

TARGET ARTICLE

The Really Fundamental Attribution Error in Social Psychological Research

John Sabini and Michael Siepmann

*Department of Psychology
University of Pennsylvania*

Julia Stein

*Department of Psychology
Bryn Mawr College*

We review the classic studies on social influence and the fundamental attribution error to determine (a) whether it is true that behavior of the sort observed in those studies is externally caused in the two senses of external causality used by attribution theorists, (b) whether laypeople have been shown to overestimate the extent to which behavior is internally caused in either of those two senses, and (c) whether there is a better way to characterize the errors people make. We conclude that (a) behavior in those studies has not been shown to be externally caused in the two senses used by attribution theorists, (b) people have not been shown to overestimate the extent to which behavior is internally caused in either of those two senses, (c) there is a different sense of internal versus external causality that better characterizes the errors people make, and (d) these literatures taken together suggest that Americans are far more disposed to preserve face and avoid embarrassment than most people had suspected.

What are the important findings of social psychology since World War II? Were one to survey social psychologists with that question, we believe that many would give something like the following answer: Social psychologists have discovered that, to a far greater degree than laypeople realize, and than social psychologists had previously realized, people's behavior is caused externally (by situations) rather than internally (by dispositions). The tendency to underestimate the degree to which behavior is externally caused has been called the Fundamental Attribution Error (FAE).¹ A prototypical answer might go on to claim that people make this error because of the way our cognitive machinery works and that the error is therefore independent of any particular content or motivation.² These

answers are certainly in line with the claims of Ross and Nisbett (1991) in a prominent attempt to say just what is important in post-World War II social psychology. In this article we argue, polemically, that giving this answer is the Really FAE.

We make three distinct arguments: (a) Social psychologists have not shown that dispositions in general are significantly less important than laypeople believe them to be; (b) social psychologists have demonstrated that certain dispositions are less important, and others more important, than had been realized; and (c) social psychologists have gathered evidence that suggests that avoiding embarrassment and saving face are more important to Americans than had been recognized. Our third argument is more speculative than the first two.

¹We believe it was Lee Ross who coined the term. Under the heading "The Fundamental Attribution Error," Ross (1977) wrote, "Our exploration of the intuitive psychologist's shortcomings must start with his general tendency to overestimate the importance of personal or dispositional factors relative to environmental influences" (p. 184).

²Two distinct research traditions are alluded to in this answer. The first, older literature has been referred to (and will be referred to here) as the social influence literature (Sabini, 1995). This literature includes at least the Milgram (1974) studies of obedience to authority, the Asch (1952, 1956) studies of conformity to a group, and the

Latané and Darley (Darley & Latané, 1968; Latané and Darley, 1968) studies of the inhibition of bystander intervention. Many social psychologists would want to add studies to this list, but few of them, we suspect, would not want to include these studies among the key studies demonstrating external control of behavior, (See Ross, 1977, as an example of someone who cites these studies as the key studies demonstrating external control.) The second literature taken to demonstrate laypeople's errors has been referred to as the literature on the FAE or Correspondence Bias (see Gilbert & Malone, 1995; Jones, 1979; Jones & Davis, 1965; and Ross, 1977, for reviews and discussions).

The three arguments are independent of one another; someone might accept the first two claims, but reject the third, for example. We offer our third argument as an alternative way of thinking about what we and Ross and Nisbett (1991) take to be the central studies of a certain tradition in social psychology. None of these arguments is new, but bringing them together in one place may be new.³

Ross and Nisbett (1991) have trod much of the ground we cover, and we agree with much of their analysis. But there is one strong difference between us. Ross and Nisbett believed that laypeople suffer from the disorder of *dispositionism*, which is, as we understand it, a pervasive tendency to overestimate the extent to which people's behavior is a product of their dispositions. Ross and Nisbett wrote; "We will show that people (1) infer dispositions from behavior that is manifestly situationally produced, (2) overlook situational context factors of substantial importance, and (3) make overly confident predictions when given a small amount of information" (p. 126). It is the first claim that we disagree with. We heartily agree with the second, and we "name names" about which factors are overlooked. We are agnostic about the third claim.

One way of putting the disagreement between Ross and Nisbett (1991) and ourselves is this: They believed that in a variety of studies, people go wrong in their predictions (understandings) because they have a general tendency to attribute behavior to dispositions rather than to situations. We think, on the other hand, that the problem people have is not that they have a general tendency to attribute one way or another, but that they underestimate the importance of certain specific factors. Ross and Nisbett certainly recognized the importance of these specific, concrete factors; indeed they devoted a chapter of their book to them and referred to them, following Lewin, as "channel factors." So we are in agreement that there are factors that laypeople tend to ignore that turn out to be very important. What is at issue is whether the FAE provides an account of why these factors are important, or why they are ignored, or even whether the FAE provides a useful characterization of the error—explanatory or not. We think it does not; Ross and Nisbett thought it does.

To our minds, the literatures on social influence, cognitive dissonance, and the FAE converge on one idea: Americans (at least) think that they should, and that they do, treat as unimportant certain motives that are in fact

not at all trivial, such as the motives to save face (for oneself and others) and to avoid embarrassment. It is in part because Americans do not acknowledge how important these motives are to them, we argue, that they miss them in predicting behavior. People in other cultures may differ from Americans not so much in that these motives are more important to them (although that may be true to some degree), but rather in whether they acknowledge these motives. We are afraid that this important possibility, derived from the literature we discuss, is being camouflaged in an overly broad internal-external distinction. We are certainly not claiming that other factors are unimportant in the literatures we are discussing; in Ross and Nisbett's (1991) language, we are not claiming that fear of embarrassment, or of losing or harming face, is the only "channel factor" operating in the studies they and we review. But we are claiming that it is a factor that runs through all of this literature. Because we argue that embarrassment plays a key role in the studies we review, we need to say a word about what it is.

Embarrassment and Face

Just what *embarrassment* is is difficult to say. (See Miller, 1996; Parrott & Harré, 1996; Sabini, 1995; Sabini, Siepmann, Stein, and Meyerowitz, 2000, for reviews.) But for our purposes, suffice it to say that embarrassment is an aversive emotional state rooted in social interaction, involving flustering and the desire to hide, flee from, or dissolve the immediate social situation. It has to do with being the center of attention and, often, with loss of face. Most important, it is something people try to avoid.

Face was introduced to American social psychology by Goffman (1955); it refers to the positive aspects of character that a person lays claim to (or is treated as having laid claim to) in a particular interaction. A *face threatening act* is an act that suggests that someone (self or other) is less worthy than their role requires them to be (Brown & Levinson, 1978). Face threatening acts commonly produce embarrassment.

Social Influence Studies

We begin by focusing on those studies commonly understood to show that behavior is under external rather than internal control; these are essentially the same studies Ross (1977) and Ross and Nisbett (1991) reviewed in arguing for the FAE. Do these studies in fact demonstrate that behavior is under situational rather than dispositional control?

³ In particular the internal-external distinction has been criticized many times (Gilbert, 1998; Kruglanski, 1975; Lewis, 1995; Locke & Pennington, 1982; Miller, Smith, & Uleman, 1981; Monson & Snyder, 1977; Sabini & Silver, 1983, 1987; Solomon, 1978; White, 1991). What has not been made so clear is the way in which the difficulties with the internal-external distinction confound and obscure our understanding of the central studies of social psychology.

Obedience

As is widely known, in the Milgram (1974) experiments on obedience to authority, most participants (65% of them in the baseline condition) obeyed the experimenter's commands to shock an innocent, protesting victim. Laypeople, Yale psychiatrists, and Yale psychology graduate students to whom the experiment was described predicted that fewer than 1% of participants would obey.

It is clear that people got something wrong when they predicted what participants would do in that experiment; it is indeed clear that they misestimated the efficacy of some cause or causes of behavior (Milgram, 1974). But just what cause or causes they misestimated is not obvious from the experimental results. Attribution theorists have been inclined to claim that the experiment shows that behavior is under external (situational) control rather than internal (dispositional) control. And indeed, in this experiment, the participants did what the experimenter ordered them to do rather than what their consciences told them to do. A particular external cause, the experimenter's orders, was found to be more important than a particular internal cause, the participant's conscience. With this we certainly agree. But attribution theorists want to make the broader claim that situational causes are more important than dispositional ones *in general* in this experiment. They certainly are not claiming merely that a specific internal cause is weaker than a specific external cause.

In any case, could we not describe the Milgram (1974) results as showing that people have a much stronger disposition to obey authorities (internal cause) than we thought and are much less responsive to the suffering of others (external cause) than we thought they were? Certainly the data are consistent with this interpretation too. The Milgram results therefore do not show that people are controlled by the situation rather than by their dispositions. We must disagree that behavior is "manifestly situationally produced" in the Milgram experiment, at least insofar as "situationally produced" is seen as an account competing with "dispositionally produced." Rather, the results show that particular dispositions (perhaps including the disposition to obey authority) are stronger than other dispositions (including the disposition to obey one's conscience). Alternatively, they show that people in this experiment were more responsive to a particular aspect of the situation (the demands of the experimenter) than to another aspect of the situation (the suffering of the victim). In our view, the results of the Milgram experiment do not at all demonstrate that dispositional or internal forces are weak relative to environmental forces; rather, they suggest that specific aspects of situations (and the particular motives those aspects

engage) exert more control over behavior than do other specific aspects of situations (and the motives they engage). As Ross (1977) put it, the reason the Milgram studies were so prominent is that they demonstrated "control by a situational factor that the reader had previously assumed to be too weak to exert such control" (p. 187). But what was that situational factor?

The reasons why participants obeyed are obviously complicated. Certainly the "slippery slope" aspect of the situation was crucial, as was the participants' distance from the suffering learner. But we think that the participants' obedience had something to do with their having a difficult time confronting the experimenter with his immorality. As argued elsewhere (Sabini & Silver, 1982), to extricate themselves from this situation, the participants needed to address the fundamental immorality of the experimenter's orders. They did not. Even those who "disobeyed" did not. They merely objected four times to continuing, at which point the experimenter stopped the experiment.

It is undeniably true that attacking the morality of an experimenter is a face-threatening act, which would typically lead to embarrassment. Do participants refrain from forcefully disobeying the experimenter because to do so they must threaten his face and experience embarrassment?

We can add one further bit of evidence. Whatever force kept participants obeying was diminished to an important degree when they communicated with the experimenter by telephone rather than face-to-face (Milgram, 1974). We presume that embarrassment is more intense in face-to-face interaction than in interaction by telephone. So embarrassment is, at the very least, a viable candidate for one cause of behavior in this experiment—one that people may underestimate. (See Milgram, 1974, and Sabini & Silver, 1982, for arguments about why embarrassment might be a leading candidate; and see Sabini, Cosmas, Siepmann, & Stein, 1999, for reasons to believe people might underestimate the prevalence of embarrassment.)

Conformity

Asch (1952, 1956) asked his participants to tell him which of three lines was the same length as a fourth line. The task was easy in the sense that participants tested alone erred only rarely. But when they were tested in the presence of six confederates who gave patently wrong answers, the situation became very difficult indeed. Almost all participants found the situation stressful (Jahoda, 1959), and most participants gave the group's wrong answer at least once.

Bystander Intervention

Asch was surprised that there was so much conformity (Asch, 1952, 1956), and perhaps from the enduring interest in this work we can infer that the results are surprising to most people. Does this research fit the FAE framework? Was behavior really under external control, and did people really think it would be under internal control? Well, people (or at least Asch) expected participants to answer in a way that was consistent with what they truly believed (an internal factor), but instead, they gave answers that were controlled, or at least strongly influenced, by what other people said (an external factor). As with the Milgram (1974) experiment, one is surprised because a description of the situation leads one to expect a particular internal cause to control behavior, but on hearing the results, one realizes that behavior was in fact controlled by a different cause, an external one. But again we ask: Does this show that behavior was caused by external as opposed to internal causes tout court, or does it show that the "wrong" internal cause was thought of as the initial hypothesis? The results are certainly compatible with the following interpretation: People are much more eager not to look like fools (a dispositional factor) than they are to follow the experimenter's instructions to call them as they see them (a situational factor). (Note that to interpret these results as consistent with the FAE, one must count the experimenter's instructions as an internal factor in this experiment, whereas in the Milgram experiment, one must count them as an external factor. As we see, there is a way to construe the internal-external distinction that makes this sensible.) Again, it is not at all clear that the layperson's belief that behavior is internally caused is an error, although it does seem that the layperson latches onto the wrong cause (whether one calls it internal or external).

Was fear of embarrassment one important cause of behavior in the Asch (1952, 1956) experiments? Participants (even the ones who told the truth, Jahoda, 1959) were made quite upset by the experience of finding themselves at odds with the group. One reason they might have been upset is that giving what they perceived to be the right answer must have made either them or the group appear to be inadequate. Someone's face must surely have suffered, and that would have caused embarrassment. Indeed, when Asch (1956) had one confederate give the wrong answer, the group (now of real participants) laughed out loud at the poor fellow—certainly an embarrassing event. If the participants believed that they would be ridiculed by the group if they called it as they saw it, they were right. Fear of embarrassment, then, is one reason people might go along with the group; it might underlie one form of what Deutsch and Gerard (1955) called *normative social influence*.

The Latané and Darley (Darley & Latané, 1968; Latané & Darley 1968) findings on bystander intervention also seem surprising, even shocking. In their experiments, participants were less likely or slower (or both) to take action in an emergency when there were other participants present than when they were alone. This was so both when someone else was at risk in the emergency and when the participants themselves were at risk. Here again a *particular* situational factor, the presence of other people, was more important than the internal factor we thought would have dominated behavior, a concern for their own and others' welfare. We do not disagree that Latané and Darley demonstrated that certain situational factors are more important than anyone thought. But, we wonder: Are the results not also consistent with the view that people are more disposed to avoid making fools of themselves (an internal cause) than we thought they were? Although Ross (1977) claimed that these experiments showed certain situational forces to be more important than had been thought, one could just as well claim that they showed certain dispositional or internal forces (motivations or desires) to be more important than had been thought.

Is fear of embarrassment an important cause (or inhibitor) of behavior in these experiments? We suspect that the reason participants did not respond more to the various provocations was their fear of looking like fools, looking "uncool," a fear we take to be fear of embarrassment, as Miller (1996) suggested, and as Sabini et al. (1999) demonstrated, people may well imagine that they dread embarrassment more than most other people do, and they may therefore underestimate the lengths to which other people would go to avoid embarrassment, even if they understand the lengths to which they themselves would go.⁴ In any event, to respond to smoke as if it signals impending disaster when others present are ignoring it threatens one's own or the others' face; if the others have a good reason for ignoring the smoke, one may look foolish or naive for worrying about it, and if they do not have a good reason, they will look foolish for having ignored it. So, again, fear of embarrassment is at least a candidate for explaining the results.

We turn now to review studies that focus directly on laypeople's mistakes in assessing causes of behavior. We begin with the research tradition referred to as *forced-compliance* studies because, we believe, at the

⁴We suggest that pluralistic ignorance is often fueled by a fear of embarrassment. (See Prentice & Miller, 1993, for a discussion of pluralistic ignorance.)

heart of those studies lies people's misunderstanding of the causes of their own behavior.⁵

Cognitive Dissonance

There are a great number of cognitive dissonance studies in which participants are induced to express attitudes that are contrary to their own for either a small amount of money or an even smaller amount of money.⁶ The participants are then asked to tell someone what their real attitudes are. The typical finding is that at least some participants come to change their views to be more in accord with the attitudes they were induced to express. The degree to which they do this seems to be related to how little they were paid for expressing counterattitudinal views: The less they are paid, the more they change (Festinger & Carlsmith, 1959; Under, Cooper, & Jones 1967).

A standard interpretation of this result is that in the lowest pay conditions, participants see themselves as not having been paid enough to lie, so they instead decide that they said what they said because they believed it (or were close to believing it). On the face of it, it would seem that they were wrong about the pay's not being enough to get them to lie, because they did indeed lie. And, again on the face of it, it is because they were wrong about how cheaply they could be bought that they distorted their beliefs.

At the heart of the matter in these experiments, as Ross (1977) pointed out, is that participants are induced to do something by some agency that they later do not understand as having been sufficient to induce them to do it. In other words, the cognitive dissonance phenomenon, however one comes to understand it, depends on participants' being wrong about a cause of their behavior. As Ross put it, "Why *does* the compliant actor in the 'one dollar' condition experience dissonance?...The answer is clear, Actors...must systematically *underestimate* the sufficiency of the particular complex of situational factors in the Festinger and Carlsmith (1959) study to produce compliance and must *overestimate* the role played by personal dispositions in producing such behavior" (p. 186).

We almost agree with this. We agree that people underestimate the impact⁷ of the particular

complex of situational factors in the study, but we also believe that what they overestimate is the impact of a *particular* internal factor, the desire to be truthful, relative to the desire to do something else. The forced-compliance experiments are indeed evidence that participants are wrong about the causes of their own behavior. But at issue is how we can best characterize their error. To answer this we need to know why they lied, especially why the participants in the \$1 condition lied. What exactly was the particular complex of situational factors that was causally efficacious?

Two hypotheses occur to us. The first is that the participants really did lie for a dollar. The second hypothesis is that the money was a red herring. In the first hypothesis, what participants were wrong about was that they thought they would not lie for a dollar, whereas in fact they would. Such people might be described as being greedier than they thought they were. So in this first hypothesis, it is true that a *particular* internal factor, the desire to tell the truth, was less important to participants than they thought it was. But it is also true that another internal factor, the desire for money, was stronger than they thought it was. Or, one *particular* external factor, how interesting the experiment was, was less important in the complex of situational factors sufficient to cause the behavior than was another external factor, the dollar.

The second hypothesis is, to our minds, more interesting. It is that the money simply deflected participant's attention from the real reason they told the lie, namely, their inability to refuse the experimenters' requests. On this account they were right in their belief that they did not lie for the small amount of money; what they are wrong about was why they really did do it. Along these lines, Kelley (1967) pointed out that people probably engage in the counterattitudinal expression not for the money but because of what he calls *diffuse social pressures* from the experimenters. In other words, the experimenters somehow manipulated the participants into writing the counterattitudinal essays, and they manipulated them in a way that the participants were not aware of, at least not retrospectively. But, again, there are both an internal and an external way of putting this error; the internal way is that participants are less able to refuse certain kinds of requests than they think they are.

In any event, as with the studies mentioned above, the forced-compliance studies do indeed demonstrate that participants are wrong about the causes of behavior. But these studies do not show that behavior in these experiments is externally caused in any sense that competes with the hypothesis that behavior is internally caused. The data are equally compatible with the hypothesis that participants simply identified the wrong internal cause.

⁵ *Forced compliance* is a bit of a misnomer; *manipulated compliance* might be a better term.

⁶ The original Festinger and Carlsmith (1959) study, it is true, offered (but did not pay) \$20, a not-so-small amount of money, especially in the 1950s. More recent studies have offered (but paid) much less money.

⁷ We are reluctant to use the word *sufficiency*. Specific situational factors are indeed sufficient causes given the right internal factors. But it is also true that specific internal factors are sufficient given the right situational causes.

Are the forced-compliance studies also examples of American participants' succumbing to concerns about embarrassment and face? Perhaps. People may do what they did not for the money but to avoid the awkwardness of refusing. But they may not realize that. Perhaps embarrassment avoidance is the real engine of behavior in the forced-compliance studies too,

F AE Studies

We turn now to review studies squarely in the FAE tradition. We ask whether they show an error on people's part, and, if so, what error? We start with the seminal Jones and Harris (1967) study on the attribution of attitudes. In the key condition of this study, the experimenters asked the participants to read an essay which was pro-Castro but which was supposedly written by a student who had been told to write a pro-Castro essay by his or her political science instructor. The participants were asked to guess the writer's true attitude toward Castro. The result was that participants did not entirely ignore the essay in inferring the writer's attitude.⁸

Specifically, the participants inferred that the pro-Castro essay writer's true attitude toward Castro was more positive than they would have inferred it to be had they not seen the essay and known only that the writer was a typical fellow student. In inferring a somewhat positive attitude, the participants are often said to have underestimated the degree to which the writer's behavior was externally (or situationally) caused—by the instruction from their teacher—and to have overestimated the degree to which it was internally (or dispositionally) caused—by the writer's true beliefs. But one could just as well say that the question participants faced was not whether the essay writing was externally or internally caused but, rather, which internal cause was active, the writer's beliefs about Castro or a desire to please the political science instructor. (Or, if you prefer, the question they faced was which external cause was active, Castro's deeds or the political science instructor's orders.) The straightforward interpretation of the results is that participants overestimated the influence of a *particular* internal cause (the writer's beliefs about Castro) compared to the influence of another, equally internal cause (the writer's desire to please the instructor).

We suggest, then, that the error participants made was, as in the other studies we have mentioned, the failure to realize how easy people (or at least students) are to manipulate. Apparently, participants in studies similar to Jones and Harris's (1967) who be-

lieved that the essay writer had a clear ulterior motive to write the essay did discount it completely (Fein, 1996; Fein, Hilton, & Miller, 1990; Hilton, Fein, & Miller, 1993). So the problem people have with the Jones and Harris experiment is not that they fail to take into account all ulterior motives. Rather, they are simply unaware of how certain motives (for example, avoidance of embarrassment) make us easy to manipulate.

The Quizmaster Study

A very famous, clever, and charming study by Ross, Amabile, and Steinmetz (1977) is also taken to be evidence for the FAE; for example, it figures prominently in Gilbert and Malone's (1995) review of the literature. Because it figures so prominently in these discussions, we will review it here, but we should warn in advance that its results actually bear very little on the Really FAE or, for that matter, on the FAE. But, as we shall see, its results, too, can be interpreted as demonstrating participants' concern with face.

In the first experiment of this study, Stanford undergraduates were asked to be either a "questioner" or a "contestant." They drew lots to decide which they would be; assignment was therefore known to the participants to be random. The questioners were to compose challenging but not impossible questions to ask the contestants. The contestants were to answer the questions. The questioners did a good job of making up difficult questions; the contestants got a mean of only 4 out of 10 questions right. The participants were then asked to judge their own and their partners' general knowledge relative to the average Stanford student's on a 100-point scale. In the second experiment, there were also participants who merely observed the contest, each observing one questioner-contestant pair.

The prominent result was that the contestants rated themselves significantly lower on the 100-point general knowledge scale than they rated their partners (41.3 vs. 66.8, $p < .001$). The questioners, however, did not rate themselves significantly higher; they rated themselves 53.5 and their partners 50.6 ($p > .05$). The observers rated the questioners as much more knowledgeable than the average Stanford student (82.1), and they rated the contestants as about as knowledgeable as the average Stanford student (48.9). This result is often said to show that participants overestimated the importance of an internal cause (general knowledge) of what they observed and underestimated the importance of an external cause (the nature of the tasks), once again displaying the FAE. Gilbert and Malone

⁸ One sometimes hears that participants in this experiment did not discount the essay. They did. They just did not discount it entirely.

(1995), for example, interpreted the experiment that way:

Surely contestants were faced with a much more difficult task than were quizmasters, and surely task difficulty was a powerful determinant of their performances. Nonetheless, observers of the game show concluded that the quizmasters were genuinely brighter than the contestants. Because observers could not actually see the "invisible jail" in which contestants were imprisoned, their impoverished understanding of the situation led them to have inappropriate expectations for the contestants' behavior—expectations that could not help but be dashed by reality. (p. 25)

First, we think it important to notice that although it is true that contestants and observers (although not questioners) rated the questioners higher than they rated the contestants, they did this not by derogating the contestants but instead by inflating the knowledge of the questioners. There really is no evidence that the observers' expectations of the contestants were dashed because of a failure to perceive an "invisible jail." Rather, their expectations of the questioners were exceeded. The tendency in discussing this study has been to suggest that observers (and contestants) failed to notice some situational constraint (invisible jail) affecting the contestants, rather than a situational facilitator affecting the questioners. But the results do not support such an interpretation,

Discussions of this experiment also often fail to appreciate how hard it is for all concerned to decide how knowledgeable about general events the questioner and contestant are. Indeed, Ross and Nisbett (1991) suggested that subjects in this study were "blind" (p. 128) to the importance of roles in this study, just as Gilbert and Malone (1995) claimed that participants had an impoverished understanding of the situation. Do the data really show that? Suppose subjects had seen perfectly clearly what Ross and Nisbett claim they were blind to, and suppose they understood the situation perfectly, then how would they have answered the experimenter's question? Would they have gotten the right answer?

Readers of the study know, of course, that since contestant and questioner were assigned randomly, on average the two groups should be roughly equal in their general knowledge. And readers are struck by the fact that the results do not come out that way. But the various participants in this study were not asked about averages. They were asked about a particular pair of people, one of whom was able to make up six questions that the other could not answer. Does that show that the questioner is smarter than the average Stanford student? Does it show that the contestant is less knowledgeable? Do you really have to be blind to the importance of roles or deficient in your understanding

of the situation to infer that the questioner is smarter than the contestant? The key fact the participants needed to know to decide who was smarter than whom was how many questions the average Stanford student can make up that the average other Stanford student cannot answer. Readers know this, but participants cannot. Lacking this critical piece of information, observers and contestants said that contestants were near the average for Stanford students, but that the questioners were smarter than the average student; observers concluded that they were much smarter. Is this evidence of blindness to the importance of roles, or merely evidence of ignorance about how many unanswerable questions the average Stanford student can make up?

Suppose participants in this experiment had realized that they lacked the crucial bit of information. Suppose the participants gave up trying to figure out what they manifestly could not figure out. What would they have done? We suggest that what they might have tried to do is to answer in a way that enhanced, or at least saved, everyone's face and at the same time was compatible with the performances. Is this what they did? We do not know, but the data are consistent with such an interpretation.

First, notice that no one—neither contestants, questioners, nor observers—denigrated the performance of the contestants. All saved the contestants' faces by saying that they were at or very near the Stanford average. Second, the questioners did not brag; they modestly said that they too were just average. The contestants and observers, however, enhanced the face of the questioners by saying that the questioners were smarter than the average Stanford student. We suggest that this result was just what Miss Manners would tell all concerned to produce, given the facts of the experiment. Because it is so socially desirable, a willingness to say that the questioners were above average does not provide good evidence for an error in thinking. A willingness to say that the contestants were below average would have been good evidence of an error in thinking, but that result was not found.

Actor and Observer

We turn now to the final study we consider, a seminal study by Nisbett, Caputo, Legant, and Maracek (1973). (See also Ross & Nisbett, 1991, for a discussion.) Better than any other, this study illustrates the central problem of attribution theory. In this study, some participants (whom the authors called *observer subjects*) saw other participants (*participant subjects*) volunteer to do something of a public-spirited nature for which act they were also paid. The observer subjects were then asked to predict whether the partici-

pant subjects would also volunteer to do another eleemosynary task for free. The participant subjects' willingness to volunteer for free was, in fact, unrelated to their willingness to do something public-spirited for pay, but the observer subjects mistakenly believed that the participant subjects' willingness to volunteer in one case was positively related to their willingness to volunteer in the other. In this experiment, too, participants (the observer subjects) were said to have erred in attributing behavior internally (to an altruistic character) when it was actually controlled externally (by the money). Ross and Nisbett (1991) wrote, "A study by Nisbett, Caputo, Legant, and Maracek (1973) showed that even such an obvious, widely appreciated situational factor as financial incentive can be slighted in explanation and prediction if there is a possibility of explaining behavior in dispositional terms" (p. 126).

But it seems just as correct to say of the results of this experiment that observer subjects attributed behavior to the wrong internal cause (or disposition), namely altruism, when they should have attributed it to the right internal cause, namely the disposition to do things for money. The problem was not that subjects attributed internally when they should have attributed externally, because they could have made correct predictions if they had made the right internal attribution—to a desire for money. Once again, the study reveals that observers are in error about something, certainly, but once again the internal-external distinction does not seem to capture the nature of the error. One can redescribe the results as showing that people err about which internal (or external) causes are important. (Embarrassment is certainly not a candidate explanation for behavior here, nor is it a candidate source of error on the part of observers.)

The Crux of the Issue

Our first claim is that the notion that behavior is really externally caused although people think it is internally caused is not adequate to capture the results we have reviewed. We have tried to show that in each case. We next examine the issue in general terms.

The problem with the FAE is that, as has been often pointed out, there is typically no right answer to the question, "Was this behavior internally or externally caused?"—at least in one sense of internal and external (Gilbert & Malone, 1995; Heider, 1958; Ross, 1977; Sabini & Silver, 1987). Consider these questions: Did Tom eat this piece of candy because he liked sweets or because the candy was sweet? Was his eating the piece of candy internally or externally caused? Typically, both are true. Indeed, it is correct, although awkward,

to say that because he liked sweets, he ate the piece of candy because it was sweet, or, conversely, to say that because the candy was sweet, he ate it because he liked sweets. As Lewin (1935) pointed out, behavior is typically a function of both the person and the environment. To see behavior as internally caused is therefore not typically an error, because behavior usually is internally caused, although it is also externally caused. (See Lupfer, Clark, & Hutcherson, 1990, for evidence that some factors, at least, that increase situational attributions do not decrease dispositional attributions, and vice versa.)

Just a Semantic Trick?

Is the fact that one can always redescribe internal causes as external, and vice versa, just some cheap semantic trick, or does it imply something deeper? We think the answer is that it implies something deeper, something that has to do with how desires and beliefs work. Desires and beliefs are facts of individuals' psychology, but they reach into the world. Consider an addict. Addicts, at least according to popular legend, are driven by their "internal" cravings. Their mental lives are taken over by something like pain. And pain causes us to be terribly internally oriented. But at the same time, addicts' lives are under the control of whatever it is they crave. They are at once controlled by their internal cravings and by their external abused substance. The key point is that the more internally controlled they are, the more externally controlled they are. Where desires are concerned, the more behavior is caused by events under the skin, the more it is caused by things outside the skin, Heider's (1958) metaphor of a boat on the river moved both by the current (external force) and by the occupant's rowing (internal force) is just plain misleading. (See Bierbrauer, 1979, and Miller, Smith, & Uleman, 1981, for a similar criticism of Heider's "hydraulic" metaphor, and see Higgins & Winter, 1993, and Reeder & Brewer, 1979, for discussions of a variety of different models linking situational and dispositional attributions.)

Why is it misleading? It is misleading because the skin, the boundary between internal and external, is not a very important boundary for this discussion. But is the skin the boundary between internal and external that attribution theorists have in mind, or is this a straw man we have created? We quote Gilbert and Malone (1995) in a recent, prominent review of attribution theories and the FAE:

Although these theories differ in both focus and detail, each is grounded in a common metaphor that construes the human skin as a special boundary that separates one set of "causal forces" from another. On the sunny

side of the epidermis are the external or situational forces that press inward on the person, and on the meaty side are the internal or personal forces that exert pressure outward, (p. 21)

But this, we suggest, is the wrong metaphor for how human desires and beliefs interact with the world. Magnets are a better metaphor. Imagine a magnet pulled toward the opposite pole of another magnet, Is the behavior of the second magnet internally or externally caused? The answer is *both*. It is by virtue of its internal magnetic properties that the magnet is attracted to other magnets, and it is by virtue of the presence of the other magnet that this attraction is actualized.

The Individual Differences Conception of Internality

Ross early on (1977) recognized the problem of considering the skin an important boundary. Indeed, he criticized Nisbett for accepting the form of an *attributional utterance* (meaning whether it mentioned something inside or outside the skin) as a way of telling whether it is internal or external. That is, in the Nisbett et al. (1973) article discussed previously, some participants were asked to indicate why they themselves were dating a particular person and others were asked why their best friend was dating a particular person. The experimenters coded the accounts given as internal or external according to what Ross called their *form*, that is, whether the accounts mentioned some psychological fact about the person the participants were asked to write about or whether they mentioned some fact about the woman the target was dating. The results were that the ratio of internal to external attributions was higher for explanations of best friends' behavior than for the participants' own behavior. Thus, participants tended to say things such as "I am dating the woman I am dating because she is beautiful" (an external attribution), but they would say that their best friend was dating the woman he was dating because he liked beautiful women (an internal attribution).

Ross (1977) pointed out that these are not competing accounts. In fact, they entail one another; unless Betty is beautiful, then Tom's liking beautiful women is not an account of his liking her. That is, Tom's liking beautiful women is an explanation of his behavior if and only if Betty is beautiful. And her being beautiful is an explanation of his behavior if and only if he likes beautiful women. But Ross did not give up on the internal-external distinction; rather, he made two suggestions. First, he proposed that the gold standard for whether subjects were making an attributional error was prediction. After

all, the observer subjects in the Nisbett et al. (1973) study were asked to predict what the participant subjects would do, and their predictions were wrong. They did, in fact, make an error. Ross therefore suggested that predictions about what people should do rather than characterizations of what they have done should be the gold standard for attributional research. Second, contrary to Gilbert and Malone's (1995) claim about the skin being the basis of internal-external distinctions, Ross proposed a very different basis for the distinction. Following Jones and Davis (1965), he suggested that what distinguishes dispositional from situational causes may be this: Behavior is situationally caused if it is what most people would do in a situation; behavior is dispositionally caused if it is unique or, at least, rare.

With regard to prediction as the gold standard, we agree that predictions can be shown to be correct or incorrect and, therefore, can serve as a standard for deciding whether an error has been made. But predictions are mute about how the error should be characterized. The issue we are raising about the Nisbett et al. (1973) study is not, after all, whether observer subjects were right in predicting what participant subjects would do; the issue is whether the observer subjects were wrong because they attributed the paid, public-spirited behavior to internal rather than external causes or because they attributed it to generosity rather than to a desire for money; either sort of error would have produced the results.

As to the individual differences interpretation of the distinction, consider Tom, who ate a piece of candy; does that mean he has a sweet tooth? Ross (1977) would argue that, no, it does not—not by itself. Having a sweet tooth is more than just occasionally enjoying a piece of candy. It entails eating more candy, or enjoying it more, or wanting to eat it more, or something like that, than the average person. Ross suggested that explanations that invoke "... a widely accepted and generally applicable S-R law" are situational, and that explanations that resort to "... an individual difference or distinguishing personality variable" are dispositional (p. 177).

On this individual difference conception of dispositional (internal) versus situational (external) causality, if someone does something they very much want to do, something they do wholeheartedly—say, eat a bowlful of chocolate ice cream—with gusto and enthusiasm, then whether that behavior was internally or externally caused depends on what other people do! If other people also love chocolate ice cream, then the behavior was situationally caused; if other people do not, then it was dispositionally caused. Ross (1977) did not merely suggest that observers would perceive behavior, such as your relishing of a new Ben and Jerry's flavor, as dispositionally caused if you were the only person who liked it and as

situationally caused if everyone liked it (cf. Kelley's, 1967, version of attribution theory)—he suggested that this is the right answer. For Ross, then, behaviors with low variance (across people) were situationally caused, whereas behaviors with high variance were dispositionally caused.

Consider in this light Gilbert and Malone's (1995) account of what they call *correspondence bias*, but we would call the FAE: "When people observe behavior, they often conclude that the person who performed the behavior was predisposed to do so—that the person's behavior corresponds to the person's unique dispositions—and they draw such conclusions even when a logical analysis suggests they should not" (p. 21). Notice that the definition of *correspondence* given outside the dashes and the one inside the dashes are not the same. If we decide that someone is eating the chocolate ice cream because he loves it (is disposed to eat it), then according to the outside-the-dash definition, we have made a correspondent inference, but according to the inside-the-dash definition, we have only done so if we also believe that chocolate ice cream is unpopular. Insufficient attention has been paid to the distinction between these distinctions.

Individual Differences and Social Influence

Based on the "internal causes are those under the skin" interpretation, the FAE was the "error" of imagining that a behavior was caused by something internal when it was really caused by something external. What is the FAE based on the individual differences interpretation? Is the evidence from social influence studies, dissonance studies, and FAE studies consistent with this version of the internal-external distinction?

Attribution theorists have not, to our knowledge, operationalized the FAE based on the individual difference view of internal causality, so we must do it for them—in terms of variance. Based on this interpretation, to make the FAE in a specific situation is to imagine that there is more variance in behavior in that situation than there actually is. We must now ask: Is this the error people make in the social influence studies discussed above, the studies typically cited in support of the importance of the FAE?

As has been argued elsewhere (Sabini and Silver, 1983), the answer is no. Consider, for example, the Milgram (1974) experiment. People asked to predict what participants would do in the Milgram experiment predicted very low variance; they expected everyone to do the same thing in that experiment, that is, to disobey. But instead, 65% obeyed and 35% disobeyed. In other words, the experiment revealed

much more, rather than much less, variance in behavior than people expected. On Ross's (1977) account of what it is for behavior to be situationally controlled, the one account that Gilbert (1998) suggested does not produce a vicious dualism, the Milgram experiment revealed that behavior is substantially more dispositionally controlled than we had thought. And this argument applies *a fortiori* to the Asch (1952, 1956) and Latané and Darley (Darley & Latané, 1968; Latané & Darley, 1968) studies. Thus, if we accept Ross's conception of what it is for behavior to be dispositionally controlled versus situationally controlled, we certainly must give up the idea that people think that it is dispositionally controlled when in actuality it is situationally controlled. The purely statistical conception makes the distinction intelligible, but it makes it so in a direction opposite to our intuitions. We have, then, two senses of the internal-external distinction before us, both of which have been used over and over by attribution theorists. One treats the skin as the important boundary. The problem with this distinction is that in using it, one must agree that behavior that is internally caused is also externally caused; his liking her because he likes beautiful women entails that she is beautiful. So, based on this account, although some causes are internal and some are external, all behavior is a function of both. The second way to make the internal-external distinction relies on individual differences. It classifies common behaviors as externally caused and distinctive behaviors as internally caused. One problem with this way of making the distinction is that it classifies what happens in the social influence experiments as examples of internally caused behavior and classifies what people expect to happen as externally caused—the reverse of the error people are accused of making.

A Third Conception of Internal and External

There is a third way of characterizing the FAE. Ross and Nisbett (1991) wrote of the Jones and Harris (1967) experiment: "The study indicates that observers are too willing to take behavior *at face value* [italics added], as reflecting a stable disposition (in this case an attitudinal disposition), even when it is made abundantly clear that the actor's behavior is under severe external constraints" (p. 126). We will make two points about this.

First, note that Ross and Nisbett (1991) were here relying on the an under-the-skin versus an above-the-skin interpretation of the FAE. We infer this because Ross and Nisbett seemed not to see subjects who attributed anti-Castro attitudes to writers

who freely wrote anti-Castro essays as succumbing to the FAE, whereas based on their individual differences account of the FAE, they should have! Because virtually everyone wrote an anti-Castro essay in that condition, the individual differences account classifies the behavior as situationally controlled; therefore, to attribute the behavior to the actors' dispositions is to succumb to the FAE. The problem is that the statistical conception of the internal-external distinction makes no distinction between two aspects of the situation: that the essay was to be about Castro, and that the essay was written under the instruction of the political science teacher. It treats both as situational causes.

Second, this interpretation relies on our having an analysis of what "at face value" means. Gilbert and Malone's (1995) under-the-skin versus above-the-skin distinction is, if nothing else, clear. How do we make the "at face value" conception clear? Certainly, in the case of the Jones and Harris (1967) study, it is easy enough to see that attributing the attitude expressed in the essay to the writer of the essay is taking the behavior "at face value," but what is the "face value" of behavior in the Milgram (1974) experiment? Is it obeying the experiment? Torturing the victim? Avoiding confrontation? Acting in a cowardly way? In any event, much work would need to be done to turn the probably quite fruitful suggestion that people are overly eager to interpret action at face value into a real theory.

A Fourth Conception of the Internal-External Distinction

The crux of the problem with the internal-external distinction is that none of these conceptions is the one that grounds our intuitions that behavior in the social influence studies is externally controlled whereas people expect it to be internally controlled. There is a distinction that will do the trick, but it is a subtle one, and it calls cognitive processing accounts of the FAE into question. This distinction is that behavior is internally caused if and only if it follows from a person's values and (correct) beliefs (see Sabini & Silver, 1987)⁹. The crucial boundary is not where Gilbert and Malone (1995) put it—at the epidermis. It is at the self. In other words, the most important distinction that we draw using the

internal-external distinction is one between causes of behavior people affirm as part of themselves and causes they reject, between *ego-syntonic* and *ego-dystonic* causes. Does this conception of the distinction support our intuitions that behavior in the social influence experiments is externally caused?

If we accept the ego-syntonic-ego-dystonic distinction, what happens to our interpretation of the Milgram (1974), Asch (1952, 1956), and Latané and Darley (Darley & Latané, 1968; Latané & Darley, 1968) studies? Based on this conception of the internal-external distinction, the classic studies of social psychology are all well seen as examples of behavior under external control. It seems right to believe that participants in all of these studies do not want to be the sort of people who torture innocent victims, conform to a group's response even when it is obviously wrong, or fail to respond promptly in an emergency when others' well-being (or even their own) is on the line. And we see now why the experimenter's instruction is an internal factor in the Asch (1952) study but an external one in the Milgram study. It is one thing to obey the instruction to report which line is longer; it is quite another to shock an innocent victim gratuitously. And, finally, we see why although everyone believes that everyone would write anti-Castro essays in the anti-Castro, free-choice condition of the Jones and Harris (1967) experiment, people are still right in imagining that (if such essays had really been written) they reflect stable, internal dispositions. Insofar as people believe that the expression of an attitude is consistent with what the person expressing the attitude really believes, then the expression is seen as internally caused, no matter how many people share it.

The strong sense of external control that pervades social influence studies follows neither from the fact that there is less variability in behavior in these situations than people expect (as Ross, 1977, would have it) nor from the fact that behavior is caused by events outside the epidermis (as Gilbert and Malone, 1995, would have it), but from the fact that people do things that they know they should not do.

Based on our conception of the internal-external distinction, what matters is not whether behavior is consistent with a person's dispositions, but whether it is consistent with the person's *regrettable* dispositions. If so, then although the behavior is consistent with dispositions, it is external—not to the person, but to the person's self, regardless of how many other people do it.

What about the chocolate ice cream eater? Does this conception cause paradoxes to break out there? No. The chocolate ice cream eater's eating is internally caused unless he is, say, dieting; then it be-

⁹ Our proposal about how to understand the internal-external distinction is closest to White's (1991) analysis. He proposed replacing the internal-external distinction with distinctions between the conscious and the unconscious, the intentional and unintentional, and things done for a reason and things done for no reason. The obedient participants in Milgram's (1974) experiment, however, did what they did intentionally, consciously, and for a reason. The "externality" of their behavior is elsewhere.

comes externally caused, And the more seriously he is dieting, the more external is the causation. The addict's behavior is externally caused—conforming to our intuitions—not because his cravings are external to either his skin or his consciousness, or, heavens knows, his dispositions—but because he rejects them as part of himself. They are foreign to his sense of self, certainly not because he lacks a disposition to take whatever he is addicted to, but because his desire, his disposition, is itself external to who he would be.

And, we suggest, the FAE studies in the Jones and Harris (1967) tradition are also connected to this sense of external control. We agree that in this sense of internal versus external, participants expect behavior to be internally caused rather than externally caused. That is, participants do not really approve of students' writing essays contrary to their important political beliefs. They expect people, or at least some people, to refuse. And that is enough to generate the Jones and Harris phenomenon. If we accept this interpretation of the internal-external distinction, then the attribution literature converges on this story: Social psychology since the Second World War has discovered that Americans' behavior is less in line with what they themselves value than they think it is, Americans are less well behaved than they think they are—less well behaved by their own lights.

But this is too broad and vague a characterization of what has really been found. After all, there are lots of ego-dystonic behaviors; the literature cited in the context of the FAE is not about excessive drinking, drug abuse, procrastination, or marital infidelity—all ego-dystonic desires. It is about a much more specific kind of failing, a much more specific kind of misbehavior.

Fear of Embarrassment as an Ego-dystonic Motive

We further suggest that all the studies we reviewed, except the Nisbett et al. (1973) and Ross et al. (1977) studies, might have involved participants' doing something (or failing to do something) out of fear of embarrassment, or of threat to their own or another's face. That is what is so interesting about these studies. They are studies in which people did things that we (and they) see as wrong, in one sense or another. But they did these wrong things without temptation from what we know tempts people; sex, food, drugs, and so on. Rather, they did wrong things for reasons we (and they) find obscure. The obscure reasons may have to do with protecting face and avoiding embarrassment.

Implications

Why do people go wrong in explaining behavior? Consistent with the cognitive revolution in psychology, social psychologists have sought strictly cognitive accounts of these errors (see Gilbert & Malone, 1995, for a review). The essence of these accounts is that people have been looking for causes in all the wrong places—internally rather than in the situation. But our account says that people do not ignore all situational causes; rather, they err because they think of the wrong aspects of situations. Saying "Look outside not inside!" will not help correct people's thinking unless you tell them what aspect of outside to look at. If we are right that people do not ignore situations tout court but rather underestimate the impact of certain aspects of situations—aspects that Ross and Nisbett (1991), following Lewin, call channel factors—then it is hard to imagine a purely cognitive account. But it is not hard to imagine that people go wrong because they have learned and absorbed a "rugged individualist" view of what people, or at least North Americans, are like.

There is further reason to believe that the view that people ignore the situation and focus on behavior cannot be right. The problem is that characterizing behavior relies on aspects of situations. Some, although surely not all, of the characterizations that are true of yelling "fire" in a crowded theater depend for their veracity on the detail of whether there is or is not a fire in the theater—and that is a fact about the situation. Typically, to decide what behavior just happened, one must put information about the muscle movements of the behaving person together with information about the situation. Therefore, whereas one might have the view that behavior engulfs some aspects of the field—perhaps motives—it cannot engulf the whole field and remain what we would call *behavior*.

Culture and Face

It has been argued that the tendency to make the FAE is a Western phenomenon. The evidence for this (Miller, 1984, 1986, 1987; Shweder & Bourne, 1982) relies too much, to our minds, on what Ross (1977) called *formal criteria* to be entirely convincing. But for the sake of argument, let us assume that there is more external attribution in other cultures. What would this mean?

It might mean that people's cognitive and even perceptual systems function differently here and there, that culture really does affect psychology at the deepest levels. But it might not. As has often been pointed out (e.g., by Nisbett & Wilson, 1977), the

search for causes begins in a theory. Insofar as Westerners have the theory that as rugged individualists they care not about saving face, avoiding embarrassment, or social harmony, then even if they *do* in fact care about these things, they will not attribute their behavior to these motives. According to this account, the reason Westerners miss some of the causes of their behavior where people in other cultures might not is that although everyone everywhere is concerned with face, embarrassment, and harmony, Americans tend not to notice this about themselves. Indeed, it is precisely to call attention to this that we wrote this article.

Summary

We argue that in the view of many social psychologists, one of the important discoveries of the second half of this century is that Americans make internal attributions for behavior more often than they should. Many social psychologists, we believe, would argue that the tendency to make this error flows from some sort of cognitive defect. Exactly what cognitive defect is responsible for the error is a matter of active investigation; Gilbert and Malone (1995), for example, proposed four candidates. These defects are held to explain why we make this sort of categorical error.

But, we argue, attribution theorists have been neither clear nor consistent in how they interpret the internal-external (disposition-situation) distinction. For example, Gilbert and Malone (1995) interpreted the distinction as (a) inside or outside the skin, (b) consistent with a person's dispositions or not, and (c) a matter of individual differences or not. But even the simplest behavior is differently classified by these various criteria: Eating a bowl of chocolate ice cream because you really like it is both internally and externally caused on account, "a"; internally caused on account, "b"; and externally caused on account, assuming that most people like chocolate ice cream, "c". Worse, the very behaviors that make the internal-external distinction important are differently classified by the three criteria. Obedience in the Milgram (1974) experiment is caused by events both inside and outside the skin. It is consistent with some of the participants' dispositions: the desire to do what the person in charge says, the desire to avoid embarrassment, and so on. But it is inconsistent with other dispositions: the desire to do the right thing, the desire to do the moral thing, the desire not to hurt someone. It is internally caused on an individual difference view of the distinction, because there is quite a lot of variance in behavior.

The problem is that neither of these interpretations grounds the sharp intuition that behavior in the Milgram (1974) experiment is externally caused. We suggest that what grounds this intuition is not the fact that participants act against all of their dispositions, but rather the sense that participants act in conflict with the dispositions they affirm. Behavior in the Milgram experiment strikes people as externally caused because participants were manipulated into doing things they would rather not have done. It is not that they did things contrary to their dispositions, their wants—what are often called their first order desires—but that they did things contrary to their second order desires, contrary to the selves they wished they were. These experiments show us that people are not as strong as we think they are.

We believe that the so-called FAE experiments (Jones and Harris, 1967; Ross et al., 1977) show this too. Given that participants seemed not to make the FAE when the possibility of ulterior motive arose but they did make it when the essay writer had been pressured to write, a plausible interpretation is that people do not think they are "better" than they are, just "stronger" than they are. We suggest that what lies at the heart of both the social influence and FAE phenomena is the fact that people underestimate the power of a specific aspect of situations (or a particular sort of disposition); they underestimate the power of the demand to avoid what Brown and Levinson (1978), following Goffman (1955), called *face-threatening acts*. In urging social psychologists to abandon the confusing, overly broad metaphor of internal versus external causality and the misleading picture of the person rowing against the tide, we echo Gilbert and Malone (1995). They decried the tendency to expand the notion of "situational constraint," and they argued that social psychologists are all too eager to invoke that notion, but that by using it promiscuously they may find that it loses all meaning. We agree with them about this. And although we disagree with them that the quizmaster study is an example of participants going wrong by missing the operation of "invisible situational constraints," we believe that the other studies we have cited here are examples of people's missing invisible situational constraints. And we suggest that being invited to write an essay about Castro in whichever direction one pleases and being told to write a negative one are not the same thing, even though both result in unanimously negative essays.

But, one might ask, when all is said and done, is it not true that the burden of this research is to show that behavior is externally controlled when we thought it was internally controlled? In the end we must agree with this, but only if one interprets *internal* to mean something very specific, that is,

to mean in line with the person's second order desires, and *external* to mean something equally specific, that is, to mean contrary to (not neutral with regard to) those second order desires. But even this is too broad, because social psychologists are not, by and large, interested in every sort of ego-dystonic behavior, just in those that arise from social influence—the kind of social influence connected to embarrassment and face work.

Notes

We thank Lee Ross and anonymous reviewers for their comments on an earlier version of this article,

John Sabini, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104-6196. E-mail: sabini@psych.upenn.edu

References

- Asch, S. E. (1952). *Social psychology*. New York: Prentice Hall
- Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, 70, 416.
- Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology*, 9, 67-84.
- Brown, P., & Levinson, S. C. (1978). *Politeness: Some universals in language use*. New York: Cambridge University Press.
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influence upon individual judgment. *Journal of Abnormal and Social Psychology*, 51, 629-636.
- Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, 70, 1164-1184.
- Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motive and the correspondence bias. *Journal of Personality and Social Psychology*, 58, 753-764.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203-210.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 89-150). Boston: McGraw-Hill.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21-38.
- Goffman, E. (1955). On facework. *Psychiatry*, 18, 213-231.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Higgins, E. T., & Winter, L. (1993). The acquisition principle: How beliefs about a behavior's prolonged circumstances influence correspondent inference. *Personality and Social Psychology Bulletin*, 19, 605-619.
- Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin*, 19, 501-512.
- Jahoda, M. (1959). Conformity and obedience. *Human Relations*, 12, 99-120.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, 34, 107-117.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219-266). New York: Academic.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3, 1-24.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium of motivation* (Vol. 15, pp. 192-240). Lincoln: University of Nebraska Press.
- Kruglanski, A. W. (1975). The endogenous-exogenous partition in attribution theory. *Psychological Review*, 82, 387-406.
- Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, 10, 215-221.
- Lewin, K. (1935). *Dynamic theory of personality*. New York: McGraw-Hill.
- Linder, D. E., Cooper, J., & Jones, E. E. (1967). Decision freedom as a determinant of the role of incentive magnitude in attitude change. *Journal of Personality and Social Psychology*, 6, 245-254.
- Locke, D., & Pennington, D. (1982). Reasons and other causes: Their role in the attribution process. *Journal of Personality and Social Psychology*, 42, 212-223.
- Lupfer, M. B., Clark, L. F., & Hutcherson, H. W. (1990). Impact of context on spontaneous trait and situational attributions. *Journal of Personality and Social Psychology*, 58, 239-249.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.
- Miller, F. D., Smith, E. R., & Uleman, J. (1981). Measurement and interpretation of situational and dispositional attributions. *Journal of Experimental Social Psychology*, 17, 80-95.
- Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, 46, 961-978.
- Miller, J. G. (1986). Early cross-cultural commonalities in social explanations. *Developmental Psychology*, 22, 514-520.
- Miller, J. G. (1987). Cultural influences on the development of conceptual differentiation in person description. *British Journal of Developmental Psychology*, 5, 309-319.
- Miller, R. (1996). *Embarrassment: Poise and peril in everyday life*. New York: Guilford.
- Monson, T. C., & Snyder, M. (1977). Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology*, 13, 89-111.
- Nisbett, R. E., Caputo, C., Legant, P., & Maracek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, 27, 154-164.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Parrott, W. G., & Harré, R. (1996). Embarrassment and the threat to character. In R. Harré & W. G. Parrott (Eds.), *The emotions: Social, cultural and biological dimensions* (pp. 39-56). Thousand Oaks, CA: Sage.
- Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology*, 64, 243-256.
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, 86, 61-79.
- Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173-220). New York: Academic.
- Ross, L., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, 35, 485-494.

REALLY FUNDAMENTAL ATTRIBUTION ERROR

- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. Philadelphia: Temple University Press.
- Sabini, J. (1995). *Social psychology* (2nd ed.). New York: Norton.
- Sabini, J., Cosmas, K., Siepmann, M., & Stein, J. (1999). Underestimates and truly false consensus effects in estimates of embarrassment and other emotions. *Basic and Applied Social Psychology*, 21, 223-241.
- Sabini, J., Siepmann, M., Stein, J., & Meyerowitz, M. (2000). Who is embarrassed by what? *Cognition and Emotion*, 14, 213-240.
- Sabini, J., & Silver, M. (1982). *Moralities of everyday life*. New York: Oxford University Press.
- Sabini, J., & Silver, M. (1983). The fundamental attributional error and the naive understanding of the Milgram obedience experiments. *Journal for the Theory of Social Behaviour*, 13, 147-152.
- Sabini, J., & Silver, M. (1987). Internal and external causes of behavior. *International Journal of Moral and Social Studies*, 2, 11-23.
- Shweder, R. A., & Bourne, E. J. (1982). Does the concept of person vary cross-culturally? In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self and emotion*, (pp. 158-199). New York: Cambridge University Press.
- Solomon, S. (1978). Measuring dispositional and situational attributions. *Personality and Social Psychology Bulletin*, 4, 589-594.
- White, P. A. (1991). Ambiguity in the internal/external distinction in causal attribution. *Journal of Experimental Social Psychology*, 27, 259-270.