

9. Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments

Lee Ross and Craig A. Anderson



Introduction to attribution theory and attributional errors

Attribution theory and intuitive psychology

Attribution theory, in its broadest sense, is concerned with the attempts of ordinary people to understand the causes and implications of the events they witness. It deals with the “naive psychology” of people as they interpret their own behavior and the actions of others. The current ascendancy of attribution theory in social psychology thus culminates a long struggle to upgrade that discipline’s conception of man. No longer the stimulus–reponse (S–R) automaton of radical behaviorism, promoted beyond the rank of information processor and cognitive consistency seeker, psychological man has at last been awarded a status equal to that of the scientist who investigates him. For in the perspective of attribution theory, people are intuitive psychologists who seek to explain behavior and to draw inferences about actors and about their social environments.

To better understand the perceptions and actions of this intuitive scientist we must explore his methods. First, like the academic psychologist, he is guided by a number of implicit assumptions about human nature and human behavior – for example, that the pursuit of pleasure and the avoidance of pain are ubiquitous and powerful human motives, or that conformity to the wishes and expectations of one’s peers is less exceptional and less demanding of further interpretation than non-conformity. The lay psychologist, like the professional one, also relies heavily upon data, albeit data that rarely satisfy formal requirements regarding randomness

This chapter draws heavily, both in content and organization, from a contribution by the first author to *Advances in Experimental Social Psychology* (1977). Permission for use of these materials is gratefully acknowledged.

or representativeness. Sometimes these data result from first-hand experiences; more often, they are the product of informal social communication, mass media, or other indirect sources. The intuitive psychologist must further adopt or develop techniques for coding, storing, and retrieving the data. Finally, he must employ various strategies for summarizing, analyzing, and interpreting the data - that is, he must employ rules, heuristics, or schemas that permit him to form new inferences. The intuitive scientist's ability to master his social environment, accordingly, will depend upon the accuracy and adequacy of his hypotheses, evidence, and analyses. Conversely, any systematic errors in existing theories, biases in available data, or inadequacies in methods of analysis, yield serious consequences - both for the lay psychologist and for the society that he builds and perpetuates. These shortcomings, explored from the vantage point of contemporary attribution theory, provide the focus of this chapter.¹

The broad outlines of attribution theory were first sketched by Heider (1944, 1958) and developed in greater detail by Jones and Davis (1965), Kelley (1967, 1971, 1973), and their associates (see Jones et al., 1971; Weiner, 1974). These theorists dealt with two closely related tasks confronting the social observer. The first task is that of causal judgment: The observer seeks to identify the cause, or set of causes, to which some particular effect (i.e., some action or outcome) may most reasonably be attributed. The second task is that of social inference: The observer of an episode forms inferences about the *attributes* or dispositions of the relevant actors and about the attributes or properties of the situations to which they have responded.

Causal judgment and social inference tasks have both been the subject of intensive theoretical and empirical inquiry and, until recently, had constituted virtually the entire domain of attribution theory. Lately, however, a third task of the intuitive psychologist has begun to receive some attention; that task is the *prediction* or *estimation* of outcomes and behavior. The intuitive psychologist not only must seek explanations and make dispositional inferences; he must also form expectations and make guesses about actions and outcomes that are currently unknown or that will occur in the future. For instance, when a presidential candidate promises to "ease the burden of the average taxpayer," we consider possible causes for the statement and implications about the candidate's personal dispositions. (Did the promise simply reflect the demands of political expediency? Can we conclude anything about the candidate's true convictions?) But we are also likely to speculate about his subsequent behavior and his views on related issues that have not yet been explored. (If elected, will he slash property taxes? Does he favor curtailment of social welfare programs?) The

¹ For a more thorough and systematic explanation of the layman/scientist parallel, the reader is referred to Nisbett and Ross, 1980.

psychology of intuitive prediction, in short, is a natural extension of attribution theory's domain.

Logically and psychologically, of course, the three attribution tasks are interdependent. Explanations for an event, and inferences about the actors and entities that figure in that event, are intimately related. And together they provide the basis for speculation about the nature of events that are currently unknown or are likely to unfold in the future. Each task, however, offers unique possibilities (and unique problems of interpretation and methodology; see Ross, 1977, pp. 175-179) for revealing the assumptions and strategies that underlie the intuitive scientist's performance. It is worth noting that in recent years the use of estimations and predictions as dependent variables in studies of lay inference has become increasingly popular. One reason for this increased popularity is particularly important. Unlike the causal judgments of dispositional inferences that follow from a perceiver's analysis of an event, estimations or predictions about new or unknown events can often be evaluated with respect to their *accuracy*. That is, one can compare predictions and estimates about events with actual observations or measurements. This permits assessment both of the relative adequacy of the intuitive scientist's attributional strategy and of the direction of specific errors or biases.

Logical attributional principles vs. self-serving biases

Contemporary attribution theory has pursued two distinct but complementary goals. One goal has been to demonstrate that social perceivers' assessments and inferences generally follow the dictates of some logical or rational model. The other goal has been to illustrate and explicate the sources of bias or error that distort those generally veridical assessments and inferences. We shall consider briefly the so-called logical, or rational, schemata employed by the intuitive psychologist and then devote the remainder of the chapter to the sources of error in his attempts at understanding, predicting, and controlling the events that unfold around him.

The "covariation" and "discounting" principles. Individuals must, for the most part, share a common understanding of the social actions and outcomes that affect them, for without such consensus, social interaction would be chaotic, unpredictable, and beyond the control of the participants. Introspection by attribution theorists, buttressed by some laboratory evidence, has led to the postulation of a set of "rules" that may generally be employed in the interpretation of behaviors and outcomes. These "common sense" rules or schemata are analogous, in some respects, to the more formal rules and procedures that social scientists and statisticians follow in their analysis and interpretation of data.

H. H. Kelley, E. E. Jones, and their associates have distinguished two

cases in which logical rules, or schemata, may be applied. In the *multiple* observation case the attributer has access to behavioral data that might be represented as rows or columns of an Actor \times Object \times Situation (or Instance) response matrix. Typically, in this domain of research summary statements are provided to the participants rather than actual responses. Thus the potential attributer learns that "Most theatergoers like the new Pinter play," or "Mary can't resist stray animals," or "The only television program that Ann watches is Masterpiece Theatre." In the *single* observation case the attributer must deal with the behavior of a single actor on a single occasion. For instance, he may see Sam comply with an experimenter's request to deliver a painful shock to a peer, or he may learn that "Louie bet all his money on a long shot at Pimlico."

The logical rules or principles governing attributions in these two cases are rather different (Kelley, 1967, 1971, 1973). In the multiple observation case the attributer applies the "covariance principle"; that is, he assesses the degree to which observed behaviors or outcomes occur in the presence, but fail to occur in the absence, of each causal candidate under consideration. Accordingly, the attributer concludes that the new Pinter play is a good one to the extent that it is liked by a wide variety of playgoers, that it is liked by individuals who praise few plays (e.g., "critics"), and that it is applauded as vigorously on the ninetieth day of its run as on the ninth.

In the single observation case the attributer's assessment strategy involves the application of the "discounting principle," by which the social observer "discounts" the role of any causal candidate in explaining an event to the extent that the other plausible causes or determinants can be identified. This attributional principle can be restated in terms of social inferences rather than causal attributions: To the extent that situational or external factors constitute a "sufficient" explanation for an event, that event is attributed to the situation and no inference logically can be made (and, presumably, no inference empirically is made) about the dispositions of the actor. Conversely, to the extent that an act or outcome seems to occur in spite of and not because of attendant situational forces, the relevant event is attributed to the actor and a "correspondent inference" (Jones & Davis, 1965) is made - that is, the attributer infers the existence and influence of some trait, ability, intention, feeling, or other disposition that could account for the actor's action or outcome. Thus, we resist the conclusion that Louie's long-shot plunge at Pimlico was reflective of his stable personal attributes to the extent that such factors as a hot tip, a desperate financial crisis, or seven pre-wager martinis could be cited. On the other hand, we judge Louie to be an inveterate long-shot player if we learn that his wager occurred in the face of his wife's threat to leave him if he ever lost his paycheck at the track again, his knowledge that he would not be able to pay the rent if he lost, and a track expert's overheard remark that the favorite in the race is "even better than the track odds suggest."

It is worth noting that the application of these two different principles places rather different demands upon the intuitive scientist. The covariance principle requires the attributer to apply rules that are essentially logical or statistical in nature and demands no further insight about the characteristics of the entities in question. Application of the discounting principle, by contrast, demands considerable insight about the nature of man and the impact of such situational forces as financial need, alcohol consumption, and a spouse's threat of abandonment. In a sense, the covariance principle can be applied by a mere "statistician," whereas the discounting principle requires a "psychologist" able to assess the role of various social pressures and situational forces and even to distinguish intended acts and outcomes from unintended ones (cf. Jones & Davis, 1965).

Evidence concerning the systematic use of common-sense attributional principles comes primarily from questionnaire studies in which subjects read and interpret brief anecdotes about the responses of one or more actors to specified objects or "entities" under specified circumstances (e.g., L. Z. McArthur, 1972, 1976). Occasional studies of narrower scope have exposed the attributer to seemingly authentic responses, encounters, and outcomes (e.g., Jones, Davis, & Gergen, 1961; Jones & DeCharms, 1957; Jones & Harris, 1967; Strickland, 1958; Thibaut & Riecken, 1955). Such research has demonstrated that attributers can, and generally do, make at least some use of the hypothesized principles or rules of thumb. What the methodologies employed to date have left ambiguous is the *degree* of the layperson's accuracy and the magnitude and direction of his errors.

Self-serving motivational biases in attribution. In speculating about possible distortions in an otherwise logical attribution system, theorists were quick to postulate "ego-defensive" biases through which attributers maintained or enhanced their general self-esteem or positive opinion of their specific dispositions and abilities (Heider, 1958; Jones & Davis, 1965; Kelley, 1967). Attempts to prove the existence of such a motivational bias have generally involved demonstrations of asymmetry in the attribution of positive and negative outcomes - specifically, a tendency for actors to attribute "successes" to their own efforts, abilities, or dispositions while attributing "failure" to luck, task difficulty, or other external factors. Achievement tasks (e.g., Davis & Davis, 1972; Feather, 1969; Fitch, 1970; Wolosin, Sherman, & Till, 1973) and teaching performances (e.g., Beckman, 1970; Freize & Weiner, 1971; Johnson, Feigenbaum, & Weiby, 1964) have provided most of the evidence for this asymmetry. It has also been shown that actors may give themselves more credit for success and less blame for failure than do observers evaluating the same outcomes (Beckman, 1970; Gross, 1966; Polefka, 1965).

Critics skeptical of broad motivational biases, however, have experienced little difficulty in mounting challenges to such research (see Miller

& Ross, 1975, also Nisbett & Ross, 1980, Chap. 10, for detailed discussions). The most telling argument against research purporting to show motivational biases is the obvious distinction between subjects' private perceptions and their public judgments. One can easily create situations where a person will publicly deny (or claim) responsibility for an event that he privately accepts (or does not accept) as his responsibility. While these public judgments may be self-serving in the sense of preserving one's public image, they do not imply the operation of ego-defensive biases in the sense of preserving one's private image (Miller, 1978).

Furthermore, asymmetries in the private attributions (were they available to researchers) of success and failure, and differences in the judgments of actors and observers may reflect other non-motivational sources of bias. As several researchers have noted, success, at least in test situations, is likely to be anticipated and congruent with the actor's past experience, whereas failure may be unanticipated and unusual. Similarly, successful outcomes are intended and are the object of plans and actions by the actor, whereas failures are unintended events that occur in spite of the actor's plans and efforts. Observers, furthermore, rarely are fully aware of the past experiences or present expectations and intentions of the actors whose outcomes they witness.

Challenges to the existence of pervasive ego-defensive biases have been empirical as well as conceptual. Thus, in some studies subjects seem to show "counterdefensive," or esteem-attenuating, biases. For example, Ross, Bierbrauer, and Polly (1974), using an unusually authentic instructor-learner paradigm, found that instructors rated their own performances and abilities as more important determinants of failure than of success. Conversely, the instructors rated their learner's efforts and abilities as more critical determinants of success than failure. In the same study these seemingly counterdefensive attributional tendencies proved to be even more pronounced among professional teachers than among inexperienced undergraduates, a result that contradicted the obvious derivation from ego-defensiveness theory that those most directly threatened by the failure experience would be most defensive.

Researchers who insist that self-serving motivational biases exist can, of course, provide alternative interpretations of studies that seem to show no motivational biases or counterdefensive biases (cf. Bradley, 1978). Indeed, in many respects the debate between proponents and skeptics has become reminiscent of earlier and broader debates in learning theory and basic perception in which the fruitlessness of the search for a "decisive" experiment on the issue of motivational influences (i.e., one that could not be interpreted by the "other side") became ever more apparent as data multiplied and conceptual analysis sharpened.

One response to this state of affairs has been to abandon motivational constructs temporarily and to concentrate upon those non-motivational

factors (i.e., informational, perceptual, and cognitive factors) that influence and potentially distort attributional judgments. Beyond the existing conceptual difficulties, empirically mixed results, and historical lessons that discourage the investigators who would search for encompassing motivational biases, there are two additional reasons for the contemporary shift. First, there is a growing conviction that a fuller appreciation of non-motivational influences might lead us to understand and anticipate those circumstances in which attributions of responsibility are likely to enhance the attributer's self-esteem and those in which such attributions are likely to attenuate his self-esteem (cf. Miller & Ross, 1975). Second, there is the increasing recognition that accurate attributions generally are apt to be more "self-serving" than inaccurate ones - that is, that distortions of causal judgment are apt to leave the organism *badly* prepared for the task of long-term survival, however pleasant the immediate consequences of certain inaccurate perceptions and influences.

The rest of this chapter deals with a limited number of such non-motivational biases (see Nisbett & Ross, 1980, for a more thorough review). It also discusses a general phenomenon that increases the "costs" of such biases - the tendency for erroneous impressions, judgments, and even broader theories to survive in the face of logically powerful data that contradict these beliefs. Let us recognize from the outset, however, that the errors and biases dealt with are not inexplicable perversities on the intuitive scientist's part. Typically, they reflect the operation of mechanisms and strategies that serve the organism reasonably well in many circumstances; otherwise they surely would not survive the learning history of the individual or the evolutionary history of the species. These errors and biases can fairly be regarded as "domain specific" failings of inferential strategies and tactics that are at least cost efficient (and probably generally quite accurate as well) in the organism's overall experience.

Non-motivational attribution biases

The fundamental attribution error

The first identified (Heider, 1958) and most frequently cited non-motivational bias, one that we shall term the *fundamental attribution error*, is the tendency for attributers to underestimate the impact of situational factors and to overestimate the role of dispositional factors in controlling behavior. As "intuitive" psychologists, we seem too often to be nativists, or proponents of individual differences, and too seldom S - R behaviorists. We too readily infer broad personal dispositions and expect consistency in behavior or outcomes across widely disparate situations and contexts. We jump to hasty conclusions upon witnessing the behavior of our peers, overlooking the impact of relevant environmental forces and constraints.

General evidence for the fundamental attribution error. Beyond anecdotes and appeals to experience, the evidence most frequently cited for this general bias (e.g., Jones & Nisbett, 1971; Kelley, 1971) involves the attributer's apparent willingness to draw "correspondent" personal inferences about actors who have responded to very obvious situational pressures. For instance, Jones and Harris (1967) found that listeners assumed some correspondence between communicators' pro-Castro remarks and their private opinions even when these listeners *knew* that the communicators were obeying the experimenter's explicit request under no-choice conditions.

A more direct type of evidence that observers may ignore or underestimate situational forces has been provided by Bierbrauer (1973), who studied subjects' impressions of the forces operating in the classic Milgram (1963) situation. In Bierbrauer's study, participants witnessed a verbatim reenactment of one subject's "obedience" to the point of delivering the maximum shock to the supposed victim. Regardless of the type and amount of delay before judging, regardless of whether they actually played the role of a subject in the reenactment or merely observed, Bierbrauer's participants showed the fundamental attribution error; that is, they consistently and dramatically underestimated the degree to which subjects in general would yield to those situational forces that compelled obedience in Milgram's situation (see Figure 1). In other words, they assumed that the particular subject's obedience reflected his distinguishing personal dispositions rather than the potency of situational pressures and constraints acting upon all subjects.

The special case of role-conferred advantages in self-presentation. The tendency of social observers to underestimate the potency of situational forces and constraints and to overestimate the role of individual dispositions has figured heavily in the strategy, conceptual analyses, and even in the professional debates of contemporary social psychology (see Nisbett & Ross, 1980; Ross, 1977). Certain special cases of this fundamental attribution error help to focus our attention on mediating processes and more specific failures of the intuitive psychologist. An experiment by Ross, Amabile, and Steinmetz (1977), dealing with evaluations made about actors who were role-advantaged or role-disadvantaged (by random assignment), is a case in point. The particular roles dealt with by Ross et al. were those of questioner and contestant in a general-knowledge quiz game. The questioner's role obliged the subject to compose a set of challenging general-knowledge questions, to pose these questions to the contestant, and to provide accurate feedback after each response by the contestant. The contestant's role was restricted to answering, or attempting to answer, the relevant questions. Both of these participants (and, in a subsequent reenactment, observers as well) were then required to rate the questioner's and the contestant's general knowledge.

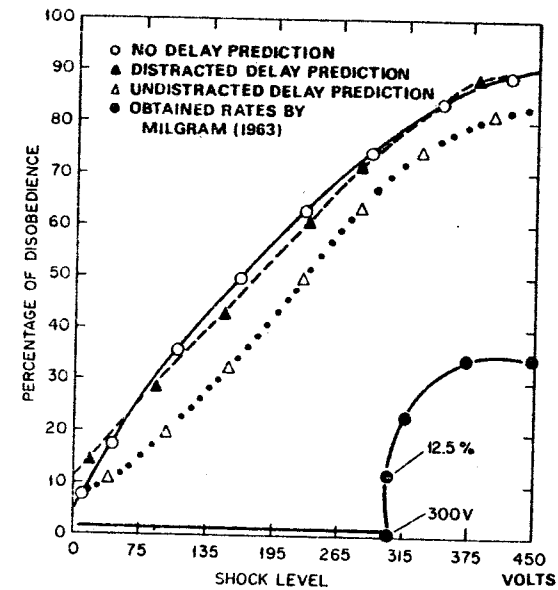


Figure 1. Comparison of predicted and actual disobedience rates.

The arbitrary assignment and fulfillment of these roles, it should be apparent, forced the participants and observers to deal with blatantly non-representative or biased "samples" of the questioners' and contestants' knowledge. The questioners' role encouraged them to display esoteric knowledge and guaranteed that they would avoid areas of ignorance; the contestants were denied such advantages in self-presentation. Indeed, there was virtually no ambiguity about the arbitrariness of the role assignment or about the differing prerogatives associated with each role, unlike many real-world situations in which social roles similarly confer advantages and disadvantages in self-display. Nevertheless, the unequal contest between questioners and contestants led to consistently biased and erroneous impressions. The participants, in a sense, simply failed to make adequate allowance for the situationally conferred advantages and disadvantages of the relevant roles. Thus, contestants rated their questioners as far superior to themselves, and uninvolved observers clearly agreed (see Figure 2). The observers, armed with the knowledge that they could no more answer the esoteric questions posed than could the contestants, recognized that the contestants were not deficient in their general knowledge. What the observers concluded, instead, was that the questioners were truly outstanding in their general knowledge. Interestingly, the questioners themselves were not misled by their encounter. An appreciation of this fact shifts our focus from the general existence of the fundamental attribution error, and the specific impact of social roles, to

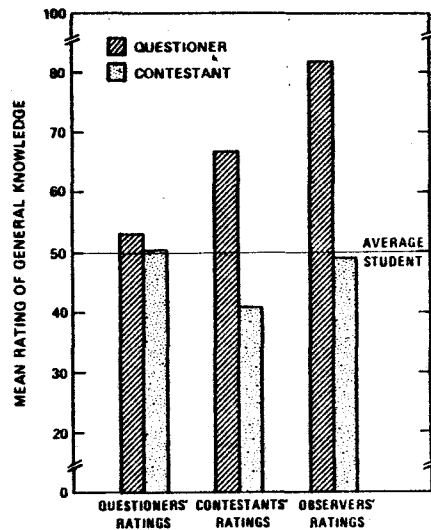


Figure 2. Ratings of questioners' and contestants' general knowledge.

the particular "data samples" upon which the various participants relied for their inferences. Unlike the contestants and the observers, the questioners were in no case forced to rely exclusively upon "biased" samples of general knowledge. Presumably, they had a great deal of additional evidence about the extent and limitations of their own general knowledge and about the unrepresentativeness of the esoteric items of information they displayed in their questions; consequently, they rated both themselves and the contestants as "average."

Both the social and theoretical implications of the Ross, Amabile, and Steinmetz demonstration should be clear. It prompts us to consider the countless social contexts in which formal or informal roles constrain interpersonal encounters and, in so doing, bias the impressions of the participants - even to the point of seeming to justify the prerogatives and limitations that are imposed by the advantaged and disadvantaged roles. It also prompts us to sharpen our focus on one of the specific failings of the intuitive scientist - his seeming insensitivity to the limited inferential value of biased data samples (see also Hamill, Wilson, & Nisbett, 1980; Nisbett & Ross, 1980, Chap. 4).

Salience or availability biases

Perhaps the most energetically researched area of attributional bias has been that involving the effects of attention and of the perceptual and cognitive factors that mediate attention. Briefly stated, it appears that whenever some aspect of the environment is made disproportionately salient or "available" to the perceiver (cf. Tversky & Kahneman, 1973, 11)

that aspect is given more weight in causal attribution. Thus, when an actor is made visually salient because of a unique racial or sexual status within a larger group (Taylor et al., 1976), because of some striking feature of appearance or dress (McArthur & Post, 1977; McArthur & Solomon, 1978), because of an instructional set (Regan & Totten, 1975), or even because of seating arrangements or other determinants of visual perspective (e.g., Storms, 1973; Taylor & Fiske, 1975), that actor is assigned disproportionate responsibility for any outcome to which she or he contributes. (See Taylor & Fiske, 1978, for a more complete review.) Indeed, a number of studies derived from "objective self-awareness" theory (Duval & Wicklund, 1972; Wicklund, 1975) have shown that actors' perceptions of their own causal roles can similarly be influenced by simple manipulations that direct their attention toward or away from the self as a social object (e.g., Duval & Hensley, 1976; Ellis & Holmes, 1979).

Recognition and understanding of how salience or availability factors affect the attributional process may help us better to understand the bases of many familiar attributional and inferential biases, perhaps even subsuming them as special cases. The fundamental attribution error, for instance, may importantly reflect the fact that actors are simply more salient than environmental features and therefore are more likely to be noticed in the attributer's initial search for causal candidates. Indeed, when situational factors and constraints are made disproportionately salient to the attributer, we might expect attributional errors that seem to be opposite to the so-called fundamental error. Thus, a supervisor can be led to incorrectly attribute a worker's trustworthy performance to an external factor - that is, the supervisor's surveillance - when that external factor is made highly salient (see Strickland, 1958). By the same token, an actor's intrinsic interest in a given task can be undermined (e.g., Deci, 1971; Lepper & Greene, 1975, 1978; Lepper, Greene, & Nisbett, 1973) if that actor is led to focus attention on an external incentive or constraint that seemingly encourages, but is in fact not necessary to encourage, performance in that task.

Consider also Jones and Nisbett's (1971) empirical generalization that actors, in accounting for their behavior, are relatively more inclined to cite situational factors and less inclined to cite dispositional factors than are observers of such behavior. To the extent that actors and observers show corresponding differences in their focus of attention - that is, actors attend to relevant features of their environment while observers focus their attention on the actors themselves - the Jones and Nisbett generalization becomes a special case of the attention/attribution generalization. Indeed, experimental evidence suggests that by manipulating actors' and observers' focus of attention, or perspective, their tendencies to cite situational versus dispositional causes can likewise be manipulated (e.g., Storms, 1973; Taylor & Fiske, 1975).

Finally, let us consider an inferential shortcoming noted by that astute

fictional detective/psychologist Sherlock Holmes – the tendency to overlook the inferential value of non-occurrences. Such informative non-occurrences are events or actions that have *not* occurred in some context, which by not occurring thereby contain potentially important information (cf. Ross, 1977). The special relevance of the relatively low cognitive availability of non-occurrences should be clear. When one searches for or considers causal candidates, non-occurrences are unlikely to be highly salient or appropriately attended to; accordingly, one is not apt to give them sufficient weight in accounting for observed actions and outcomes. Jill is more likely to attribute Jack's anger to something she has "done" than to something she has failed to do, simply because the former is apt to be more salient to her than the latter. Indeed, assuming that the sins of omission are apt to be less salient than the sins of commission, Jack is apt to make the same error in accounting for his own anger.

The false consensus or egocentric attribution bias

The final non-motivational, or "informational," bias to be considered in this chapter relates to people's estimates of social consensus – the perceived commonness or oddity of the various responses they witness. Unlike the professional psychologist, who relies upon well-defined sampling techniques and statistical procedures for making such estimates, the layperson must rely upon intuitions and subjective impressions based on limited access to relevant data. The possibilities for bias in such estimates, and in the various social inferences or attributions that reflect such estimates, are thus legion. The specific attributional bias that we shall consider here concerns people's tendency to perceive a "false consensus" – that is, to see their own behavioral choices and judgments as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant, and inappropriate.

References to "egocentric attribution" (Heider, 1958; Jones & Nisbett, 1971), to "attributive projection" (Holmes, 1968), and to specific findings and phenomena related to false consensus biases have appeared sporadically in the social perception and attribution literatures (cf. Katz & Allport, 1931; Kelley & Stahelski, 1970). Perhaps the most compelling evidence, however, has been provided in a series of studies by Ross, Greene, and House (1977).

In the first study reported, subjects read descriptions of hypothetical conflict situations of the sort they might personally face and were to (a) estimate the commonness of the two possible response alternatives; (b) indicate the alternative they, personally, would follow; (c) assess the traits of the "typical" individual who would follow each of the two specified alternatives.

The estimates and ratings demonstrated the "false consensus" effect; subjects estimated that the alternative they chose would be relatively more

common than the unchosen alternative. An obvious corollary to the false consensus proposition is that the intuitive psychologist judges those responses that differ from his own to be more revealing of the actor's stable dispositions than those responses that are similar to his own. The Ross, Greene, and House (1977) data support this prediction; subjects made relatively more confident and extreme predictions about the typical person who would perform the subject's unchosen alternative than about the typical person who would perform the subject's chosen alternative.

The term *relative* is critical in this formulation of the false consensus bias and it requires some clarification. Obviously, the man who would walk a tightrope between two skyscrapers, launch a revolution, or choose a life of clerical celibacy recognizes that his choices would be shared by few of his peers. The false consensus bias would, however, lead him to see his personal choices as less deviant than they would seem to us who would not walk tightropes, launch revolutions, or become celibate clerics. Similarly, the present thesis concedes that for some response categories virtually all raters' estimates may be biased in the same direction. The incidence of infant abuse, for instance, might be underestimated by abusing and non-abusing parents alike. The relative terms of the false consensus hypothesis lead only to the prediction that abusing parents will estimate child abuse to be more common and less revealing of personal dispositions than will non-abusing parents.

In a final demonstration by Ross, Greene, and House (1977) the hypothetical questionnaire methodology was abandoned and subjects were confronted with a real and consequential conflict situation. Subjects were asked to walk around campus for 30 minutes wearing a large sandwich-board sign bearing the message "EAT AT JOE'S." The experimenter made it clear to subjects that they could easily refuse to participate in the sandwich-board study but that he would prefer them to participate and thereby "learn something interesting while helping the research project." Subjects were subsequently asked to make their own decision about taking part in the study, to estimate the probable decisions of others, and to make trait inferences about particular peers who agreed or refused to participate.

The results using this "real" conflict situation (Table 1) confirmed the findings of earlier questionnaire studies dealing with hypothetical responses. Overall, subjects who agreed to wear the sandwich-board sign estimated that 62% of their peers would make the same choice. Subjects who refused to wear the sign estimated that only 33% of their peers would comply with the experimenter's request. Furthermore, as predicted, compliant and non-compliant subjects disagreed sharply in the relative strength of inferences they were willing to make about one peer who agreed and one who refused to wear the sandwich board. Compliant subjects made more confident and more extreme inferences about the personal characteristics of the non-compliant peer; non-compliant subjects made stronger inferences about the compliant peer.

Table 1. The false consensus effect: Raters' estimates of commonness and trait inferences regarding two behavioral alternatives

| | Estimated commonness of agreement (%) | Estimated commonness of refusal (%) | Strength of trait inferences* | |
|---|---------------------------------------|-------------------------------------|---------------------------------------|--|
| | | | About subject who agrees to wear sign | About subject who refuses to wear sign |
| Subjects who agree to wear sign (n = 48) | 62 | 38 | 120.1 | 125.3 |
| Subjects who refuse to wear sign (n = 32) | 33 | 67 | 139.7 | 106.8 |

*Sum of ratings for four traits; higher number indicates more confident and more extreme inferences by rater.
Source: Summarized from Ross, Greene, & House (1977).

Some broad implications of the Ross, Greene, and House (1977) demonstrations for our conception of the intuitive psychologist should be clear. Lay estimates of deviance and normalcy, and the host of social inferences and interpersonal responses that accompany such estimates, are systematically and egocentrically biased in accord with the layperson's own behavioral choices. More generally, it is apparent that attributional analyses may be distorted not only by errors in the intuitive psychologist's eventual analysis of social data but also by earlier biases in sampling or estimating such data.

Several non-motivational factors appear to play a role in producing false consensus phenomena. Principal among these are (a) selective-exposure and availability factors, and (b) factors pertaining to the resolution of situational ambiguity.

Selective-exposure factors underlying false consensus are fairly straightforward. Obviously, we know and associate with people who share our background, experiences, interests, values, and outlook. Such people *do*, in disproportionate numbers, respond as we would in a wide variety of circumstances. Indeed, our close association is determined, in part, by feelings of general consensus, and we may be inclined to avoid those whom we believe unlikely to share our judgments and responses. This exposure to a biased sample of people and behavior does not demand that we err in our estimates concerning the relevant populations, but it does make such errors likely. More subtle and more cognitive in character are the factors that increase our ability to recall, visualize, or imagine paradigmatic instances of behavior. In a given situation the specific behaviors we have chosen or would choose are likely to be more readily retrievable from memory and more easily imagined than opposite behaviors. In Kahneman and Tversky's (1973; 4) terms, the behavioral choices we favor may be more cognitively "available," and we are apt to be misled by this ease or difficulty of access in estimating the likelihood of relevant behavioral options.

A second non-motivational source of the false consensus effect arises from the intuitive psychologist's response to ambiguity - both about the nature and magnitude of situational forces and about the meaning and implications of various response alternatives. Attempts to resolve such ambiguity involve interpretation, estimation, and guesswork, all of which can exert a parallel effect on the attributor's own behavior choices and upon his predictions and inferences about the choices of others. Thus, subjects who anticipated and feared the ridicule of peers for wearing the "EAT AT JOE'S" sign and who regarded the experimenter's wishes and expectations as trivial were likely to refuse to wear the sign, to assume similar refusals by their peers, and to draw strong inferences about the traits of any subject who chose to wear the sign. Opposite priorities, of course, would have produced opposite personal choices and opposite social estimates and inferences.

In summary, the false consensus bias both reflects and creates distortions in the attribution process. It results from non-random sampling and retrieval of evidence and from idiosyncratic resolution of ambiguous situational factors and forces. In turn, it biases judgments about deviance and deviates, and, more generally, promotes variance and error in the interpretation of social phenomena.

Belief perseverance in the face of empirical challenges

The intuitive psychologist's various shortcomings – those described in this chapter and elsewhere (see Nisbett & Ross, 1980) – can lead him to hold beliefs about himself, about other people, or even about the nature of the social world, that are premature and in many cases erroneous. As long as they remain private and are not acted upon, such beliefs may seem inconsequential – merely tentative in nature and adjustable to new input. A gradually increasing body of theory and research, however, can now be marshaled to suggest the contrary.

It appears that beliefs – from relatively narrow personal impressions to broader social theories – are remarkably resilient in the face of empirical challenges that seem logically devastating. Two paradigms illustrate this resilience. The first involves the capacity of belief to survive and even be strengthened by new data, which, from a normative standpoint, should lead to the moderation of such beliefs. The second involves the survival of beliefs after their original evidential bases have been negated.

Belief perseverance and polarization in the face of new data

Individuals, social factions, interest groups, and even nations often hold differing beliefs about pressing social or political issues. Such divergences in opinion are hardly surprising. Given the informal and often purely intuitive basis on which such opinions are formulated, and given the role that social communications (often highly biased ones) play in shaping our beliefs, honest disagreements are inevitable. But what happens when the holders of divergent viewpoints are allowed to examine relevant evidence – especially when that evidence is relatively formal in nature and is identical for all concerned parties?

An optimistic expectation is that the contending factions would narrow the gap between their beliefs. This narrowing might consist of change toward the position justified by the relevant evidence, if such evidence were consistent and compelling; alternatively, it might consist of change toward greater moderation or mutual tolerance, if the relevant evidence were mixed or inconclusive. A less optimistic expectation is that the contending factions would remain unmoved; that is, they would disregard the new evidence and hold fast to their original positions. A recent

experiment by Lord, Lepper, and Ross (1979) suggests an even more disheartening result (disheartening, at least, for those who hope or expect the objective data of the social scientist to dampen the fires of social dispute).

Lord et al. (1979) first selected subjects who either supported capital punishment and believed it to be an effective deterrent (proponents) or opposed capital punishment and believed it not to be a deterrent (opponents). The subjects were presented, in a counterbalanced design, with two purportedly authentic empirical studies. One seemingly provided empirical support for their position; the other seemingly opposed that position. At strategic points in the reading of these two studies, the two groups completed ratings dealing both with their evaluations of the two studies and with their own changes in attitudes and beliefs. These ratings dramatically revealed the capacity of theory-holders to interpret new evidence in a manner that strengthens and sustains their theories. First, both proponents and opponents of capital punishment consistently rated the study that supported their beliefs as “more convincing” and “better conducted” than the study that opposed those beliefs. Second, and in contrast to any normative strategy imaginable for incorporating new evidence relevant to one's beliefs, the net effect of reading the two studies was to polarize further the beliefs of the death penalty opponents and proponents. The manner in which this polarization occurred was particularly illuminating (see Figure 3). Upon reading a brief statement of a result that supported their own viewpoint, subjects' beliefs became considerably more extreme; these changes were maintained or enhanced when the subjects considered details about the procedure and data. By contrast, upon reading a brief result statement that opposed their own viewpoint, subjects became only slightly less extreme; and upon reading the relevant details concerning procedures and data the subjects tended to revert to the beliefs they had held before ever learning of the study's existence. In fact, many individual subjects who had read both the results summary and the procedural details of a study that opposed their belief ultimately became more convinced of the correctness of that belief! No such effects occurred when the same results and procedures were read by subjects whose initial views were supported.

Obviously, professional scientists frequently are guilty of the same offense as intuitive ones. Again and again one sees contending factions that are involved in scholarly disputes – whether they involve the origins of the universe, the line of hominid ascent, or the existence of ego-defensive attribution biases – draw support for their divergent views from the same corpus of findings. Later in this chapter we shall consider the processes underlying such phenomena in more detail and comment more specifically on the normative status of the scientist's willingness to process evidence in the light of his existing theories and expectations. First it is necessary to consider a second general class of perseverance phenomena.

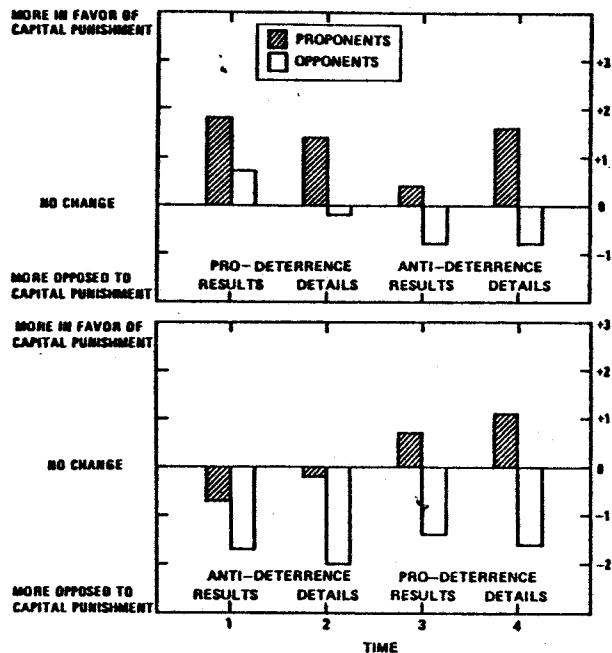


Figure 3. Top panel: Attitude changes on capital punishment relative to start of experiment as reported across time by subjects who received pro-deterrence study first. Bottom panel: Attitude changes on capital punishment relative to start of experiment as reported across time by subjects who received anti-deterrence study first.

Belief perseverance after evidential discrediting

Sometimes, one's beliefs are threatened not by new data but rather by challenges to the "formative" evidence for such beliefs – that is, to the information or analysis that led one to form the belief in the first place. At an anecdotal level it is easy to cite such instances. Sally fails miserably in her first attempts to learn how to skate and then finds out that her borrowed skates were much too large to give her the needed ankle support. Mary assumes that John's bouquet of flowers reflects thoughtfulness, a romantic nature, and a certain conventionalism, only to learn subsequently that John's father owns a flower shop. Broader theories or beliefs about the world similarly can be challenged. A baby-sitter decides, on the basis of his experience with a single infant who cried all night, that bottle feeding produces colicky babies, only to discover that the infant in question was suffering from a high fever. Or a scientist discovers that a classic experiment that figured in the emergence of a particular theory was tainted with some severe artifact or by outright fraud.

The perseverance hypothesis, in its general formulation, suggests that the social perceivers identified above would persist in their initial assess-

ments to an unwarranted and inappropriate degree. But the terms of such a contention are obviously too general and too vague to be testable. When, precisely, can we infer that a social perceiver is "inappropriately" persisting in an impression or belief whose basis has been undermined? To explore the perseverance hypothesis experimentally, what clearly is required is a paradigm that permits us to specify precisely how much perseverance and how much change might be warranted.

One such paradigm is suggested by the dilemma of the social psychologist who has made use of deception in the course of an experiment and then seeks to debrief the subjects who had been the target of such deception. The psychologist reveals the totally contrived and inauthentic nature of the information presented presuming that this debriefing will thereby eliminate any effects such information might have exerted upon the subjects' feelings or beliefs. Many professionals, however, have expressed public concern that such experimental deceptions may do great harm that is not fully undone by conventional debriefing procedures (e.g., Kelman, 1972; A.G. Miller, 1972; Orne, 1972; Silverman, 1965).

A series of experiments by Lepper, Ross, and their colleagues (see also earlier studies by Walster et al., 1967; Valins, 1974) have used the total discounting, or debriefing, paradigm to explore the phenomenon of belief perseverance in the face of evidential discrediting. We may begin discussing this work by outlining a pair of experiments by Ross, Lepper, and Hubbard (1975) that dealt with subjects' postdebriefing impressions about their own abilities at a particular task or about the abilities of a peer.

Postdebriefing perseverance of personal impressions. The procedure employed by Ross et al. (1975) was quite straightforward. Subjects first received continuous false feedback as they performed a novel discrimination task (i.e., distinguishing authentic suicide notes from fictitious ones). In the first experiment reported this procedure was used to manipulate the subjects' perceptions of their own performance and ability. A second experiment introduced observers, who formed social impressions as they witnessed the false feedback manipulation. In both experiments after this manipulation of first impressions had been completed, the experimenter totally discredited the "evidence" upon which the actors' and/or observers' impressions had been based. Specifically, the actor (overheard in Experiment 2 by the observer) received a standard debriefing session in which he learned that his putative outcome had been predetermined and that his feedback had been totally unrelated to actual performance. Before dependent variable measures were introduced, in fact, every subject was led to explicitly acknowledge his understanding of the nature and purpose of the experimental deception.

Following this total discrediting of the original information, the subjects completed a dependent variable questionnaire dealing with the actors' performances and abilities. The evidence for postdebriefing

Table 2. *Postdebriefing perceptions of the actor's performance and ability*

| | Actor's own perceptions | | | Observer's perceptions of actors | | |
|----------------------------------|-------------------------|---------|----------|----------------------------------|---------|----------|
| | Success | Failure | <i>t</i> | Success | Failure | <i>t</i> |
| Estimated initial number correct | 18.33 | 12.83 | 5.91*** | 19.00 | 12.42 | 4.43*** |
| Predicted future number correct | 18.33 | 14.25 | 4.23*** | 19.08 | 14.50 | 2.68* |
| Rated ability at task | 5.00 | 3.83 | 2.65* | 5.33 | 4.00 | 3.36** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

Source: Summarized from Experiment 2 of Ross, Lepper, & Hubbard (1975).

impression perseverance was unmistakable for actors and observers alike. On virtually every measure (i.e., objective estimates of the actor's just-completed performance, estimates for performance on a future set of discrimination problems, and subjective estimates of the actor's abilities) the totally discredited initial outcome manipulation produced significant "residual" effects upon actors' and observers' assessments (see Table 2).

Follow-up experiments have since shown that a variety of unfounded personal impressions, once induced by experimental procedures, can survive a variety of total discrediting procedures. For example, Jennings, Lepper, and Ross (1980) have demonstrated that subjects' impressions of their ability at interpersonal persuasion (having them succeed or fail to convince a confederate to donate blood) can persist after they have learned that the initial outcome was totally inauthentic. Similarly, in two related experiments Lepper, Ross, and Lau (1979) have shown that students' erroneous impressions of their "logical problem solving abilities" (and their academic choices in a follow-up measure two months later) persevered even after they had learned that good or poor teaching procedures provided a totally sufficient explanation for the successes or failures that were the basis for such impressions.

Postdebriefing perseverance of discredited theories. A recent series of experiments by Anderson, Lepper, and Ross (1980) have extended the domain of perseverance demonstrations from personal impressions to broader beliefs about the world. Anderson et al.'s studies first manipulated and then attempted to undermine subjects' theories about the functional relationship between two measured variables: the adequacy of firefighters' professional performances and their prior scores on a paper and pencil test of risk preference. In one particularly pertinent variation, the formative evidence consisted of only one pair of specific cases - i.e., one successful

and one unsuccessful firefighter with appropriately discrepant scores in their respective tests of risk-taking preferences. Interestingly, such minimal data did suffice to produce strong theories, on the subjects' part, about the probable relationship between the relevant measures. More important, however, was the finding that such theories survived the revelations that the cases in question had been totally fictitious and the different subjects had, in fact, received opposite pairings of riskiness scores and job outcomes. Indeed, when comparisons were made between subjects who had been debriefed and those who had not been, it appeared that over 50% of the initial effect of the "case history" information remained after debriefing.

In summary, it is clear that beliefs can survive potent logical or empirical challenges. They can survive and even be bolstered by evidence that most uncommitted observers would agree logically demands some weakening of such beliefs. They can even survive the total destruction of their original evidential bases. While much work remains to be done in specifying the precise limits and exploring inevitable exceptions to such phenomena, it is clear that the costs of the layperson's attributional biases and other inferential shortcomings are apt not to be corrected but instead to be compounded by subsequent experience and deliberations. The question that must at last be addressed, therefore, is *how* and *why* does such perseverance occur? That is, what cognitive mechanisms underlie the unwarranted persistence of our impressions, beliefs, and broader social theories?

Possible mechanisms underlying belief perseverance

Biased search, recollection, and assimilation of information. There can be little doubt that our beliefs influence the processes by which we seek out, store, and interpret relevant information. Indeed, without prior knowledge and corresponding preconceptions, our understanding of everyday experience would demand considerably more time and effort, and in all likelihood that understanding would be greatly diminished. But an inevitable consequence of our willingness to process evidence in the light of our prior beliefs is the tendency to perceive more support for those beliefs than actually exists in the evidence at hand.

Such "confirmation biases" (see Einhorn & Hogarth, 1978; Hamilton, 1979; Hastie & Kumar, 1979; Wason & Johnson-Laird, 1972) have long been noted by philosophers of science (e.g., Bacon, 1620/1960). Perhaps most noteworthy is the theory holder's response to equivocal or ambiguous data. As Lord et al. (1979) have documented, potentially confirmatory evidence is apt to be taken at face value while potentially disconfirmatory evidence is subjected to highly critical and skeptical scrutiny. Thus, two

consequences follow: First, any pattern of evidence processed in this fashion, even evidence that is essentially random, will tend to *bolster* the initial belief. Second, once evidence has been processed in this fashion it gains the capacity to *sustain* the prior belief when that belief is subjected to new empirical disconfirmation or to attacks on its original evidential basis.

The role of biased assimilation has been shown fairly convincingly, we think, for the case where the theory holder is confronted with new data (i.e., Lord et al., 1979). But the role of this mechanism in the discounting or debriefing paradigm is perhaps less obvious, and we are forced to rely on speculation rather than hard data. We suggest that the subject who forms an initial impression about himself, about another person, or about some functional relationship is apt to search his memory and the immediate situation for additional data relevant to that impression. Such data, then, are apt to be recalled and regarded as pertinent or probative only to the extent that they confirm the impression at hand. Thus a subject who has succeeded or failed at a given task recalls similar successes or failures at related tasks – and decides upon their relevance to the present case – on the basis of the congruency of the relevant outcomes. Similarly, a subject who has come to believe that variables X and Y are functionally related will recall, and give credence to, cases that confirm rather than challenge that presumed relationship. Once again, such biased searching, recollection, and assimilation not only bolster one's initial belief, they also produce a pattern of biased evidence that remains highly available to sustain the belief in question when its *initial* basis is attacked or even destroyed. The critical assumption here is that people do not constantly update or reevaluate the evidence relevant to their beliefs. They do not commonly decide "now that my prior hypothesis has been undermined somewhat I must go back and reassess all of the evidence that I ever considered in the light of that hypothesis."

The formation of causal explanations. People do more than merely note evidence relevant to their impressions or beliefs. They also engage in causal analysis or explanation (Heider, 1958). That is, they try to *account for* the characteristics of self or others, or for the functional relationships that they have come to believe exist. Thus, the subject who believes herself a superior or inferior discriminator of suicide notes in the Ross et al. (1975) study might search for some aspect of her background that would account for such a talent or deficiency. Similarly, the subject who is induced to believe in a positive or negative relationship between firefighting ability and risk preference will have little difficulty in postulating a logical basis for either relationship. Once again, this process not only buttresses an initial impression or belief, it is apt to sustain that impression or belief in the face of subsequent challenges or attacks.

Evidence for the operation of this perseverance mechanism comes

primarily from two debriefing studies demonstrating that when subjects are explicitly required to formulate such explanations, prior to debriefing, the magnitude of the perseverance effect is increased. In the Anderson et al. (1980) study one group of subjects was explicitly instructed to explain the positive or negative relationship suggested by the two firefighter cases. As predicted, this manipulation greatly enhanced the relevant perseverance effect. In fact, subjects who explained the basis for a positive or for a negative relationship before being debriefed were only trivially less certain of that relationship than subjects who received no debriefing. Similar results were obtained by Ross, Lepper, Strack, and Steinmetz (1977), who found that subjects induced to explain outcomes in the lives of clinical patients (whose earlier case histories they had read) continued to regard such outcomes as relatively likely even when they learned that the explained events were inauthentic and had been contrived by the experimenter.

Behavioral confirmation or "self-fulfilling" hypotheses. The two research paradigms used by Ross, Lepper, and their colleagues to investigate perseverance phenomena lack one element that may be critical to many everyday situations. Specifically, subjects in those studies lacked the opportunity to act upon their beliefs. Such actions are important partially because they can increase the psychological costs or "dissonance" (Festinger, 1957) involved in changing one's beliefs (cf. Ashmore & Collins, 1968; Collins & Hoyt, 1972; Hovland, Campbell, & Brock, 1957). Furthermore, such actions create new data relevant to those beliefs. Not only may this new data be processed in a biased manner, but the data themselves may also be biased in a direction that tends to confirm the relevant hypothesis.

The idea of self-confirming, or self-fulfilling, hypotheses is not a new one to social scientists. The famous but controversial "Pygmalion" studies by Rosenthal and Jacobson (1968), which dealt with the impact of teachers' expectations upon the "blooming" of their students' abilities and performances, is a case in point. However, a recent series of studies by Snyder and his colleagues have considerably advanced our appreciation and understanding of such phenomena by demonstrating the manner in which subjects' expectations, or the hypotheses they are led to test, can generate "objective support" for those expectations or hypotheses (e.g., Snyder & Swann, 1978a, 1978b; Snyder, Tanke, & Berscheid, 1977).

Concluding remarks: Beliefs do change!

Our foregoing discussion of phenomena and mechanisms should not make the reader lose sight of the fact that beliefs about ourselves, our political leaders, and even our scientific theories *do* change. In part such change may simply be the result of brute force. Even if logical or empirical challenges have less impact than might be warranted by normative

standards (see Ross & Lepper, 1980) they may still get the job done. In part, such change may reflect the fact that formal methods of hypothesis testing sometimes are deliberately employed to protect us from the dangers of informal ones. But we suspect there is more to the story, for there is evidence that prior theories can sometimes be overcome without massive amounts of disconfirming evidence or decisive well-controlled experiments. Thus, the changes in outlook and belief that can be wrought by vivid, concrete, first-hand experience (see Nisbett & Ross, 1980) and the effectiveness of groups and leaders that accomplish dramatic political or religious conversions offer inviting targets for future research.

REFERENCES

- Anderson, C. A., Lepper, M. R., and Ross, L. Theory perseverance: The persistent influence of discredited information. Journal of Personality and Social Psychology, in press, 1980.
- Ashmore, R. D., and Collins, B. E. Studies in forced compliance X: Attitude change and commitment to maintain publicly a counter-attitudinal position. Psychological Reports, 1968, 22, 1229-1234.
- Bacon, F. The new organon and related writings. New York: Liberal Arts Press, 1960. (Originally published, 1620.)
- Beckman, L. Effects of students' performance on teachers' and observers' attributions of causality. Journal of Educational Psychology, 1970, 61, 75-82.
- Bierbrauer, G. Effect of set, perspective, and temporal factors in attribution. Unpublished doctoral dissertation, Stanford University, 1973.
- Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. Journal of Personality and Social Psychology, 1978, 36, 56-71.
- Collins, B. E., and Hoyt, M. F. Personal responsibility-for-consequences: An integration and extension of the forced-compliance literature. Journal of Experimental Social Psychology, 1972, 8, 558-593.
- Davis, W. L., and Davis, D. E. Internal-external control and attribution of responsibility for success and failure. Journal of Personality, 1972, 40, 123-136.
- Deci, E. Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology, 1971, 18, 105-115.
- Duval, S., and Hensley, V. Extensions of objective self-awareness theory: The focus of attention-causal attribution hypothesis. In J. H. Harvey, W. J. Ickes, and R. F. Kidd (Eds.), New directions in attribution research, Vol. 1, 1976.

- Duval, S., and Wicklund, R. A. A theory of objective self awareness. New York: Academic Press, 1972.
- Einhorn, H. J., and Hogarth, R. M. Confidence in judgment: Persistence in the illusion of validity. Psychological Review, 1978, 85, 395-416.
- Ellis, R. J., and Holmes, J. G. Focus of attention in social interaction: The impact on self evaluations. Unpublished manuscript, University of Waterloo, 1979.
- Feather, N. T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. Journal of Personality and Social Psychology, 1969, 13, 129-144.
- Festinger, L. A Theory of Cognitive Dissonance. Stanford, CA: Stanford University Press, 1957.
- Fitch, G. Effects of self-esteem, perceived performance, and chance on causal attributions. Journal of Personality and Social Psychology, 1970, 16, 311-315.
- Freize, I., and Weiner, B. Cue utilization and attributional judgments for success and failure. Journal of Personality, 1971, 39, 591-606.
- Gross, A. Evaluation of the target person in a social influence situation. Unpublished doctoral dissertation, Stanford University, 1966.
- Hamill, R., Wilson, T. D., and Nisbett, R. E. Ignoring sample bias: Inferences about collectives from atypical cases. Unpublished manuscript, Univ. of Michigan, 1980.
- Hamilton, D. L. A cognitive attributional analysis of stereotyping. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. 12. New York: Academic Press, 1979.
- Hastie, R., & Kumar, P. A. Person memory: Personality traits as organizing principles in memory for behavior. Journal of Personality and Social Psychology, 1979, 37, 25-38.

- Heider, F. Social perception and phenomenal causality. Psychological Review, 1944, 51, 358-373.
- Heider, F. The psychology of interpersonal relations. Wiley, 1958.
- Holmes, D. S. Dimensions of projection. Psychological Bulletin, 1968, 69, 248-268.
- Hovland, C. I., Campbell, E. H., and Brock, T. The effects of "commitment" on opinion change following communication. In C. I. Hovland et al., The order of presentation in persuasion. New Haven: Yale University Press, 1957.
- Jennings, D. L., Lepper, M. R., and Ross, L. Persistence of impressions of personal persuasiveness: Perseverance of erroneous self assessments outside the debriefing paradigm. Unpublished manuscript, Stanford Univ., 1980.
- Johnson, T. J., Feigenbaum, R., and Weiby, M. Some determinants and consequences of the teacher's perception of causation. Journal of Experimental Psychology, 1964, 55, 237-246.
- Jones, E. E., and Davis, K. E. From acts to dispositions: The attribution process in perception. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, vol. 2, Academic Press, 1965.
- Jones, E. E., Davis, K. E., and Gergen, K. Role playing variations and their informational value for person perception. Journal of Abnormal and Social Psychology, 1961, 63, 302-310.
- Jones, E. E., and DeCharms, R. Changes in social perception as a function of the personal relevance of behavior. Sociometry, 1957, 20, 75-85.
- Jones, E. E., and Harris, V. A. The attribution of attitudes. Journal of Experimental Social Psychology, 1967, 3, 1-24.
- Jones, E. E., Kanouse, D. E., Kelley, H. H., Nisbett, R. E., Valins, S., & Weiner, B. (Eds.). Attribution: Perceiving the causes of behavior. Morristown, N. J.: General Learning Press, 1971.

- Jones, E. E., and Nisbett, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones et al. (Eds.), Attribution: Perceiving the causes of behavior. Morristown, N.J.: General Learning Press, 1971.
- Kahneman, D., and Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.
- Katz, D., and Allport, F. Student's attitudes. Syracuse: Craftsman Press, 1931.
- Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), Nebraska Symposium on Motivation, vol. 15. Lincoln: University of Nebraska Press, 1967.
- Kelley, H. H. Attribution in social interaction. In E. E. Jones et al. (Eds.), Attribution: Perceiving the causes of behavior. New York: General Learning Press, 1971.
- Kelley, H. H. The process of causal attribution. American Psychologist, 1973, 28, 107-128.
- Kelley, H. H., and Stahelski, A. J. Social interaction basis of cooperators' and competitors' beliefs about others. Journal of Personality and Social Psychology, 1970, 16, 66-91.
- Kelman, H. C. The human use of human subjects: The problem of deception in social psychological experiments. In A. G. Miller (Eds.), The social psychology of psychological research. New York: Macmillan, 1972.
- Lepper, M. R., and Greene, D. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. Journal of Personality and Social Psychology, 1975, 31, 479-486.
- Lepper, M. R., and Greene, D. (Eds.), The hidden costs of reward. Morristown, NJ: Lawrence Erlbaum Associates, 1978.

- Lepper, M. R., Greene, D., and Nisbett, R. E. Undermining children's intrinsic interest with extrinsic reward: A test of the overjustification hypothesis. Journal of Personality and Social Psychology, 1973, 28, 129-137.
- Lepper, M. R., Ross, L., and Lau, R. Persistence of inaccurate and discredited personal impressions: A field demonstration of attributional perseverance. Unpublished manuscript, Stanford University, 1979.
- Lord, C., Lepper, M. R., and Ross, L. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 1979, 37, 2098-2110.
- McArthur, L. Z. The how and what of why: Some determinants and consequences of causal attribution. Journal of Personality and Social Psychology, 1972, 22, 171-193.
- McArthur, L. Z. The lesser influence of consensus than distinctiveness information on causal attributions: A test of the person-thing hypothesis. Journal of Personality and Social Psychology, 1976, 33, 733-742.
- McArthur, L. Z., and Post, D. Figural emphasis and person perception. Journal of Experimental Social Psychology, 1977.
- McArthur, L. Z., and Solomon, L. K. Perceptions of an aggressive encounter as a function of the victim's salience and the perceiver's arousal. Journal of Personality and Social Psychology, 1978, 36, 1278-1290.
- Milgram, S. Behavioral study of obedience. Journal of Abnormal and Social Psychology, 1963, 67, 371-378.
- Miller, A. G. (Ed.), The social psychology of psychological research. New York: Macmillan, 1972.
- Miller, D. T. What constitutes a self-serving attributional bias? Journal of Personality and Social Psychology, 1978, 36, 1221-1223.
- Miller, D. T., and Ross, M. Self-serving biases in the attribution of causality: Fact or fiction? Psychological Bulletin, 1975, 82, 213-225.

- Nisbett, R. E., and Ross, L. Human inference: Strategies and shortcomings. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1980.
- Orne, M. T. Demand characteristics and their implications for real life: The importance of quasi-controls. In A. G. Miller (Ed.), The social psychology of psychological research. New York: Macmillan, 1972.
- Polefka, J. The perception and evaluation of responses to social influences. Unpublished doctoral dissertation, Stanford University, 1965.
- Regan, D. T., and Totten, J. Empathy and attribution: Turning observers into actors. Journal of Personality and Social Psychology, 1975, 32, 850-856.
- Rosenthal, R., and Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart, & Winston, 1968.
- Ross, L. The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, vol. 10, New York: Academic Press, 1977.
- Ross, L. D., Amabile, T. M., and Steinbetz, J. L. Social roles, social control, and biases in social-perception processes. Journal of Personality and Social Psychology, 1977, 35, 485-494.
- Ross, L., Bierbrauer, G., and Polly, S. Attribution of educational outcomes by professional and non-professional instructors. Journal of Personality and Social Psychology, 1974, 29, 609-618.
- Ross, L., Greene, D., and House, P. The false consensus phenomenon: An attributional bias in self perception and social perception processes. Journal of Experimental Social Psychology, 1977, 13, 279-301.
- Ross, L., Lepper, M. R., and Hubbard, M. Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 1975, 32, 880-892.

- Ross, L., Lepper, M. R., Strack, F., and Steinmetz, J. L. Social explanation and social expectation: The effects of real and hypothetical explanations upon subjective likelihood. Journal of Personality and Social Psychology, 1977, 35, 817-829.
- Silverman, I. Motives underlying the behavior of the subject in the psychological experiment. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.
- Snyder, M., and Swann, W. B. Hypothesis-testing processes in social interaction Journal of Personality and Social Psychology, 1978a, 36, 1202-1212.
- Snyder, M., and Swann, W. B. Behavioral confirmation in social interaction: From social perception to social reality. Journal of Experimental Social Psychology, 1978b, 14, 148-162.
- Snyder, M., Tanke, E. D., and Berscheid, E. Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. Journal of Personality and Social Psychology, 1977, 35, 656-666.
- Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' point of view. Journal of Personality and Social Psychology, 1973, 27, 165-175.
- Strickland, L. H. Surveillance and trust. Journal of Personality, 1958, 26, 200-215.
- Taylor, S. E., and Fiske, S. T. Point of view and perceptions of causality. Journal of Personality and Social Psychology, 1975, 32, 429-445.
- Taylor, S. E., and Fiske, S. T. Saliency, attention and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), Advances in Experimental Psychology, vol. 11. New York: Academic Press, 1978.
- Taylor, S. E., Fiske, S. T., Close, M., Anderson, C., and Ruderman, A. Solo status as a psychological variable: The power of being distinctive. Unpublished manuscript, Harvard University, 1979.

- Thibaut, J. W., and Riecken, H. W. Some determinants and consequences of the perception of social causality. Journal of Personality, 1955, 24, 113-133.
- Tversky, A., and Kahneman, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973, 5, 207-232.
- Valins, S. Persistent effects of information about internal reactions: Ineffectiveness of debriefing. In H. London & R. E. Nisbett (Eds.), Thought and feeling: Cognitive modification of feeling states. Chicago: Aldine, 1974.
- Walster, E., Berscheid, E., Abrahams, D., and Aronson, E. Effectiveness of debriefing following deception experiments. Journal of Personality and Social Psychology, 1967, 6, 371-380.
- Wason, P. C., and Johnson-Laird, P. N. Psychology of reasoning: Structure and content. London: Batsford, 1965.
- Weiner, B. Achievement motivation and attribution theory. Morristown, N.J.: General Learning Press, 1974.
- Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. 9. New York: Academic Press, 1975.
- Wolosin, R. J., Sherman, S. J., and Till, A. Effects of cooperation and competition on responsibility attribution after success and failure. Journal of Experimental Social Psychology, 1973, 9, 220-235.