. In similar manner, a second set of identical trials was run in which the original response word of each pair was presented first and the subjects wrote down the original stimulus word for each pair. Thus, all subjects had two forward learning trials and two backward trials.

Picture cues. For the "perceptual aesthetics" task the experimenter explained that he was interested in seeing what combinations of colors were aesthetically pleasing to people. He said that he would show eight color slides, most of which were pictures from magazines. Each slide was shown for fifteen seconds. After viewing each one the subjects rated them on a five-point scale ranging from "very pleasing" to "very displeasing."

¹ When group means rather than individual scores were used as the unit of analysis, conclusions were unaffected. The resulting *t* values were virtually identical, though *p* values were slightly higher when group means were used due to a loss of degrees of freedom.

Word association. Next the word-association task was begun. The experimenter explained, "I will say the name of a category, and I want you to write down on your answer sheet the first thing in that category that comes to your mind. For example, if I said, 'novel', you might write down 'War and Peace'." The experimenter then named the eight categories, in the order that they are listed in Table 1, pausing after each one to give the subjects a chance to write down their responses.

Open-ended awareness probe. Next, the experimenter handed out a sheet of paper containing blanks numbered from one to eight, with an inch of space between each number. The experimenter explained that he was interested in finding out "why you said what you did for each category." He asked the subjects to introspect for each category, and to write down all the reasons they could think of. They were paced through the eight categories: The experimenter said the first category, gave the subjects 1.5 minutes to write down their reasons, then said the second category, etc.

Direct awareness probe. For the direct probe, the experimenter explained that he was interested in finding out if any of the words or pictures they had seen earlier influenced their responses to the categories. A sheet of paper identical to the one used in the open-ended awareness probe was handed out, and the subjects were again paced through the eight categories. They were asked to indicate for each one whether or not any of the words or pictures had influenced their response, and if so, to indicate which ones.

After all the answer sheets had been collected, the subjects were asked to recall, on another blank piece of paper, as many of the word cues they had memorized earlier as they could.

Results

The effect of the word cues either alone or in combination with picture cues was to double the probability that subjects would give the target response on the word-association task. Every one of the eight pairs of word cues increased the probability of the intended target response.

Since every subject responded to categories in four cue conditions (word cued, picture cued, both word and picture cued, and uncued), and since there were two categories in each condition, four indexes were constructed for each subject. A score of 0, 1, or 2 was assigned for each cue condition, depending on whether the subject gave 0, 1, or both possible target responses in that condition. (E.g., in one condition subjects were given word cues for the categories, "U. S. City" and "Foreign Country." If a subject gave the target response for both categories, i.e., "New York" and "Italy," he received a score of 2 for word cues. If he gave only one of these target responses he was given a score of 1, and if he gave neither he received a score of 0. Each subject received such a score for the word-cued categories, the picture-cued categories, the word- and picture-cued categories, and the uncued categories.)

These scores were used as the basis for dependent t tests contrasting each experimental condition with the control, uncued condition. This score was available for each subject for each condition, thus the degrees of freedom for each contrast is 80. A baseline of 10% of word-association responses in the uncued control condition were target responses. This fraction increased to 19% with word cues ($t = 2.35, p < .025$)² and to 20% with both word and picture cues ($t = 3.67, p < .0005$). The effectiveness of the picture cue manipulation was marginal (16% target responses, $t = 1.60, .10 < p < .15$). Since picture cues by themselves did not unambiguously increase the probability of target responses, and since picture cues did not significantly increase the probability of target responses when combined with word cues, results for picture cues will not be discussed further, except to point out that subjects only very rarely identified the pictures as having influenced their associations.

It is important to note that the word cues were quite retrievable by subjects at the time they were questioned about the reasons for their responses. The mean

² All p values are based on two-tailed tests.

number of words recalled at the end of the experiment was 14.4 out of a possible 16.

Despite the potent effect of the word cues on the word associations, and despite their recallability, they were not very salient as explanations for giving the target response. Table 2 presents the reasons given by subjects for their associations in response to the open-ended question. Subjects were not at a loss for reasons as to why they gave the target responses. 1.85 reasons for each target response were given on the average, and for only 3% of the target responses were subjects unable to find any reasons at all. But the word cues came in last as a percent of all reasons. A total of 94 reasons for target responses were given in the word-cued conditions, yet only two of these referred to the word cues as being responsible for a target response.

The responses to the direct probe, in which subjects were specifically asked if the words or pictures might have influenced their responses, tell a superficially somewhat different story. Of the subjects who gave a target response to a word-cued category, 31% reported that the word cues had influenced them to give the target response. There is good reason to suspect, however, that these reports of influence do not really represent accurate awareness of the role played by the word cues in generating the target responses. For each of the word-association categories, the

observed frequency of target responses in the word-cued conditions was compared to the expected frequency based on the control condition. The former frequency minus the latter gives an estimate of the number of target responses in each word-association category that were prompted by the word cues. This estimate could then be compared to the number of "aware" subjects, i.e., those who reported that the target response was prompted by the word cues. The number of "aware" subjects divided by the number of influenced subjects gives a kind of "awareness ratio," that is, the percent of subjects influenced who reported influence. If all subjects who gave the target response due to the word cues were aware of this influence, this ratio should be approximately unity. In fact, however, this ratio ranges from 0 to 2.43 across the eight word-association categories, and only three of the eight ratios were within .40 of unity. In other words, for some of the target words, none of the subjects reported any influence of the word cues while for others, many more subjects reported an influence than were probably influenced! This fact gives a quite different impression of subjects' ability to report on the influence of word cues than is given by the blanket average of reported awareness across all eight categories.

It was our suspicion that the subjects who reported an influence of the word

Table 2. Reported Reasons for Giving Target Responses (As Percent of Total Reasons).

Reason for Target Response	Cue Condition				Row Total
	Words and Pictures	Words Only	Pictures Only	Neither	
1) Distinctive Feature of Object	16%	13%	28%	10%	17%
2) Long-time Association with Object	8%	22%	11%	15%	14%
3) Object is Well Known	15%	13%	11%	10%	13%
4) Recent Experience with Object	15%	9%	17%	15%	13%
5) Object Heads its Class List (e.g., January)	10%	11%	11%	5%	10%
6) Liking for Object	20%	7%	6%	5%	10%
7) Personal Meaning of Object	4%	9%	8%	5%	7%
8) Anticipated Experience with or Current Desire for Object			0%	20%	5%
9) Disliking for Object	4%	7%	0%	5%	4%
10) Can't Think of Any Reason	0%	4%	0%	10%	3%
11) Picture Cues	4%	0%	8%	0%	3%
12) Word Cues	2%	2%	0%	0%	1%
	100%	100%	100%	100%	100%

Note: N of subjects = 81, N of total reasons given = 150.

cues were in fact merely giving a judgment about how plausible it seemed that the word cues would elicit the target response. In order to check this, we gave the word-pair cues and the target words to a different sample of 47 male introductory psychology students and asked them to judge how likely the word cues would be to elicit the target words in a word-association task. The rank order correlation of the judgments of these "observer" subjects was .55 with the "awareness" ratios for the original subjects. The rank order correlation of the judgments was $-.45$ with the *actual* strength of the cuing effect. Thus it seems highly likely that when subjects in the original experiment reported an influence of word cues on target responses they were merely reporting the *plausibility* of such an influence based on the judged similarity of cue and target words, in a manner similar to the judgments of observer subjects.

*Experiment 2: Position Effects on
Judgments of the Quality of
Consumer Goods*

The question of people's ability to report accurately on the causes for their evaluative judgments and behavior has been examined only very rarely. One of the few reports of such an investigation that the authors were able to find was by Laird (1932), who conducted an experiment in which housewives rated the quality of four pairs of silk stockings. Unbeknownst to the subjects, all of these stockings were identical, except that each had a different scent applied to it. The stockings with one particular scent were chosen much more often than the others; yet, Laird reported, the housewives were not even aware that there were any scents on them. Thus Laird's experiment would appear to supply evidence that people are poor at reporting the influence of a particular stimulus manipulation (in this case, of a scent) on their evaluative judgments. We attempted to replicate this experiment, with the hope that it would duplicate the demonstration that people are not aware of the reasons for their evaluations. As it turned out, the scents we applied had no effect on people's choices, probably

because we used only a small amount and were situated in a drafty location near a store entrance. The experiment did provide an interesting test of people's ability to report on causal factors influencing their evaluations, however, since, as it happened, the left-to-right position of the stockings had an effect on evaluations.

Method

The experiment took place in a large bargain store in a shopping center. Card tables were set up with signs attached to the tables that read, "Institute for Social Research—Consumer Evaluation Survey—Which is the Best Quality?" Four identical nylon stocking pantyhose (Agilon ®, cinnamon color) were hung from racks on the tables, approximately three feet (.91m) apart, with the top of the stockings just below eye level.

Subjects were passersby who voluntarily approached the stocking display and made a judgment as to which one they thought was the best quality. A total of 52 people participated in the study, 50 of whom were female.

Both the scent and the position of the stockings were counterbalanced. When subjects approached the display, one of the two male experimenters asked which stocking they thought was the best quality. When the choice was made, the experimenter recorded it and said, "Could you tell me why you chose that one?" After responding, subjects were asked if there were any other reasons for their choice. All reasons were recorded.

Results

The scents had no effect on stocking choice, but the position of the stockings on the table had a substantial effect. The further their position was to the right, the more likely the stockings were to be chosen as being of the best quality. Stockings in position A, the left-most position, were chosen by 12% of the subjects, stockings in position B by 17% of the subjects, in position C by 31% of subjects, and in position D by 40% of the subjects ($\chi^2 = 10.62$, 3 df, $p < .025$).

Despite the fact that stockings were

identical, few subjects appeared to experience difficulty in making a choice. Only two subjects voiced a suspicion that the stockings were identical. Only ten subjects expressed an initial hesitancy in choosing one of the stockings over the others, and all of these responded to a single probe to try a second time to make a judgment. Most of the subjects walked up to the table, fanned out the top of the stocking to examine the weave, examined the heel and toe, kneaded the stocking between their fingers, and briskly announced their choices. Only a quarter of the subjects required any prompting to explain the basis of their choices. Most of the subjects promptly responded that it was the knit, weave, sheerness, elasticity, or workmanship that they felt to be superior. A total of 80 reasons, for an average of 1.54 per subject, were recorded. Not a single subject mentioned the position of the stockings as a reason for the choice. When subjects were directly probed as to whether the position of the stockings could have influenced their decisions, all but one denied that such a thing could have had any influence. The subject who admitted the likelihood of influence said that she was currently taking three psychology courses and knew all about the power of order effects, and therefore felt she probably had been influenced. (In fact, however, this subject showed no evidence of the right-ward bias. She chose the stockings in position B.)

The position effect is of some interest in its own right (if only to merchandisers), and to our knowledge has not been demonstrated previously. We do not know why the effect occurs, though for what it is worth, it is our speculation that it is in fact a temporal order effect rather than a spatial position effect. Subjects almost universally "read" the stockings from left to right, probably because they were labeled A, B, C, D. By the time subjects reached object D, they had probably found something they did not like about one or two of the previous objects in the array, but could not remember precisely which one or ones. Then when judging D, if they found nothing wrong with it, they were likely to choose it, since they knew at least some of the others, but not

exactly *which* ones, were less preferred. If they did find something wrong with D, then they would be likely to shift to C if they could not specifically recall anything wrong with C. What matters for present purposes though, is not why the effect occurs but *that* it occurs and that subjects do not report it or recognize it when it is pointed out to them.

REPORTING THE INFLUENCE OF INEFFECTIVE STIMULUS FACTORS

Experiment 3: The Emotional Impact of Literary Passages

The fact that a good critic can often cause us to completely change our minds about why we liked or did not like a particular artistic product ought to alert us to the possibility that we do not have very good access to the precise means by which an artistic product exerts its effects. Yet most people, when questioned about why they do or do not like a book, movie, or painting, readily provide explanations. Such causal analyses are usually satisfactory to both speaker and hearer. Experiment 3 suggests that such confidence may be ill-founded.

Overview

Subjects were requested to read a selection from the novel, *Rabbit, Run*, by John Updike (pages 213–220 in the Fawcett Crest 1960 paperback edition). This selection describes an alcoholic housewife who has just been left by her husband and is desperately trying to make it through the morning, tending her two young children and cleaning up her house, preparing for a visit from her mother. In her drunken state she attempts to give her baby daughter a bath, and the baby accidentally drowns. The selection is well-written, and has a substantial emotional impact even when read out of context.

Two brief (2-sentence) passages from this selection were either deleted or included, with the idea that one of them, when absent, would significantly reduce the emotional impact of the selection. The baby, Rebecca, is generally referred to in the book in abstract terms, as simply a baby, except for one passage in which the

child is described physically in a much more concrete, visual fashion than anywhere else in the selection. It was anticipated that deletion of this passage (passage 1) would decrease the impact of the total selection. The second brief passage (passage 2) concerned the messiness of the house. It was anticipated that deletion of passage 2 would have no effect on the overall impact of the selection.

Method

The setting was a large conference room, furnished with couches and comfortable chairs. The subjects were 76 male and 76 female introductory psychology students at The University of Michigan, drawn from the psychology department subject pool.³ Subjects participated in groups of 10 to 15 persons. The experimenter told the subjects that the experiment was concerned with the "impact of different types of information when presented in different media forms." He continued, "We've been presenting various types of information to people using video tape, film, sound recordings, and written materials, and seeing what kind of impact this information has on people." In this session, he said, the subjects would be reading a brief selection from a novel and viewing a documentary film.

The selections from *Rabbit, Run* were then handed out, with a questionnaire at the end. The subjects were given a brief description of who the characters were and were instructed to read the selection page by page at their own pace and not to look ahead. They were not informed until the end of the questionnaire that the passage was from *Rabbit, Run*, at which time they were asked if they had read it before. Ten indicated that they had, and were not included in the analyses.

There were four different versions of the selection. One was the original selection as it appeared in the novel, one had only passage 1 deleted, one had only passage 2 deleted, and one had both passages deleted. All four versions were handed out

in a random order during each experimental session. The subjects rated the emotional impact of the selection on a seven-point scale ranging from "had no impact at all" to "had an extremely powerful impact." Subjects then read, or reread, passages 1 and 2. Subjects who had previously read a particular passage were asked, "how much you think it contributed to the overall emotional impact of the selection?" This was rated on a seven-point scale ranging from "greatly increased the impact" to "greatly decreased the impact." Subjects who had not previously read the passage were asked how they thought the passage would have contributed to the impact of the selection had it been included.

Results

As expected, passage 2 had little or no effect on the impact of the overall selection. Contrary to expectations, however, passage 1 also had no effect. As seen in Table 3, the impact of the selection was very similar regardless of whether either passage was included. One should be cautious about accepting the null hypothesis of no differences between means, but as seen from the Scheffé 95% confidence intervals in Table 3, it is clear that if there are differences, they are very small ones. There is only a 5% chance that the effects of passages 1 or 2 are greater than 1/2 to 2/3 of a scale point in either direction, and the center of both confidence intervals is close to zero.

It may be seen in Table 4, however, that subjects believed that both passages increased the impact of the selection. The two columns on the left of Table 4 show the percent of subjects reporting that passages 1 and 2, respectively, increased, decreased, or had no effect on the impact of the selection. Only a third of the subjects reported that passage 1 had no effect, and two and a half times as many subjects reported that passage 1 increased the impact as reported that it decreased the impact. Subjects were virtually unanimous in their opinion that passage 2 increased the impact of the selection. It should be noted that for neither passage 1 nor passage 2 was there a correlation between reported

³ There were no significant differences between males and females on the dependent measures. Thus, the data for the two sexes were combined.

Table 3. Means and Confidence Intervals for the Reported Impact of the Selection from *Rabbit, Run*.

	Means			Upper & Lower Limits of 95% Confidence Intervals for Differences Between the Marginal Means	
	Passage 2			Column 1— Column 2	Row 1— Row 2
	Present	Deleted	Row Means		
Passage 1					
Present	4.61	4.63	4.62	-.57, .63	-.49, .71
Deleted	4.55	4.46	4.51		
Column Means	4.58	4.55			

Note: Scale ranged from 1-7; 1 = "Had no impact at all," 7 = "Had an extremely powerful impact." All cell n's = 38.

effect of the passage and reported impact of the selection. Thus it would be difficult to argue that the passages actually increased the impact of the selection for some subjects and decreased it for others, and that both types of subjects reported accurately on these effects.

The columns on the right in Table 4 show the predictions about the probable effect of passages 1 and 2 by subjects who did not receive the passages. It may be seen that subjects' predictions about the effects of the passages closely resemble the reported effects of the passages by subjects who were actually exposed to them. And, of course, the predictions do not at all resemble the actual effects of the passages.

Thus subjects in this experiment were shown to be incapable of either reporting correctly on the effects of a particular passage on the impact of a literary selection or of accurately predicting these effects.

Experiment 4: The Effect of Reassurance on Predictions About Willingness To Take Electric Shock

It is common for both laypeople and social scientists to ask a person why he

performed a particular behavior or chose a particular course of action. Neither the layman nor the social scientist is often disappointed. A confident analysis of the stimulus situation at the time of action is usually forthcoming and the role of particular stimulus components in determining action is often briskly delineated. Experiment 4 suggests that, confidence notwithstanding, such reports may be in error.

In a recent series of experiments by Nisbett and Borgida (1975) subjects read a description of a study in which male students were requested to take large amounts of electric shock, and were then asked to predict how much shock they themselves would take if they were to be placed in the experimental situation. The procedural protocol included the "reassurance" that the shocks would do no "permanent damage." After subjects had made their predictions about the amount of shock they thought they would take, they were asked what had been the effect on their predictions of the statement concerning permanent damage. Most subjects felt that this had played a part in their predictions about their behavior, and 90 percent of those who believed the factor

Table 4. Percent of Subjects Reporting or Predicting Influence of the Passages on the Impact of the Selection.

Response	Reported Influence of:		Predicted Influence of:	
	Passage 1	Passage 2	Passage 1	Passage 2
Increased Impact	47%	86%	42%	71%
Had No Effect	34%	9%	30%	16%
Decreased Impact	18%	5%	28%	13%

Note: n = 76 for each column.

to be important said that it had *increased* the amount of shock they thought they would take. This was surprising, since the phrase had actually been included in the original experiment (Nisbett and Schachter, 1966) in order to frighten subjects about the shock. The phrase "no permanent damage" had been intended to call up the specter of temporary damage, not to allay fears about the permanence of the effects.

Since the conflict between the intentions of the original experimenters and the report of Nisbett and Borgida's subjects was so great, the Nisbett and Borgida study was replicated. The procedure for one group of subjects was identical to that in the Nisbett and Borgida study, whereas for another group the "reassuring" phrase about permanent damage was left out. This made it possible to test the effect on subjects' predictions of inclusion of the phrase. The anticipation was that inclusion of the phrase would actually produce less willingness to take shock while subjects would assert that it had produced more willingness to take shock.

Method

Subjects received the same protocol describing the Nisbett and Schachter (1966) shock experiment that subjects in the Nisbett and Borgida (1975) experiment received. The gist of the shock experiment is that subjects are told that the effect of a drug on skin sensitivity is being studied. The test of skin sensitivity would be reactions to electric shock. The study is described as important because electrical accidents are becoming more frequent and the use of electroshock therapy has become widespread; yet little is known about the effects of intense electric shock. It is stressed that the shocks must be intense if anything is to be learned, and for experimental subjects the original phrase, "but of course they won't do any permanent damage," was included. For control subjects, this phrase was deleted. Subjects were then asked to indicate how much shock they thought they would take if they were to participate in the experiment. They answered on a six-point scale with end points labeled "would have re-

fused to participate" and "would have gone all the way to a jolt causing entire arm to jerk." Then subjects in the experimental condition ($n = 35$), who had received the "no permanent damage" phrase, were asked in open-ended form what had been the effect, if any, of the experimenter's statement that the shock would cause no permanent damage. These answers were coded into reports of increased willingness to take shock, decreased willingness, and no effect. Subjects in the control condition ($n = 40$) were shown the phrase and asked what effect, if any, the phrase would have had on their predictions had it been included. Subjects were run in groups, with both versions of the questionnaire handed out in each session in a random order.

Results

Contrary to expectations, the phrase concerning permanent damage had no detectable effect on predictions about shock-taking. Experimental subjects, who received the phrase, predicted they would take about the same amount of shock as did control subjects who did not receive the phrase (3.30 vs. 3.48, respectively). The distributions of predicted responses, moreover, were almost identical for the two groups.

Despite the fact that inclusion of the phrase had no detectable effect, subjects who received the phrase were for the most part of the opinion that it did have an effect. Table 5 presents the reported effects of the phrase on willingness to take shock for the Nisbett and Borgida subjects, and for the comparable subjects in the experimental condition of Experiment 4; and presents the beliefs of control subjects about the likely effects of the phrase had it been included in the description they received. Only the responses of subjects who predicted they would take at least some shock, but less than the full amount, are included in Table 5, since subjects who predicted they would not participate in the experiment logically could not say that the phrase increased the amount of shock they were willing to take, and subjects who said they would be willing to take the full amount of shock logi-

Table 5. Number of Subjects Reporting or Predicting Increased, Decreased, or Unaffected Willingness to Take Shock Due to the "Reassurance".

Reported or Predicted Effect of Reassurance	Nisbett and Borgida Subjects	Experiment 4 Experimental Subjects	Experiment 4 Control Subjects
Increased	32	8	6
Decreased	6	2	3
No Effect	31	10	9

cally could not say that the phrase decreased the amount of shock they were willing to take.

It may be seen in Table 5 that the majority of Nisbett and Borgida subjects felt that the phrase had influenced their willingness to take shock, and that of those who believed the phrase had an influence, the overwhelming majority (84%) believed the effect was to increase their willingness to take shock. Similarly, half of the experimental subjects in Experiment 4 felt that the phrase had had an effect, and of these, 80% felt the effect was to increase their willingness. Finally, control subjects' predictions about the effects of inclusion of the phrase were similar to the incorrect reports of experimental subjects. Half believed the phrase would have had an effect and of these, two-thirds believed the phrase would have increased their willingness to take shock. Thus, again we find little correspondence between the reported and actual effects of a stimulus on a response.

DISCUSSION

Possible Methodological Confounds in the Present Research

Each of the present studies casts doubt on people's ability to report accurately about the effects of stimuli on their judgments and behavior. Collectively, they raise very serious questions about the validity of such reports in previous research and about the scientific utility of asking people for predictions about their behavior in various situations. Some of the studies are themselves subject to methodological criticisms, however, and these should be briefly discussed.

Social desirability confounds. It could be argued that some of the verbal reports were produced by a desire on the subjects'

part to look good—i.e., sensible and reasonable—in the experimenter's eyes or in their own eyes. It certainly seems likely, for example, that subjects would have been motivated to deny a serial position effect on their judgment of garment quality, had they known of its existence. On the other hand, if subjects were aware of their propensity to show such an effect, why should they not have counteracted it and deliberately eliminated its impact? It seems scarcely likely that subjects would have made a judgment of garment quality that was knowingly contaminated by the position effect. It is much more plausible to accept their denials at face value—that they did not in fact know that the position of the garment had affected their judgment. Social desirability effects are not very compelling explanations for the verbal report results of the other experiments because, simply put, it is hard to see why subjects would be motivated to report one effect vs. another. It scarcely seems embarrassing to report either that one's verbal associations were or were not affected by word pairs previously memorized (Experiment 1), that one's reaction to a literary passage was or was not affected by a particular paragraph (Experiment 3), or that one's predictions about the amount of shock one would take in an experiment were or were not affected by a "reassurance" about the effects (Experiment 4). It should be emphasized, however, that the methodological implications of the present research would not be affected if the verbal reports were the result of social desirability or self-esteem maintenance motives. If subjects base their reports primarily on such motivational considerations even for such inconsequential matters as the present judgments and behaviors, then their reports about matters that affect

them more deeply could scarcely be relied upon.

Experimenter demand effects. The traditional experimenter demand effect, where experimenters hold a certain hypothesis and communicate this to compliant subjects, could not possibly account for the present results. The experimenters did indeed hold strong hypotheses about how subjects would behave in each of the experiments and about the verbal reports that subjects would give, but in virtually no instance did subjects' behavior and reports correspond to experimenter expectations. An experimenter demand critique would have to take a different tack; to wit, that the subjects incorrectly perceived the experimenter's hypothesis and then tried to provide results consistent with their inaccurate perceptions. Or perhaps they correctly perceived the experimenter's expectations and then deliberately traduced them. We suspect that the reader will agree that it would take a genius or psychic, and a particularly perverse genius or psychic at that, to deduce the experimenter's expectations from the instructions and then to behave in such a fashion as to provide the opposite results. It is almost equally difficult to imagine what cues would have sufficed to allow a putatively compliant subject to deduce an expectation that was precisely the opposite of the experimenter's actual expectations. Again, of course, the methodological implications of the present work would be unaffected if the results were somehow due to prescience or perversity on the part of subjects. If subjects provide verbal reports about the causes of their behavior to researchers primarily because that is what they think the researcher wants to hear (or what the researcher doesn't want to hear), then such reports are quite valueless as indicators of the subjects' true reasons for their evaluations or behavior.

THE BASIS OF VERBAL REPORTS ABOUT CAUSAL EFFECTS ON JUDGMENTS AND BEHAVIOR

If it is granted that the present results reflect a fundamental incapacity to report accurately about the effects of stimuli on

responses, then the question arises as to what it is that people do when they think they are introspecting about the effects of stimuli upon their responses. Nisbett and Wilson (1977) have discussed this question at length, and their arguments will be briefly summarized here.

It will be recalled that *predictions* about stimulus effects made by subjects in Experiments 1, 3, and 4 who were not exposed to particular stimulus components were quite similar to the erroneous *reports* about their effects given by subjects who were actually exposed to them. The former subjects, who simply read a description of the critical stimuli, were operating as "observer subjects" (Bem, 1967). Since subjects and "observer subjects" give similar answers, it seems likely that the subjects themselves are not, as Bem put it in a similar context, drawing on "a fount of privileged self-knowledge" (Bem, 1967:186). In all likelihood, subjects do not even attempt to examine their personal "founts" when asked about the causes of their evaluations or behavior. Instead subjects, like observer subjects, probably rely on a *priori causal theories* about the effects of particular stimuli on particular responses. In Abelson's (1968) terms, this judgment is an assessment of the extent to which a given stimulus "psychologically implies" the given effect. In Tversky and Kahneman's (1974) "cognitive heuristics" terminology, subjects and observer subjects make "representativeness" judgments, that is, they assess whether the effect is similar to the stimulus or is representative of the sorts of effects the given stimulus would be expected to produce.

Subjects in the present experiments therefore probably did not even attempt to examine searchingly the data of personal experience when asked about the causes of their evaluations and behavior. Instead they went about the task of judging cause and effect in essentially the same manner as observers—by making judgments about the plausibility or representativeness of potential causes and given effects. Thus, a "reassurance" seems a plausible cause of taking more shock, while the position of stockings in an array does not seem a plausible reason for liking stockings.

It should be emphasized that this view does not contend that people will *always* be wrong in their verbal reports about the effects of stimuli on their responses. Even though there may be little or no access to the cognitive processes mediating these effects, there will be times when the *a priori* causal theories invoked by people will correspond to the actual causes that influenced the responses. Nisbett and Wilson (1977) discuss at length the conditions that should affect accuracy and inaccuracy in self-reports. In general, reports should be accurate whenever (a) the influential stimulus is a plausible cause of the response, (b) is highly available in memory, and (c) there are few or no plausible but noninfluential factors available. Nisbett and Bellows (1977) have shown recently that subjects are indeed sometimes correct about the influences on their behavior. In their study, however, subjects were correct if and only if observer subjects, operating only with an impoverished description of the stimulus materials, were also correct in their predictions about the stimulus influences. Thus, even when subjects were correct in their reports, there is no reason to assume that introspection was of any assistance to them. Instead, subjects, like observers, were simply applying a causal theory correctly in those instances.

IMPLICATIONS FOR EXPERIMENTAL AND SURVEY RESEARCH

The experiments reported here call into question the accuracy of verbal reports about cognitive processes, and question the use of these reports by social scientists. Such reports may well be based more on people's *a priori* causal theories about stimulus effects than on direct examination of their cognitive processes, and will be inaccurate whenever these theories are inaccurate. The present results indicate that inaccuracy may be common and imply that investigators should not place themselves at the mercy of such verbal reports.

On the other hand, we are most certainly not proposing that researchers, either in experimental studies or in survey studies, cease to ask their subjects ques-

tions about their perceived reasons for their behavior. Even though such verbal reports are highly suspect as indicators of causal influences, they can be very useful in other ways. In experimental work, for example, the responses to such questions can provide valuable information about how subjects perceived (or misperceived) the stimulus situation, leading to a revision of the experimental design. In survey research, where the worlds of the respondents and researchers are often quite different, questions about perceived effects of stimuli can often lead to the generation of new hypotheses, or to the inclusion of questions that never would have occurred to the researcher. Thus, we are not suggesting that researchers turn a deaf ear to the reports of their subjects about stimulus effects. Such reports can be very useful, so long as they are not considered to be accurate assessments of the actual influence of stimuli on their responses.

REFERENCES

- Abelson, R. P.
1968 "Psychological implication." In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, and P. H. Tannenbaum (eds.), *Theories of Cognitive Consistency: A Sourcebook*. Chicago: Rand McNally.
- Bem, D. J.
1967 "Self-perception: An alternative interpretation of cognitive dissonance phenomena." *Psychological Review* 74:183-200.
- Brown, R.
1962 "Models of attitude change." In R. Brown, E. Galanter, E. H. Hess, and G. Mandler (eds.), *New Directions in Psychology* (Vol. 1). New York: Holt, Rinehart and Winston.
- Kelman, H.
1966 "Deception in social research." *Transactions* 3:20-24.
- Laird, D. A.
1932 "How the consumer estimates quality by subconscious sensory impressions." *Journal of Applied Psychology* 16:241-246.
- Nisbett, R. E., and N. Bellows
1977 "Verbal reports about causal influences on social judgments: Private access vs. public theories." *Journal of Personality and Social Psychology* 35:613-624.
- Nisbett, R. E., and S. Borgida
1975 "Attribution and the psychology of prediction." *Journal of Personality and Social Psychology* 32:932-943.
- Nisbett, R. E., and S. Schachter
1966 "The cognitive manipulation of pain." *Journal of Experimental Social Psychology* 2:227-236.

SYSTEMATIC THEORY OF POWER ATTRIBUTION

- Nisbett, R. E., and T. D. Wilson
1977 "Telling more than we can know: Verbal reports on mental processes." *Psychological Review* 84:231-259.
- Personality and Social Psychology Bulletin
1977 Volume 3 (Fall). Washington D.C.: Society for Personality and Social Psychology.
- Tversky, A., and D. Kahneman
1974 "Judgment under uncertainty: Heuristics and biases." *Science* 185:1124-1130.
- Winer, B. J.
1962 *Statistical Principles in Experimental Design*. New York: McGraw-Hill

Social Psychology
1978, Vol. 41, No. 2, 131-148

Towards a Systematic Theory of Power Attribution*

STAN A. KAPLOWITZ
Michigan State University

This paper develops and discusses a large number of general propositions about the consequences of power attribution and about how people make such attributions. In discussing the consequences of power attributions, it specifies when the appearance of power is useful to an actor and when it is not. In discussing how power is attributed, we focus on the effect of information about outcomes on power attributions. Among the questions dealt with are the following: (1) Does benefiting greatly or benefiting slightly from an outcome lead to a greater amount of attributed power? (2) Does winning on an important decision indicate great power or great effort? (3) Does transferring resources to others indicate that one is wealthy or that one is being coerced? The final section looks at ways in which actors and observers differ in the power attributions they will make.

From the middle 1950's to the middle 1960's, power was one of the most debated topics in American social science. Some (e.g., Dahl, 1957; French, 1956; Harsanyi, 1962; Simon, 1957) tried to give precise definitions of power and proposed methods of measurement. Others studied power in American local communities and in America society as a whole.

Different students of power in American society came to very different conclusions. Some (Hunter, 1953; Mills, 1956; Domhoff, 1967, 1970) concluded that power was concentrated in the hands of a small, unified elite. Others (Dahl, 1961; Polsby, 1963; Rose, 1967; McFarland,

1969) saw a pluralistic power structure in which many different groups have power. The disparate results and some of the theoretical analyses of the concept of power led some social scientists (e.g., Riker, 1964; March, 1966; Wagner, 1969; and Wolfinger, 1971) to suggest that the concept of power is unmeasurable, has little predictive value, and ought to be abandoned.

More recently, Nagel's (1975) path-analytic approach has offered new possibilities for measuring power accurately and getting beyond the inconclusive debates of the past. But even when they lack the extensive information that Nagel's approach requires, laymen and social scientists are quite ready to attribute power. This indicates that such attributions have important consequences and are worthy of further study.

Michener *et al.* (1973) and Schopler and Layton (1974) have already done some theoretical and experimental work on how people attribute power to others. The work of Winter (1973) and McClelland (1975) deals with some of the conditions under which people attribute power to

* I would like to thank James L. Phillips, Eileen Thompson, James Gard, James McKee, Richard C. Hill, Sheldon Levy, Jack Nagel, William Ewens, William Gamson, Christopher Vanderpool, Joey Reagan, Michael Doubek and anonymous referees for their insightful comments on an earlier version. My wife Toba provided valuable editorial assistance. An earlier version of this was presented at the meetings of the Midwest section of the Peace Science Society (International) in Chicago, May 1975. Address communications to the author at the Department of Sociology, Michigan State University.