

In creating this electronic reprint, we have attempted to keep the style, pagination, and format as close to the published form as possible. Nonetheless, some errors may have occurred. If you discover an error, please contact Craig Anderson using the following email address: caa@iastate.edu.

This electronic reprint is provided as a courtesy. Please do not post or distribute this reprint in any fashion that may violate the copyright of the original publisher or the authors. Thank you for your interest in this work.

THE DEVELOPMENT, PERSEVERANCE, AND CHANGE OF NAIVE THEORIES

CRAIG A. ANDERSON AND JAMES J. LINDSAY
University of Missouri—Columbia

Naive theories—knowledge structures with a causal or explanatory component—are examined in terms of initial development, resistance to change, and consequences. Three types of psychological processes are proposed to underlie the perseverance of naive theories. Illusory correlations occur when a biased sample of relevant events becomes distinctive by virtue of the numerical frequency of relevant events, the cognitive associations of relevant events, or the image-generating properties of relevant events. Data distortions occur when behavioral confirmation, biased attribution and recall, and biased assimilation processes operate. Use of some form of the availability heuristic underlies many of these effects. A model of naive theory perseverance and change is proposed. The model suggests where interventions might reduce potential biases that typically arise from use of naive theories. Evidence of successful use of several such debiasing techniques is presented.

Social perception requires use of naive theories. Strictly data-driven perception, in the social sphere or elsewhere, would produce the “great blooming, buzzing confusion” described by William James over 100 years ago (1890). But use of naive theories can produce distortions. Upon observing the same event, individuals with different prior theories do not “see” the same event. Hastorf and Cantril (1954) reported wildly divergent perceptions of devious and dirty play in a Princeton-Dartmouth football game, as reported by fans of these two teams. More recently, Sedikides and Anderson (1992) demonstrated that perceptions of a Cold War defector described in a news story were jointly influenced by the perceivers’ *causal knowledge structures* about American and Soviet culture, and whether the story was described as a Soviet or American defection.

In some cases, theory-based divergences in perception signal clearly

Address correspondence to Craig Anderson, Dept. of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211-2500; or, e-mail: psycaa@showme.missouri.edu.

that someone is in error. It cannot be the case that each football team played dirtier than the other. In other cases, naive theories increase perception accuracy. To the extent that the Sedikides and Anderson (1992) participants had accurate naive theories about differences between conditions in the Soviet Union versus the United States, the “same” defection story should have been perceived differently as a function of whether it concerned an American defecting to the Soviet Union or vice versa. In most naturalistically occurring contexts, however, there is no clearly right or wrong social perception.

If naive theories were always accurate, interest in them would be minor. However, they are frequently inaccurate. Researchers from many domains of psychology have realized the importance of understanding how naive theories develop, persist across time and context, and how they occasionally change. In this article we examine these issues using a modified version of Anderson, Krull, and Weiner’s (1996) model of explanation processes as a framework for integrating the literatures on naive theories and theory perseverance. We also discuss techniques designed to reduce or eliminate naive theory biases.

KNOWLEDGE STRUCTURES, NAIVE THEORIES, AND SOCIAL THEORIES

KNOWLEDGE STRUCTURES

The knowledge structure construct—packets of organized information about the world—can be traced to Bartlett’s early memory experiments. He showed that individuals categorize new information in meaningful ways based on prior knowledge structures (Bartlett, 1932). More recent investigators have further shown that identification of an object depends (in part) on how well the features of the object match the knowledge structure representation, or “basic features” of the category (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). For example, the basic features for the category “bird” include “feathers, flies, warm blooded, small.” A stimulus having these features (e.g., robin) will be easily and quickly classified as a bird. Objects having only some of the features (e.g., penguin) will be more difficult to identify.

Knowledge structures can also “drive” feature detection. When a category is used frequently, the perceiver more often “detects” the basic features of that category in ambiguous stimuli, even when they may not actually be present (e.g., Reicher, 1969). For example, aggressive people “see” hostility in the actions of others, even when there is no hostility (Crick & Dodge, 1994; Dill, Anderson, Anderson, & Deuser, 1997).

Knowledge structures include information about relations among the

features. Detecting a "hostile act" requires perception of the features *harmful behavior*, *intent to harm*, and a *causal link* from the intention to the behavior. Similarly, the concepts of love, hate, fairness, and success require perceptions of multiple features and specific relations among them. Indeed, relational elements are merely another type of feature in knowledge structures.

NAIVE THEORIES

Naive theories are knowledge structures with a special twist. They not only describe what features go with what; they also have causal (explanatory) features. In earlier work, we (Sedikides & Anderson, 1992) referred to such theories as "causal knowledge structures." The causal elements give naive theories their utility and predictive power. For instance, naive theories about age—related likes and dislikes help us choose appropriate gifts for the young and the old. But, this causal link can produce harmful effects as well, as in the case where "knowledge" of age—related abilities produces unwarranted ageism in employment settings.

SOCIAL THEORIES

Social theories are merely naive theories with a social (i.e., human interaction) component (Anderson, Lepper, & Ross, 1980). We use both terms synonymously in this article for two reasons. First, most current research on naive theories involves human interactions in some way. Second, the basic principles underlying them are the same, regardless of whether the content includes human interactions. That is, naive theories and social theories are both forms of knowledge structures that have a causal component. The causal component may be explicit, as in high-level debates about the likely outcomes of various poverty programs. It also may be implicit, as in many people's vaguely formed social theories about race differences in intelligence.

RELATED CONCEPTS

Other concepts are also related to naive theories. Scripts, schema, and implicit personality theories are very similar to—and sometimes indistinguishable from—the naive/social theory construct (e.g., Abelson, 1981; Bower, Black, & Turner, 1979; Levy & Dweck, 1997; Schneider, 1973). Attitudes are not naive theories, however, because attitudes need not include organized beliefs and causal links. For example, Petty and Cacioppo (1981, p. 7) define "attitude" as a "general and enduring positive or negative feeling about a person, object or issue" (See also

Zimbardo & Leippe, 1991). Simple attitudes may exist independent of any naive theory, but complex attitudes are always closely linked to at least one. The attitude *is* the stored affective reaction to the cognitive contents of the theory. So, the attitude "I hate Dartmouth football players" is derived (in part) from the naive theory beliefs that (a) their football team plays dirty (b) because of their low moral character. Thus, processes underlying the development, perseverance, and change of naive theories are relevant to understanding similar attitudinal phenomena.

PREDICTION

Naive theories allow people to explain events that have already occurred. More importantly, they predict possible futures. Successful social intercourse—with friends and enemies alike—requires knowing how they will behave in various situations and subsequently creating the right situation to produce the desired effects on their future actions. Some naive theories are specific to a given individual (e.g., person impression), but they are not relegated to idiosyncratic personal beliefs. Skitka and Tetlock (1993; see also Sniderman, Hagen, Tetlock, & Brady, 1986) demonstrated that liberals and conservatives differ in their beliefs about underlying causes of poverty. The naive theories of these groups "explain" poverty in different ways (roughly, social forces vs. laziness) and therefore create different attitudes and preferred solutions.

These examples illustrate that it is the causal relational link among beliefs that yields differential attributions, expectations, attitudes, and behaviors. In the Sedikides and Anderson (1992) defection study, attributions for and memories about the defectors were driven by the different causal links between features about the Soviet and American societies (e.g., poor vs. good living conditions) and personality traits that would account for a defection. Similarly, cultural stereotypes for ethnic groups (Dovidio, Evans, & Tyler, 1986; Devine, 1989; Wittenbriink, Hilton, & Gist, 1997) produce biased attributions, expectations, and behaviors—sometimes without awareness (Gilbert & Hixon, 1991). People possess many naive theories about how various personality traits go together (e.g., Anderson & Sedikides, 1991; Levy & Dweck, 1997), about characteristics necessary for success in particular occupations (e.g., Anderson, Lepper, & Ross, 1980), even about the malleability of traits (Dweck, Hong, & Chiu, 1993).

Thus, researchers from multiple domains have demonstrated that individuals tend to organize information into meaningful categories consisting of multiple features—frequently including causal links among the features. Such naive theories greatly simplify social perception and lead to generally high levels of accuracy in the inferences,

attributions, and expectations that are generated from the categorization process. Sometimes naive theories regarding perceptual biases help people correct the bias (e.g., Wegener & Petty, 1995; Petty, Wegener, & White, 1997). Sometimes the advantages of naive theory use are outweighed by resulting biased inferences, attributions, and behaviors.

HOW DO NAIVE THEORIES DEVELOP?

DIRECT AND INDIRECT EXPERIENCE

Understanding how naive theories develop is crucial to understanding theory perseverance and change. Many naive theories are learned primarily (or initially) through direct experience. Naive theories about the abstract category "falling down" have many features such as causes of falling down (someone pushed), causes of the cause (intent to harm), immediate consequences (experience pain), and potential responses (come up swinging). Though most of the basic features of "falling down" are learned through direct experience—beginning with childhood attempts to stand up, some elaborations are learned through indirect experiences, such as watching pratfalls in slapstick comedy routines. Other theories (e.g., stereotypes) are developed primarily through indirect experience. For instance, most naive theories regarding "Nazis" have been developed indirectly through the media, even among many people who lived through the World War II era.

Many naive theories, such as racial stereotypes, develop from both types of experience. Virtually all U.S. citizens learn the predominant stereotypes of various racial groups in the United States, even without direct experience with members of some of the races. Most people also have had at least some direct experiences with many racial groups in school, at work, or in other social settings.

CAUSAL EXPLANATIONS

Some naive theories develop out of a need to understand or predict some event. The belief perseverance literature (e.g., Anderson et al., 1980; Ross, Lepper, & Hubbard, 1975) has demonstrated that merely presenting people with some event outcome information is sufficient in some circumstances to create naive theories that are resistant to change. For instance, Anderson et al. (1980) gave some participants case histories suggesting that people who make risky decisions are better firefighters than those who make conservative ones. Other participants received information suggesting the opposite. Later, after all participants learned that the initial information was fictitious, the two groups of participants

maintained their initial naive theories about the relation between risk preference and firefighting ability. Other results showed that only those who created causal explanations for the supposed risk preference/firefighting ability relations fell prey to the perseverance effect (Anderson et al., 1980). Further work found that inducing people to think in causal terms (by using vivid case histories) and asking people to create possible explanations for hypothetical causal relations can produce naive theories resistant to change (Anderson, 1983; Anderson & Sechler, 1986).

COMPETING THEORIES AND INDIVIDUAL DIFFERENCES

The large number of potential sources of naive theories makes quite probable the occurrence of competing or contradictory theories within the same individual. For instance, Devine's work demonstrates that most people are aware of cultural stereotypes for minority groups, even though they may refuse (or attempt to refuse) to make judgments on the basis of these naive theories. This suggests that many people possess naive theories based on the cultural stereotype, as well as an opposing theory for these minority groups (Devine, 1989; Yzerbyt, Leyens, & Corneille, 1997). These contradictory information sources are both "available" in memory, although they are not necessarily equally accessible. One difference between prejudiced and nonprejudiced people may well be the relative accessibility of competing naive theories of race. Individual differences in aggressiveness (e.g., K. Anderson, Anderson, Dill, & Deuser, in press; Crick & Dodge, 1994; Dill et al., 1997) and depression/attributional styles (e.g., Abramson, Metalsky, & Alloy, 1989; Anderson & Riger, 1991) also appear to be largely the result of accessibility differences in competing naive theories.

The multiplicity of applicable naive theories for a particular stimulus situation also forms the basis for various priming effects. For instance, the weapons effect (Berkowitz & LePage, 1967), in which provoked aggression is higher when a weapon is in sight versus out of sight, presumably operates by temporarily increasing the accessibility of aggression-related naive theories.

PROCESSES UNDERLYING THE PERSEVERANCE OF NAIVE THEORIES

A fascinating feature of naive theories is their ability to survive empirical disconfirmations. Incorrect naive theories, such as racial stereotype explanations of achievement-test score differences, would be relatively harmless if those theories were easily changed. Similarly, incorrect theories about how HIV is (and is not) spread would be a minor social issue

if direct public persuasion campaigns were effective. The long history of racial prejudice and the short history of HIV ignorance both attest to the perseverant nature of naive theories.

We believe that there are several processes contributing to theory perseverance. Some are behavioral, some are purely cognitive, and some have their origins in motivation. In all cases, however, we believe that ultimately there is a cognitive mechanism involving causal thinking at work. Below, we describe a wide variety of processes that can produce theory perseverance. In the next section, we present an expanded version of the general model of explanation processes recently proposed by Anderson, Krull, and Weiner (1996), and show how the theory perseverance processes can be integrated into this model. Finally, we derive from the model several techniques designed to reduce theory perseverance, and show how well (or poorly) they have worked.

Perseverance producing processes can be classified into three broad categories. One set involves perceptual and memory processes that lead social perceivers to see stronger support for their naive theories than is actually present in the relevant social data. These are the *illusory correlation* processes. The second set involves processes that in some way change the data that are brought to bear on evaluation of a person's naive theory. These are referred to as *data distortion* processes. The third and final set involves heuristic judgment processes that tend to bolster a priori naive theories. These are the *available elements* processes. These three categories are not mutually exclusive. We view them as fuzzy sets containing processes that share family resemblances.

ILLUSORY CORRELATION

Illusory correlations are cases in which people perceive a relation between two variables when in fact there is none, or when the perceived relation is stronger than actually exists. For example, people presented with occupation-trait information tend to overestimate the frequency with which the pair "artist-creative" was presented and underestimate the frequency of "artist-wealthy" pairings (Slusher & Anderson, 1987; cf. Hamilton & Rose, 1980).

There are at least three different processes that can yield illusory correlations. One is based on the numerical distinctiveness of the different types of stimuli involved. This process does not result from use of a priori naive theories. Another process is based on associative links between the paired variables; these sometimes involve a priori naive theories. The third process is based on self-generated (imagined) events whose content is always biased by a priori naive theories.

Numerical Distinctiveness. Hamilton and Gifford (1976) pioneered the work on illusory correlations based on numerical distinctiveness (see Hamilton & Sherman, 1989, for a review). Participants read behavioral statements consisting of a name, a group designation (A or B) and a desirable or undesirable behavior. However, one group was twice as large as the other, and therefore had twice as many information statements. Because the ratio of desirable to undesirable behaviors was the same (2:1), participants should perceive that group membership (A or B) is uncorrelated with behavior. Yet, people reported liking the majority group better than the (numerical) minority group. The numerical distinctiveness explanation notes that it is relatively easier to remember undesirable behaviors by members of the numerical minority because both types of information (undesirable behaviors, minority group) stand out due to their relative infrequency. This "theory-free" illusory correlation process has important implications for theory perseverance in real world contexts. Even if there were no true differences in violent crime perpetration rates between minority and majority groups, the perceived correlation between race and violent crime garnered from news reports would be biased solely by the numerical distinctiveness of minority race and occurrence of murder.

Associative Links. In one of the earliest studies of illusory correlation, the Chappmans (Chapman, 1967; Chapman & Chapman, 1967; 1969) demonstrated that after being presented with a series of stimuli (word pairs, clinical symptoms and diagnoses), participants' frequency estimates tended to be based on a priori associations, rather than actual number of stimulus presentations.

In the social arena Hamilton and colleagues demonstrated illusory correlations based on associative links that are clearly driven by naive theories (e.g., Hamilton & Rose, 1980). In the original study, after reading about persons of various occupations and traits, participants were asked to estimate the frequency with which the various jobs were paired with the traits in the sentences. The results indicated that occupation-trait pairs that were linked by the naive theories (stereotypes) of the participant population were judged as having occurred more frequently than unlinked occupation-trait pairs (see also McConnell, Sherman, & Hamilton, 1994). The perseverance of naive theories can be based on this associative-link-based illusory correlation. Exposure to real data that contradicts the naive theory may result in little change because the perceiver "sees" the frequency data as confirmatory.

Self-Generated Data. The least studied process underlying illusory correlation has been labeled "imaginal confirmation" (Slusher & Anderson, 1987). In this pair of studies, participants were presented with a number

of occupation-trait sentences of the form, "Arthur, a wealthy lawyer, is taking a swim in his back yard pool." Some of the sentences included occupation-related traits, some included occupation-unrelated traits, and others completely lacked trait information. These trait-less sentences contained a context in which stereotype-relevant features could be imagined (e.g., "John, a lawyer, is standing in front of his house"). For each sentence, the participant's task was to imagine the described scene. Later, participants judged the frequency of all possible occupation-trait pairings. The actual frequency was the same for all pairings.

Three findings are relevant here. First, people spontaneously imagined scenes that were congruent with occupational stereotypes. For example, participants imagined that John the lawyer had an expensive house. Second, frequency judgments for stereotypic occupation-trait pairs were higher than for nonstereotypic pairs, replicating other association-link illusory correlation findings. Of most importance was the third finding: traitless imagination trials increased judged frequency. On traitless trials participants imagined stereotypic-congruent scenes. Later they could not distinguish normal trials (in which the stereotypic trait was actually presented) from traitless trials, and therefore overestimated the frequency of stereotype congruent, occupation-trait frequencies.

We suspect that this type of self-generated illusory correlation occurs with great frequency in naturalistic settings. When we read about a violent crime in a newspaper story we are likely to fill in missing information with stereotype congruent default values (e.g., race, sex, age, drug status, etc.). In fact, those with the strongest and most extreme stereotypes are also the most likely to use their default values, and to have difficulty in distinguishing fact from self-generated fiction.

DATA DISTORTION

Data distortions are instances in which some action by the perceiver—behavioral or mental—fundamentally changes the data being used to assess the validity of a naive theory. Several different processes produce such distortions, typically in a theory-congruent direction.

Behavioral Confirmation. Snyder, Tanke, and Berscheid (1977) demonstrated that a perceiver's naive-theory-based expectations can produce perceiver behaviors that elicit expectation-congruent behaviors by the target person. Male research participants expected to interact with either an attractive or unattractive woman. Phone conversations between the men and their unseen female partners were recorded. Men who believed they were interacting with an attractive woman judged their female

phone partner as having more desirable traits (e.g., sociable, poised) than did the men who expected to interact with an unattractive woman. Independent raters who listened to the audiotaped conversations rated the men in the attractive female partner condition as more sociable, sexually warm and permissive, interesting, bold, and outgoing than men in the unattractive partner condition. Independent raters of the female interactants, using only the phone recordings, rated those in the attractive condition as having more stereotypically attractive characteristics (i.e., sociability, poise, etc.) than women in the unattractive condition (see also, Snyder & Haugen, 1994). Thus, behavioral confirmation processes can lead to perseverance of naive theories by artifactually creating new data in the form of biased social interactions.

Biased Attributions and Reconstructive Memory. In two experiments, Sedikides and Anderson (1992) showed how biased attribution and reconstructive memory processes combine to produce data distortions supporting the initial naive theory. The first experiment revealed that American research participants have country-specific naive theories about defection. The second experiment examined differences in attribution about the defector. Moreover, recall was assessed both immediately and one week following exposure to the defection story. The Soviet defector was seen as more intelligent, competent, warm, moral, and sympathetic than the American defector—consistent with participants' naive theories. Participants also rated the Soviet defection as largely attributable to problems with the country, whereas the American defection was rated as largely attributable to personality problems. After one week, memory intrusions—facts that weren't actually in the story—were frequent and congruent with the defection naive theories. Data distortion, in the form these memory intrusions, made the data appear supportive of the initial naive theories.

Outcome Bias. Allison, Mackie, and Messick (1996) demonstrated that outcome knowledge—knowing how an event turned out—can also lead to data distortions. A series of studies has demonstrated that people sometimes use the outcome of an election as a diagnostic tool for inferring distribution of opinions of voters, and forget the actual distribution of the votes. Such a data omission process can obviously contribute to naive theory perseverance.

Assimilation Bias #1: Perceived Relevance and Quality of New Data. Lord and colleagues (e.g., Lord, Lepper, & Preston, 1984) have studied the effects of naive theories about the efficacy of capital punishment laws on evaluation of new data. Participants from both perspectives (pro- and anti- efficacy) were presented with two studies: one "showed" that capital punishment laws decrease murder rates, whereas the other

“showed” the opposite effect. These two sets of results were crossed with the methodologies of the studies that supposedly gave rise to them. The main finding was that those who believed that capital punishment laws effectively deter murder perceived studies supporting capital punishment as methodologically superior to studies that supposedly found no benefits of capital punishment. The opposite pattern was found for participants holding anti-capital punishment beliefs. Recent experiments have replicated this effect and ruled out several important alternative interpretations (Ratcliff, Scarborough, Lord, Ross, & Lepper, 1996). This assimilation bias distorts information by methodologically ruling out contradictory data.

Assimilation Bias #2: Perceived Covariation Strength of New Data. Perceived covariation strength is the judgment people make about how strongly two variables covary with each other. Covariation detection, however, is quite difficult, especially in the kinds of “noisy” environments characteristic of social perception (Alloy & Tabachnik, 1984; Crocker, 1981). There are problems involving identification and assessment of the independent and dependent variables, memory load problems, as well as incomplete data. Thus, it is not too surprising that distortions in covariation strength judgments are easy to find and produce in noisy environments, whether they are primarily perceptual (e.g., Lewicki, Hill, & Sasaki, 1989), verbal (e.g., Chapman, 1967), or social (e.g., Hamilton & Rose, 1980). What happens if incomplete data, memory load, and variable identification and assessment problems are eliminated?

Although several early studies found no evidence of biased assimilation under such relatively easy conditions (Anderson & Kellam, 1992; Anderson & Sechler, 1986), more recent work found clear evidence of biased assimilation when the target naive theories are strongly held. Anderson (1995) examined peoples’ perceptions of covariation in scatterplots that either clearly contradicted or clearly supported their naive implicit personality theories about how various traits go together. Participants’ covariation judgments were strongly influenced by their naive theories even though scatterplots greatly reduce task difficulty. This second type of biased assimilation of new data also helps people to maintain their naive theories in the face of contradictory evidence.

AVAILABLE ELEMENTS

This third set of perseverance producing processes involves use of some type of availability heuristic in making judgments under uncertainty (for an overview see the edited volume by Kahneman, Slovic, & Tversky,

1982).¹ In many judgment contexts, the final judgment is based (at least partly) on how easily, quickly, or frequently particular elements can be brought to mind. In different judgment contexts the type of element involved in the availability judgment varies. More importantly, in many contexts the availability of the key elements is influenced by naive theories, sometimes directly and sometimes indirectly.

Availability of Confirming Instances. This perseverance producing process involves cases in which some judgment or decision is based on how easily, quickly, or frequently specific instances of one type versus another are brought to mind. The relevance of this “confirming instances” process to perseverance of naive theories has already been discussed in the illusory correlation section, though no explicit mention of the availability heuristic was made. For instance, numerical distinctiveness makes instances of infrequent behaviors (e.g., violent crimes) stand out, especially when committed by infrequent people (e.g., numerical minorities).

Availability of Arguments. Creating arguments for one side of a theoretical or policy position can produce a resilient belief in that position. In the early theory perseverance studies (Anderson, 1983; Anderson, New, & Speer, 1985), some participants were induced to think about and explain data suggesting that risky firefighters were better at their job than their conservative counterparts. Others received the conceptually opposite “theory” and data. Even after the evidential base of these theories had been destroyed (the experimenter admitted that the original data were fictitious), participants maintained the theory that they had been randomly assigned to explore. Mediation analyses revealed that this theory perseverance effect was directly linked to the availability of causal arguments supporting vs. contradicting the experimentally induced social theory.

More recently we have found that merely explaining the *hypothetical* existence of a relation between two variables—such as the effects of covered versus uncovered food rewards on children’s ability to delay gratification—is sufficient to produce social theories that are resistant to change (Anderson & Kellam, 1992; Anderson & Sechler, 1986). Furthermore, one study (Anderson & Sechler, *Experiment 2*) demonstrated that naive theories created by a randomly assigned, hypothetical explanation

1. Within this literature, elements that first come to mind are said to be more “available” than elements that are effortful, slow, or infrequent. In other literatures, this is referred to as “accessibility,” and the term “availability” is reserved for indicating a two-level distinction between whether or not a particular element can be retrieved or brought to mind (Higgins, 1996). Because most studies in this domain use “availability” to connote the same thing as “accessibility,” we will stick with that convention here.

selected (accumulated) for testing. As we have noted, there are many ways a prior theory can produce data that essentially confirm the theory.² If the perceiver has a preferred theory in mind, judgments of relevance can be biased. This is true for old information recruited from memory, as well as for new information sought in the present situation. In other words, the motives or needs of the perceiver can influence what information is brought to bear. Data judged irrelevant (perhaps due to biased judgments of methodological quality) are not even considered. Another important feature of data accumulation is that it can also promote change. The accumulated data may strongly contradict the initial naive theory. How this happens, and the consequences of this happening, are best considered in the next stage.

In the second stage—data analysis—information brought to bear is considered, and an attempt is made to integrate it all into a plausible account. Along with other scholars working in this area we believe this stage to be relatively free of motivational biases (Anderson & Slusher, 1986; Anderson & Weiner, 1992; Cheng & Novick, 1990; Novick, Fra-tianne, & Cheng, 1992; Rusbult & Medlin, 1982). In other words, the person considers the implications of the currently available data. In the end, a tentative data-based judgment is reached.

This decision is then subjected to a test—Is it satisfactory? It may not satisfy the perceiver for any of a variety of reasons. It may not sufficiently fit the “facts” brought to bear at the data accumulation stage. It may not meet the particular goals of the perceiver. The person may dislike the conclusions for any of a variety of reasons. If so, the process cycles back to the beginning and begins again if (and only if) there is sufficient time, cognitive resources, and motivation. In the next iteration the person may modify the criteria for “accumulating” relevant data, which in turn may produce a different judgment at the conclusion of the data analysis stage.

We believe that such repeated cycles can occur quite rapidly, and that in well-practiced domains it can occur without awareness. Of course, there are cases where these processes are slow, deliberate, and predominantly conscious. Major changes in world view, such as religious conversions, which usually occur after a prolonged period of discomfort and doubt, are changes in naive theories at a most general conceptual level. For those old enough to remember various incidents in Alabama Governor George Wallace’s career (especially his stand against desegrega-

2. Of course, as also noted earlier some of these processes can lead to the creation of the initial naive theory. The distinctiveness of a biased set of cases, caused by the numerical distinctiveness or the associative links processes can operate in the Direct Learning system to produce new (but faulty) naive theories.

tion of public universities), his conversion from a diehard racist to a foe of racism is another example of this type of naive theory change.

At some point, the perseverance and change processes cease when a satisfactory conclusion is reached or because the other constraints (time, cognition, motivation) prevent further cycling. Then, the “best” version of the naive theory is adopted and used. As with all knowledge structures, naive theories may yield adaptive or maladaptive decisions, judgments, and behaviors.

SOLUTIONS TO KNOWLEDGE STRUCTURE-INDUCED BIASES

Despite the space we have devoted to the various judgmental biases, we do not mean to imply that naive theory-based judgments are necessarily *wrong*, or that judgments are always driven by naive theories. Most people successfully negotiate complex social situations most of the time, indicating that common use of naive theories is usually sufficiently accurate in everyday interactions. Furthermore, some naive theories promote attention to and corrective action for potentially biasing effects of certain context features (e.g., Wegener & Petty, 1995).

Nonetheless, there are many instances in which dependence on naive theories results in maladaptive judgments—as in cases where the judgment produces costly or harmful outcomes. For example, an educator who maintains a naive theory regarding race and intellectual abilities may be less inclined to devote time to enhancing the instruction of African-American students. In this section we apply knowledge about perseverance-producing processes to the goal of *counteracting* the negative impact of naive theories. Several strategies seem useful.

First, we note differences between formal and informal judgment contexts. Formal contexts are those in which people are explicitly engaged in making important decisions or judgments. Usually these contexts are well defined, both in terms of what kinds of evidence ought to be used, as well as what kinds of decisions are possible. The search committee is a prototypical example—the members know the main criteria for hiring decisions. Furthermore, many formal decisions are made repeatedly across time, and erroneous judgments in these contexts may be very costly (e.g., medical diagnoses).

Informal contexts are those everyday situations in which countless decisions are made, very often without awareness. These judgments may also yield costly outcomes. For example, many cases of sexual harassment stem from incorrect naive theories about how men are supposed to treat women, and how to interpret women’s responses to that type of treatment. Such informal contexts are pervasive and ambiguous. The decision makers usually are unaware that they are making judgments or

that they are using a naive theory to guide their judgments and behaviors. For this reason, reducing particular naive theory errors is more difficult in informal contexts, and must rely on changes in the underlying naive theory.

REDUCING BIAS IN INFORMAL (EVERYDAY) CONTEXTS

Inspection of the perseverance process model indicates three contexts in which bias-reducing interventions should focus. First, interventions can focus on the development and maintenance of the naive theory. In the race-biased educator example, for instance, changing the naive theory (racial stereotype) may be the best option for reducing harmful effects. Second, interventions can focus on reducing the impact of the perseverance processes on the data accumulation stage. Third, interventions can focus on preventing the person from using the naive theory in the judgment task.

Making Other Theories Available. One general strategy that can aid in all three of these endeavors is to increase the availability of an opposing naive theory. This can be done through a "counter-explanation" process in which the person imagines and explains how a different relation is (or might be) true. Several paradigms have found that this strategy works (e.g., Anderson & Sechler, 1986; Hirt and Markman, 1995; Lord, Lepper, & Preston, 1984; Snyder & Swann, 1978). Similarly, one may attack bias by increasing the availability of competing causal scenarios. We know of no studies that have directly tested this "counter-imagination" procedure, but existing research suggests that imagining alternative scenarios should work.

Finally, we have used the availability of causal explanations and scenarios to develop more effective persuasive communications involving the spread of HIV (Slusher & Anderson, 1996). Specifically, we showed that a persuasive communication designed to change beliefs about the inability of mosquitoes and casual contact to spread HIV was more effective when it included causal explanations than when it merely contained standard statements of fact. Furthermore, we showed that this greater persuasive effectiveness increased participants' willingness to commit to working with persons with AIDS.

Motivating Accuracy. A second general strategy for reducing bias in everyday judgments involves increasing the motivation of perceivers to come up with the "right" answers. Hirt & Markman (1995) and Schuette & Fazio (1995) have cited Kruglanski's (1989) *fear of invalidity* notion as a possible explanation for reliance on naive theories for judgments. Those individuals who are unconcerned with the validity of their judgments may resort to a "quick and dirty" method of assess-

ment (i.e., depend on their naive theories or attitudes) rather than conduct a more effortful evaluation of the data. Indeed, Schuette & Fazio (1995) indicated that participants who were told that their judgments were going to be evaluated by others made more stimulus-based judgments. A similar effect was found by Hirt and Markman (1995). However, increasing pressure to be accurate does not always improve accuracy (e.g., Lord et al., 1984).

ADDITIONAL STRATEGIES IN FORMAL CONTEXTS

Individuals in formal, decision-making settings also would benefit from adopting the informal-context strategies outlined earlier. But in many formal contexts, additional tools are available for reducing use of naive theories. Wittenbaum and Stasser (1995) suggest that group discussions of subjective decision tasks reduces the likelihood of any one theory-driven bias having a big impact. We recommend use of decision committees composed of people with diverse perspectives. This increases the odds that varying viewpoints and naive theories would be available.

The most effective procedures for removing naive theory effects involve removing subjective judgments of the human judges. In some contexts, potentially biasing cues such as race or gender can be removed before "paper" credentials are examined. The first author routinely removes names from applicant information prior to judging each applicant's suitability for a faculty position or for graduate admissions. Careful consideration of this example shows, however, that in some cases it is not possible to keep the judges ignorant of information deemed irrelevant but possibly biasing. A search for a senior level faculty person is difficult (almost impossible) without knowledge of the applicants' race and gender.

Many repetitive formal decisions (hiring, admissions, etc.) can be improved by using standard regression models to remove both the human judge and the theory-driven biases from the judgment processes (Dawes, 1971; Meehl, 1954). More interestingly, Dawes (1979) has demonstrated that linear models based on regression weights chosen through nonoptimal means also yield better results than does human judgment alone. One method of obtaining weights for such *improper linear models* is by "bootstrapping." This consists of asking experts to generate valid (and legal) predictor variables and to generate a set of weights for those predictors. Dawes's research has shown that predictions utilizing these weighting schemes are consistently superior to the judges' intuitive predictions. In fact, even randomly chosen weights (only the sign of the weight, \pm , is decided by the expert judges) provide

more accurate prediction of outcomes than do the intuitions of those same experts.

CONCLUSION

In most cases naive theories serve a heuristic function: They allow perceivers to quickly and effortlessly categorize events and arrive at judgments. However, when a person derives his or her naive theory from invalid data and continues to maintain that belief despite contradictory evidence, then use of such a theory in everyday and formal decision-making produces bad judgments.

We have shown how several cognitive, motivational, and behavioral processes yield belief perseverance. We have also suggested that efforts to prevent the perseverance of naive theories should attempt to increase the availability of alternate theories and to increase perceivers' fear of invalidity. In formal decision-making settings, in which individuals presumably have the time and motivation to make their decision rules explicit, we suggest that judges adopt Dawes' strategy of allowing decisions to be made statistically (Dawes, 1979).

One final note concerns the role of education in reducing biases. There is now a considerable literature on the effects of different types of training on the ability to solve everyday problems that have a statistical character (see the collection of articles by Nisbett, 1993). For instance, both graduate and undergraduate psychology students improve more on everyday statistical reasoning abilities than do natural science majors. Even short term "training" sessions can improve reasoning skills of this type. Therefore, our final recommendation for reducing naive theory biases in judgments within all contexts is to improve our society's general reasoning ability by including more statistics and reasoning content in public schools.

REFERENCES

- Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36*, 715-729.
- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review, 96*, 358-372.
- Allison, S. T., Mackie, D. M., & Messick, D. M. (1996). Outcome biases in social perception: Implications for dispositional inference, attitude change, stereotyping, and social behavior. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 53-93). New York: Academic Press.
- Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review, 91*, 112-149.
- Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories:

- When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19*, 93-108.
- Anderson, C. A. (1995). Implicit personality theories and empirical data: Biased assimilation, belief perseverance and change, and covariation detection sensitivity. *Social Cognition, 13*, 25-48.
- Anderson, C. A., & Godfrey, S. (1987). Thoughts about actions: The effects of specificity and availability of imagined behavioral scripts on expectations about oneself and others. *Social Cognition, 5*, 238-258.
- Anderson, C. A., & Kellam, K. L. (1992). Belief perseverance, biased assimilation, and covariation detection: The effects of hypothetical social theories and new data. *Personality and Social Psychology Bulletin, 18*, 555-565.
- Anderson, C. A., Krull, D. S., & Weiner, B. (1996). Explanations: Processes and consequences. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic mechanisms* (pp. 271-296). City: Publisher.
- Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037-1049.
- Anderson, C. A., New, B. L., & Speer, J. R. (1985). Argument availability as a mediator of social theory perseverance. *Social Cognition, 3*, 235-249.
- Anderson, C. A., & Riger, A. L. (1991). A controllability attributional model of problems in living: Dimensional and situational interactions in the prediction of depression and loneliness. *Social Cognition, 9*, 149-181.
- Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50*, 24-34.
- Anderson, C. A., & Sedikides, C. (1991). Thinking about people: Contributions of a typological alternative to associationistic and dimensional models of person perception. *Journal of Personality and Social Psychology, 60*, 203-217.
- Anderson, C. A., & Slusher, M. P. (1986). Relocating motivational effects: An examination of the cognition-motivation debate on attributions for success and failure. *Social Cognition, 4*, 270-292.
- Anderson, C. A., & Weiner, B. (1992). Attribution and attributional processes in personality. In G. Caprara & G. Heck (Eds.), *Modern personality psychology: Critical reviews and new directions* (pp. 295-324). New York: Harvester Wheatsheaf.
- Anderson, K. B., Anderson, C. A., Dill, K. E., & Deuser, W. E. (in press). The interactive relations between trait hostility, pain, and aggressive thoughts. *Aggressive Behavior*.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Berkowitz, L., & LePage, A. (1967). Weapons as aggression-eliciting stimuli. *Journal of Personality and Social Psychology, 7*, 202-207.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177-220.
- Cacioppo, J. T., Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.
- Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Behavior, 6*, 151-155.
- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology, 72*, 193-204.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271-280.

- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545-567.
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101.
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90*, 272-292.
- Dawes, R. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26*, 180-188.
- Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Dill, K. E., Anderson, C. A., Anderson, K. B., & Deuser, W. E. (1997). Effects of aggressive personality on social expectations and social perceptions. *Journal of Research in Personality, 31*, 272-292.
- Dovidio, J. F., Evans, N. E., & Tyler, R. B. (1986). Racial stereotypes: the contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22-37.
- Dweck, C. S., Hong, Y., & Chiu, C. (1993). Implicit theories: Individual differences in the likelihood and meaning of dispositional inference. Special issue: On inferring personal dispositions from behavior. *Personality and Social Psychology Bulletin, 19*, 644-656.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509-517.
- Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology, 43*, 89-99.
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392-407.
- Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology, 39*, 832-845.
- Hamilton, D. L., & Sherman, S. J. (1989). Illusory correlations: Implications for stereotype theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 59-82). New York: Springer-Verlag.
- Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49*, 129-134.
- Higgins, E. T. (1996). Knowledge activation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford.
- Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69*, 1069-1086.
- James, W. (1890). *The principles of Psychology*. New York: Holt, Rinehart & Winston.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. London: Cambridge University Press.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum Press.
- Levy, S. R., & Dweck, C. S. (1998). Trait- versus process-focused social judgment. *Social Cognition, 16*, 151-172.

- Lewicki, P., Hill, T., & Sasaki, I. (1989). Self-perpetuating development of encoding biases. *Journal of Experimental Psychology, 118*, 323-337.
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*, 1231-1243.
- Mackie, D. M., & Allison, S. T. (1987). Group attribution errors and the illusion of group attitude change. *Journal of Experimental Social Psychology, 23*, 460-480.
- McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). On line and memory based aspects of individual and group target judgments. *Journal of Personality and Social Psychology, 67*, 173-185.
- Meehl, P. E. (1954). *Clinical versus statistical predictions A theoretical analysis and revision of the literature*. Minneapolis: University of Minnesota Press.
- Nisbett, R. E. (Ed.). (1993). *Rules for reasoning*. Hillsdale, NJ: Erlbaum.
- Novick, L. R., Fratianne, A., & Cheng, P. W. (1992). Knowledge-based assumptions in causal attribution. *Social Cognition, 10*, 299-334.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: Brown.
- Petty, R. E., Wegener, D. T., & White, P. H. (1998). Flexible correction processes in social judgment: Implications for persuasion. *Social Cognition, 16*, 93-113.
- Ratcliff, C. D., Scarberry, N. C., Lord, C. G., Ross, L., & Lepper, M. R. (1996). *Biased assimilation, attitude polarization, and behavior polarization: Lord, Ross, and Lepper (1979) revisited*. Unpublished manuscript.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 3*, 382-407.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880-892.
- Rusbult, C., & Medlin, S. (1982). Information availability, goodness of outcome, and attributions of causality. *Journal of Experimental Social Psychology, 18*, 292-305.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79*, 294-324.
- Schutte, R. A., & Fazio, R. H. (1995). Attitude accessibility and motivation as determinants of biased processing: A test of the MODE model. *Personality and Social Psychology Bulletin, 21*, 704-710.
- Sedikides, C., & Anderson, C. A. (1992). Causal explanations of defection: A knowledge structure approach. *Personality and Social Psychology Bulletin, 18*, 420-429.
- Skitka, L. J., & Tetlock, P. E. (1993). Providing public assistance: Cognitive and motivational processes underlying liberal and conservative policy preferences. *Journal of Personality and Social Psychology, 65*, 1205-1223.
- Slusher, M. P., & Anderson, C. A. (1987). When reality monitoring fails: The role of imagination in stereotype maintenance. *Journal of Personality and Social Psychology, 52*, 653-662.
- Slusher, M. P., & Anderson, C. A. (1996). Using causal persuasive arguments to change beliefs and teach new information: The mediating role of explanation availability and evaluation bias in the acceptance of knowledge. *Journal of Educational Psychology, 88*, 110-122.
- Sniderman, P., Hagen, M. G., Tetlock, P. E., & Brady, H. E. (1986). Reasoning chains: Causal

- models of policy reasoning in mass publics. *British Journal of Political Science*, 16, 405-430.
- Snyder, M., & Haugen, J. A. (1994). Why does behavioral confirmation occur? A functional perspective on the role of the perceiver. *Journal of Experimental Social Psychology*, 30, 218-246.
- Snyder, M., & Swann, W.B., (1978). Hypothesis testing processes in social interaction. *Journal of Personality and Social Psychology*, 36, 1202-1212.
- Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, 35, 656-666.
- Wegener, D. T., & Petty, R. T. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68, 36-51.
- Wittenbaum, G. M., & Stasser, G. (1995). The role of prior expectancy and group discussion in the attribution of attitudes. *Journal of Experimental Social Psychology*, 31, 82-105.
- Wittenbriink, B., Hilton, J. L., & Gist, P. L. (1998). In search of similarity: Stereotypes as naive theories in social categorization. *Social Cognition*, 16, 31-55.
- Yzerbyt, Leyens, & Corneille, 1997
- Zimbardo, P. G., & Leippe, M. R. (1991). *The psychology of attitude change and social influence*. New York: McGraw-Hill.