Psychological Bulletin
1994, Vol. 116, No. 1, 117-142

0033-2909/94/$3.00

Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations

Timothy D. Wilson and Nancy Brekke

We define mental contamination as the process whereby a person has an unwanted response because of mental processing that is unconscious or uncontrollable. This type of bias is distinguishable from the failure to know or apply normative rules of inference and can be further divided into the unwanted consequences of automatic processing and source confusion, which is the confusion of 2 or more causes of a response. Mental contamination is difficult to avoid because it results from both fundamental properties of human cognition (e.g., a lack of awareness of mental processes) and faulty lay beliefs about the mind (e.g., incorrect theories about mental biases). People's lay beliefs determine the steps they take (or fail to take) to correct their judgments and thus are an important but neglected source of biased responses. Strategies for avoiding contamination, such as controlling one's exposure to biasing information, are discussed.

To survive in today's world, we have to be aware of a great many physical contaminants and do our best to avoid or neutralize them. The air we breathe is fouled by industrial pollutants, automobile emissions, and radon gas. The water many of us drink is contaminated by lead from old city pipes or pollutants seeping into groundwater. Many foods are contaminated with pesticides and hormones. We are bombarded with electromagnetic radiation from computer screens and innocent-looking household appliances. Incurable diseases, such as acquired immunodeficiency syndrome (AIDS), are spreading.

The topic of this article is *mental contamination.* According to the *Random House Dictionary* (1968), to contaminate means "to render impure or unsuitable by contact or mixture with something unclean, bad, etc." (p. 289). In other words, a desirable state is rendered less desirable by contact with some unwanted agent. We believe that many mental errors are profitably viewed as tainted by the intrusion of internal or external agents. More formally, we define mental contamination as the process whereby a person has an unwanted judgment, emotion, or behavior because of mental processing that is unconscious or uncontrollable. By unwanted, we mean that the person making the judgment would prefer not to be influenced in the way he or she was (we expand on this definition shortly).[1]

For example, teachers would rather not give a student a high grade because the student is physically attractive, yet there have been repeated demonstrations of such halo effects (e.g., Landy & Sigall, 1974). Most would prefer that their consumer decisions be unaffected by advertising for products because they know that advertisements are often one sided or false. Yet, there is ample evidence that advertising has powerful effects (e.g., Abraham & Lodish, 1990; Liebert & Sprafkin, 1988; Ryan, 1991). It is unlikely that people would want their decision about whether to act cooperatively or competitively toward another person to be influenced by a news broadcast they had just heard about a prosocial or antisocial act, and yet there is a good deal of evidence for such priming effects (e.g., Hornstein, LaKind, Frankel, & Manne, 1975).

In this article, we classify and evaluate research on mental contamination. Our review is meant to be more illustrative than exhaustive in that our goal is to reconceptualize the area of mental bias rather than simply to review or summarize it. Thus, we do not catalogue each of the hundreds of studies in this vast area of research, but we discuss the major findings in terms of our reconceptualization.

Our argument can be summarized as follows: The analogy of contamination is a useful one for the study of mental bias for two main reasons. First, it focuses attention on the difficulty of avoiding many biases. Something that is contaminated is not easily made pure again, which we believe is an apt metaphor for many mental biases. We argue that, because of a lack of awareness of mental processes, the limitations of mental control, and the difficulty of detecting bias, it is often very difficult to avoid or undo mental contamination. By emphasizing the ubiquity of

The writing of this article and the research it reports were supported by National Institute of Mental Health Grant MH41841 to Timothy D. Wilson and National Science Foundation Grant SES-8708007 to Nancy Brekke. We would like to thank Kate Etling, Daniel Gilbert, Sara Hodges, Chris Houston, Suzanne LaFleur, Steven Neuberg, Richard Nisbett, Fritz Strack, and several anonymous reviewers who provided valuable insights on a draft of this article.

Correspondence concerning this article should be addressed either to Timothy D. Wilson, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903-2477 (electronic mail: tdw@virginia.edu), or to Nancy Brekke, Department of Psychology, Lake Forest College, 555 North Sheridan Road, Lake Forest, Illinois 60045 (electronic mail: brekke@lfmail.lfc.edu).

[1] We should note that we are using the term *contamination* differently than Rozin's use of the similar terms *contagion* and *to contaminate* (Rozin, 1990; Rozin & Nemeroff, 1990). Rozin has focused on people's primitive beliefs about contagion between physical objects, such as the belief that when two objects come into contact, the properties of one object are transferred to the other object. In contrast, we are concerned with the extent to which people's thoughts, feelings, and beliefs are contaminated by unconscious or uncontrollable mental processes.

mental contamination, we stand in contrast to some who have questioned the generalizability or seriousness of mental errors documented in the laboratory (e.g., Cohen, 1981).

Second, the contamination metaphor focuses attention on the strategies people use to avoid mental biases. Just as people are motivated to take steps to avoid pollution of the physical environment, so are they motivated to avoid mental pollution. To predict whether people will have a biased judgment, we suggest, it is necessary to understand the strategies people use to protect their minds. For example, to understand the extent to which people are influenced by advertising, it is important to consider the steps they take to avoid being influenced, such as switching the channel when an advertisement comes on. Clearly, people who engage in such a strategy will be less influenced than people who do not. Similarly, to predict whether a college professor will be biased by the physical attractiveness of a student when grading the student's paper, it is important to consider the steps the professor takes to avoid such a bias, such as grading the paper while unaware of who wrote it. Professors who grade in such a way will be less influenced by the attractiveness of their students than those who do not.

What determines the strategies people adopt to avoid mental contamination? We suggest that these strategies stem in large part from lay beliefs about the mind, such as people's beliefs about their susceptibility to such things as advertising and halo effects. The study of lay beliefs about mental bias can be considered to be part of the growing field of metacognition, which has examined people's cognitions about their cognitions (Nelson, 1992; Wellman & Gelman, 1992). Most of this literature has focused on the development of beliefs about the mind, such as how children's understanding of their memories changes with age (e.g., Flavell, 1979). One of our goals is to extend this literature to a consideration of adults' beliefs about mental errors, namely people's theories about the ways in which their judgments can be biased. These beliefs, we suggest, influence the strategies people use to avoid mental biases, as well as the success of these strategies.

We begin with a discussion of what mental contamination is not, that is, how this topic fits into a broader scheme of biases in human judgment. We then discuss the determinants of mental contamination in some detail, including a discussion of different types of mental contamination.

Failure of Rule Knowledge and Application

In the last 20 years, cognitive and social psychologists have documented numerous errors in reasoning, biases in judgment, and flawed heuristics (e.g., Arkes, 1991; Baron, in press; Einhorn & Hogarth, 1978; Kahneman, Slovic, & Tversky, 1982; Kruglanski, 1989b; Nisbett & Ross, 1980; Ross, 1977; Sherman & Corty, 1984; Tversky & Kahneman, 1974). We suggest that the numerous instances of biases in human reasoning are of two general types: those that result from the failure to know or apply an explicit rule of inference—the *failure of rule knowledge or application*—and those that result from *mental contamination* (cases whereby a judgment, emotion, or behavior is biased by unconscious or uncontrollable mental processes). In the first case, people are inappropriately influenced by information (e.g., they draw inferences from very small samples of data) because they are unaware of the proper inferential rules (e.g., the law of large numbers). Thus, they might not recognize that their judgment is in any way biased or inappropriate as a result of their ignorance of the proper rule. In the second case, people know that they do not want a stimulus to influence their judgment (e.g., a student's physical attractiveness influencing the grade they assign to the student's paper), but they are either unaware that they have been influenced or unable to avoid the influence. That is, people know how they want to form a judgment but nonetheless end up biased. We begin by discussing the first of these causes of biased judgment: the failure to know or apply a useful rule or algorithm.

There are many demonstrations of cases whereby human reasoning goes wrong because of an ignorance or misapplication of normative rules of inference. We refer to rules that are consciously known and can be deliberately applied, such as the principle of "sunk costs," according to which only future costs and benefits, rather than past costs, should be considered when making a choice (Arkes & Blumer, 1985; Mishan, 1976; Morgan & Duncan, 1982). People who have never learned these rules or who know them but fail to apply them correctly will make more judgmental errors than people who know the rules and when to apply them (Larrick, Morgan, & Nisbett, 1990).

Such rules, we suggest, are like those people learn in other, problem-solving domains such as mathematics. Consider people's knowledge about how to compute the area of a parallelogram (Baron, in press). People who know that the correct formula is to multiply the base by the height will make fewer errors than people who do not know this rule. Of course, even if a rule is learned, people have to know when it applies and when it can be generalized to similar problems. As noted by Wertheimer (1959), people often fail to apply the base-by-height rule to parallelograms that are rotated such that their height exceeds their width, and they apply the rule inappropriately to figures such as trapezoids. Nonetheless, the learning of such rules increases the accuracy of relevant judgments. Similarly, people who have learned the principle of sunk costs are less likely to make certain kinds of errors in reasoning than those who have not (Larrick et al., 1990). There are many examples of such explicit rules in the realm of human judgment, some of which are listed in Appendix A. In each case, human inference suffers when people do not apply an inferential rule either because they do not know the rule or fail to apply it properly.

If the inferential rules summarized in Appendix A can be consciously learned and deliberately applied, it ought to be possible to teach people the rules and when to apply them. This assumption is commonly made in the literature on debiasing (e.g., Arkes, 1991; Einhorn & Hogarth, 1978; Fischhoff, 1982; Kahneman & Tversky, 1982) and has been confirmed in several studies by Nisbett and his colleagues (e.g., Fong, Krantz, & Nisbett, 1986; Larrick et al., 1990; Nisbett, Fong, Lehman, & Cheng, 1987). For example, Nisbett and his colleagues have studied people's understanding of the law of large numbers and people's knowledge of when to apply this rule to everyday inferences about the social world. This rule is similar to learning how to compute the area of a parallelogram, we suggest, in that it is deliberately applied to everyday problems.[2] Nisbett and his

[2] We do not mean to imply that these rules are completely conscious

colleagues have found that people do appreciate this rule to some extent but that it is possible to improve human reasoning by explicitly teaching people when to apply the rule (e.g., Fong et al., 1986; Larrick et al., 1990; Lehman, Lempert, & Nisbett, 1988; Nisbett et al., 1987).

We do not mean to minimize the difficulty of knowing exactly when to apply such rules. Consider the use of base-rate information when making probabilistic judgments. There is some controversy over exactly which base rates are appropriate when making many social judgments and exactly how to integrate these base rates with individuating information about a stimulus (e.g., Birnbaum, 1983; Cohen, 1977). Our point is that despite the difficulty of teaching reasoning, it is bound to be easier to teach people to use conscious rules that can be applied deliberately than to teach them to recognize and correct problems in mental processing that occur outside of awareness. That is, there is a large class of mental procedures, to which we now turn, that are much more difficult to verbalize and control, with important implications for how easy it is to avoid mental errors.

Mental Contamination

Earlier we defined mental contamination as the process whereby a person ends with an unwanted judgment, emotion, or behavior because of mental processing that is unconscious or uncontrollable (again, "unwanted" in the sense that the judgment maker would prefer not to be influenced by the mental processes in question). The process of mental contamination is portrayed in Figure 1. Later we elaborate on each stage of this diagram at some length; for now, consider the following example: Professor Jones is grading papers from a small seminar. When grading Hernandez's paper, undesirable mental processes are triggered by the fact that Jones dislikes Hernandez and knows that Hernandez is a member of a minority group. That is, Jones's dislike and prejudice taints her evaluation, such that the evaluation is more negative than it would otherwise be. Furthermore, assume that Jones would agree that this is unfair and would prefer not to be influenced by her prejudice. This example, then, fits our definition of mental contamination in that Jones's judgment was influenced by unwanted agents (her prejudice).

The first thing to note about this example is that it does not easily fit into our category of the failure to know or apply an algorithm. At a very general level, Jones knows the rule "Don't be biased when grading papers," but this is not a rule that is easily applied in the way that the law of large numbers or sunk costs can be. That is, the rule does not specify a set of procedures by which one can improve the judgment such that teaching people this rule is unlikely to reduce mental errors. In all of our examples of mental contamination, there are no specific algorithms that, if known, can be applied to correct one's responses. Jones's biased judgment stems not from an ignorance of a rule but from a lack of awareness that her judgment is biased or the inability to control this bias.

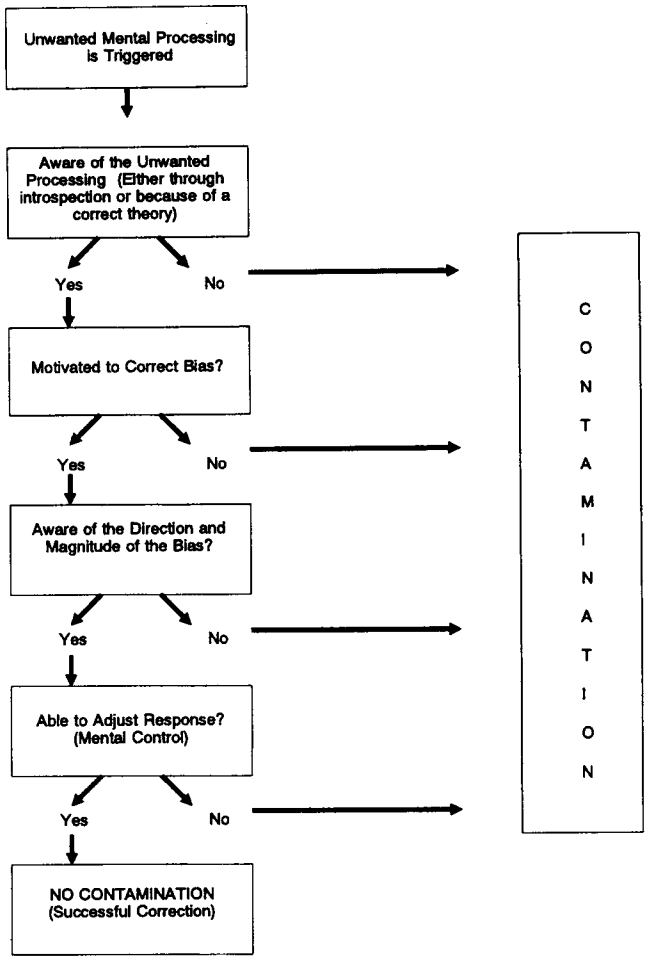

Figure 1. The process of mental contamination and mental correction.

More specifically, the processes responsible for mental contamination are shown in Figure 1. To avoid a contaminated judgment, people must satisfy four conditions. First, they must be aware of the unwanted mental process. Such awareness can occur through direct introspective access to the process; for example, Jones might directly observe the fact that her dislike for Hernandez is tainting the evaluation. In the absence of direct access, people might be aware of this process because of a theory they have learned (Nisbett & Wilson, 1977b). Even if Jones cannot directly observe the cognitive processing leading to a prejudiced judgment, she might suspect that it is occurring because of a theory about halo effects.

Second, people must be motivated to correct the error. Jones may know that her grade is biased and prefer that it not be bi-

and verbalizable. A person could recognize that a sample of 5 people is likely to yield more biased results than a sample of 100 people without fully verbalizing the law of large numbers. These rules can operate at the intuitive level of statistical heuristics (Nisbett, Krantz, Jepson, & Kunda, 1983), even if people are unable to state the rules in formal terms. Nonetheless, these rules are more easily verbalized and more able to be applied deliberately than the processes contributing to mental contamination, as we discuss shortly. Indeed, one of the dependent measures in many of the Nisbett studies on teaching reasoning was the extent to which people's explanations of their reasoning mention the appropriate rule (e.g., Nisbett et al., 1983). Thus, it is useful to consider them more conscious and more controllable in relative terms.

ased but believe that it is not worth her time and energy to try to correct it. Third, even if motivated to correct the error, people must be aware of the direction and magnitude of the bias. Jones may know that the grade is biased but have little knowledge of how much ("Should I change my grade from a C to a B−? Or to a B+?"). Finally, even if Jones has satisfied the first three conditions (awareness of the unwanted processing, motivation to correct it, and knowledge of the direction and extent of the bias), she might not have sufficient control over her responses to be able to correct the unwanted mental processing. Jones may know that biased processing has led to a lowering of Hernandez's grade from a B to a C but may simply be unable to escape the impression that the paper is truly mediocre.

As implied by this example, we are rather pessimistic about people's ability to avoid or correct for mental contamination. We suggest that the nature of human cognition, as well as the nature of lay theories about the mind, makes it difficult to satisfy all of the conditions necessary to avoid contaminated responses. In the following sections, we review research that speaks to each of these claims, beginning with fundamental properties of human cognition. Before beginning this review, however, we need to take a slight detour to consider the question of the definition of biased judgment.

Definitions of Mental Bias

It is important to note that our two general classes of mental errors—the failure to know or apply a rule and mental contamination—use different definitions of bias. The first class uses normative standards, defined as "consensus among formal scientists that the rule is appropriate for the particular problem" (Nisbett & Ross, 1980, p. 13). If people do not use the rule that experts agree is correct, their responses are called into question. By contrast, we used a subjective criterion to define mental contamination: A judgment, emotion, or behavior is said to be contaminated if it was influenced in an "unwanted" way (i.e., unwanted by the person whose judgment, emotion, or behavior it is).

There has been considerable debate over the viability of each of these types of definitions of mental errors (e.g., Baron, in press; Cohen, 1981; Conee & Feldman, 1983; Einhorn & Hogarth, 1981; Funder, 1987; Goodman, 1965; Hastie & Rasinski, 1988; Kahneman & Tversky, 1983; Kruglanski, 1989b; Nisbett & Ross, 1980; Stich, 1990; Stich & Nisbett, 1980; Swann, 1984). Some theorists prefer normative standards (e.g., Baron, in press; Nisbett & Ross, 1980), whereas others prefer more subjective, pragmatic definitions, similar to our definition of mental contamination (e.g., Kahneman & Tversky, 1983; Stich, 1990).

The criterion of accuracy adopted, we suggest, depends in part on the goals of the researcher. If the goal is to make definitive statements about which judgments are inaccurate, according to the most up-to-date normative models of human inference and statistical reasoning, then it is best to compare human reasoning with the standards of experts on reasoning. This goal is generally the one adopted by researchers on rule knowledge and application. The rules that are assessed or taught are those that experts argue are advantageous for people to use (e.g., the rule of sunk costs or knowledge of the gambler's fallacy). If a subjective criterion of accuracy were to be used, whereby people defined for themselves the rules of inference that were most appropriate, it would be difficult to prescribe ways in which human inference could be improved. For example, there would be no grounds to question the judgment of someone who believed in the gambler's fallacy, even if there were consequential, negative behaviors that followed from the use of this principle (Stich & Nisbett, 1980).

A disadvantage of using a normative criterion, however, is that it shifts attention away from lay beliefs about bias and from the defensive strategies people use to protect themselves from bias. As noted earlier, one of our central points is that to understand when and why human judgments go wrong, one must understand both the nature of the cognitive processes producing biased judgments and the defensive strategies people adopt when they believe that their judgments might be contaminated. To study these strategies, one must consider people's beliefs about how valid their responses are (i.e., with the subjective criterion of bias). If people do not believe that their responses are biased, they are unlikely to invoke strategies to avoid or undo mental bias, even if the experts find fault with their judgments. Conversely, if people are concerned about a source of bias on their judgments, they might well take consequential steps to overcome this bias, even if their judgments were perfectly acceptable according to normative standards.

In short, just as theories of social perception are concerned with people's beliefs about the social world—regardless of their accuracy—we are concerned with people's beliefs about the validity of their judgments, regardless of whether these beliefs correspond to normative standards of judgment. Beliefs about one's judgments, we hope to show, have consequential effects on people's behavior, independent of their accuracy. In this sense, our definition of bias is a pragmatic one in that it is concerned with the utility of inferences to the person making the inferences (McArthur & Baron, 1983; Stich, 1990; Swann, 1984).

Using a subjective criterion has a potential disadvantage: There may be large individual differences in beliefs about mental bias, making it difficult to study and make generalizations about specific mental phenomena. If everyone uses a different standard to judge his or her mental errors, it would be difficult to build theories about the nature of biased reasoning. An inference that is abhorrent to one person might be endorsed with enthusiasm by another. However, this objection may not be as serious as it seems. As noted by Kahneman and Tversky (1983), many of the shortcomings in human inference that have been demonstrated actually meet both criteria: They are wrong according to normative standards, and respondents, "after some reflection or discussion" (p. 509), agree that the judgments are wrong. Thus, in many areas there may not be wide individual differences in beliefs about what constitutes a valid judgment.

Even when there are individual differences in lay beliefs about contamination, we believe that to predict whether a judgment will be biased, it is necessary to take these beliefs into account. Suppose, for example, that one wants to predict whether a college professor will exhibit a halo effect, whereby his or her grade is biased by a student's physical attractiveness. To make such a prediction, it is necessary to consider the professor's beliefs about the appropriateness of halo effects. A professor who believes that such halo effects are improper is more likely to take

steps to avoid them (e.g., by grading papers blindly). Thus, without understanding people's personal definitions of bias, one cannot predict the strategies they will take to debias themselves, strategies that may well change the nature of their judgment. We discuss people's lay beliefs about contamination and their defensive strategies shortly. First, however, we need to consider some fundamental properties of human cognition that contribute to mental contamination, independent of people's lay beliefs.

Causes of Mental Contamination: Properties of Human Cognition

Awareness of unwanted processing: The difficulty of detecting mental contamination. One reason we have used the term *contamination* is to draw an analogy to contamination in the physical realm (e.g., to demonstrate that people's thoughts and feelings can become contaminated in much the same way as can their bodies by viruses and bacteria). One difference between physical and mental contamination, however, concerns their ease of detection. It is much more difficult to detect mental contamination, particularly at the individual level.

One impediment to such a recognition is people's limited access to their mental processes. We do not have the space to review the literature on introspective access to mental processes completely or to document the controversies that have ensued over the issue of awareness (e.g., Ericsson & Simon, 1980; Nisbett & Wilson, 1977b; Smith & Miller, 1978). Nor do we wish to exaggerate claims made about what is inaccessible to conscious processing (see Wilson & Stone, 1985, for a related discussion). Despite the stormy history of the literature on unconscious processing, the idea that people are unaware of a substantial amount of their mental processing has a firmer toehold in social and cognitive psychology than ever before (Jacoby, Lindsay, & Toth, 1992; Kihlstrom, 1987; Lewicki, 1986; Posner & Rothbart, 1989). Because of this limited access, mental processes leading to contamination can occur unobserved.

For example, when teachers assign a C to a student's paper, they probably believe that they have given it a fair and unbiased evaluation, even if they were biased by how much they like the student (Nisbett & Wilson, 1977a). When people decide that someone is adventurous instead of hostile, they are unaware that their judgment is "contaminated" by the fact that they had just memorized words having to do with adventurousness (Higgins, Rholes, & Jones, 1977). When they form an evaluation of someone, what they experience subjectively is usually the final product (e.g., "This guy is pretty attractive"), not the mental processes that produced this product, such as the operation of a halo effect (e.g., people do not consciously think, "Well, I like this guy, so I guess I'll boost my perception of how attractive he is"; see Nisbett & Wilson, 1977a).

The same is often true, of course, in the physical realm: People cannot directly observe the processes that cause physical contamination, such as the invasion of their cells by rhinoviruses. In the physical realm, however, there are often observable symptoms of contamination, even when the process of contamination is unobservable. Although people cannot observe rhinoviruses, a stuffed-up nose tells them they have a cold. If one is wondering whether a gallon of milk is fresh or spoiled, a quick whiff will reveal the answer. There are seldom such observable symptoms, such as smell, temperature, or physical appearance, indicating that a human judgment is contaminated. As a result, people are often unaware that their judgment is "spoiled," in Jacoby and Kelley's (1987) terms. Human judgments—even very bad ones—do not smell.

Sometimes, we acknowledge, there are clues that a judgment has been contaminated. This is most likely when people have an unexpected reaction to a stimulus. If Jim enjoys a movie much less than he expected, he might be suspicious that he was influenced by the people behind him slurping their drinks. If one of Jane's students has been doing very poorly and Jane gives his or her term paper the highest mark in the class, she might be suspicious that she was biased by the fact that she was very fond of the student. These clues are not definitive, however, and can be misleading. Jim has no way of knowing how he would have felt about the movie in the absence of slurping noises, and Jane has no way of knowing what grade she would have given the student if she did not like that student. Furthermore, it is likely that many cases of mental contamination do not have such large effects that people have very unexpected reactions. It is probably more common for the effects to be subtle (e.g., giving a favorite student a B instead of a C). In these cases, the bias is likely to go completely unnoticed, without a moment's doubt about the validity of the judgment.

In the physical realm, of course, some forms of contamination are also symptomless. Some diseases have virtually no symptoms, and some chemical and biological contaminants are difficult to detect. One cannot open the basement door and see radon gas seeping out; nor are there observable signs that tap water has dangerously high levels of lead. Our point is that there are almost never symptoms of mental contamination, whereas such signs often accompany physical contamination. Even when they do not, instruments have been developed to measure many kinds of physical contamination. One can put a radon detector in the basement and send off samples of tap water to labs that will test the lead content. Unfortunately, there are few such devices that measure how much people's judgments and inferences are biased.

The only device available is the experimental method, whereby people are randomly assigned to a condition in which they are exposed to a potential contaminant or a control condition in which they are not. By comparing the average judgments of the two groups, researchers can determine whether a certain kind of information has affected people's judgments. Indeed, this method was used in all of the documented cases of mental contamination we describe in this article. Theoretically, it would be possible, using within-subjects designs with many trials, to assess how an individual is affected by a particular type of information. One could, for example, have people make repeated judgments of other individuals, varying the kinds of primes they received before making their judgments. By averaging across trials, the extent to which each person was influenced by the primes could be estimated.

Within-subjects designs, however, require that the researcher average each person's judgments over several examples of the stimulus (e.g., of the target people being rated in a priming experiment), such that it is impossible to assess how much one judgment of one stimulus is contaminated. Essentially, re-

searchers have two choices: They can average across people to estimate the effects of an experimental manipulation on judgments of one stimulus (using a between-subjects design), or they can average across stimuli to estimate the effects of an experimental manipulation on one person's judgments (using a within-subjects design). They can detect mental contamination in the "average person" on judgments of one stimulus or contamination in one person on judgments of the "average" stimulus. They cannot detect mental contamination in one person's judgment of one stimulus.

Even the procedures available, of course, require a good deal of control, usually in the laboratory. They are of little help to the layperson trying to untangle what influenced his or her judgments in everyday life (e.g., why he or she gave a B to a favored student's term paper). In everyday life people do not have the proper control conditions, with random assignment, that would enable them to determine how biased their judgments are, even in the aggregate. Thus, the experimental method is of little help to the layperson who is trying to determine the extent to which a particular judgment is biased. The fact that mental contamination is difficult to detect makes it hard to avoid or eliminate, for the simple reason that if people are unaware that their judgment is biased, they will not try to debias it.[3]

The difficulty of mental control. Even if people are aware that an unwanted mental process is tainting their judgments and recognize the direction and magnitude of the resulting bias, they need to be able to control their responses sufficiently to correct the bias (see the fourth stage in Figure 1). Unfortunately, however, a considerable amount of recent research suggests that people's ability to control their thoughts and feelings is limited (Wegner & Pennebaker, 1993). Although people can stop certain kinds of mental operations from occurring, such as simple arithmetic operations (Logan, 1988, 1989), many other kinds of mental processes are extremely difficult to control once they are set in motion, especially those that occur outside of awareness (Bargh, 1989; Logan, 1989; Wegner, 1989, 1992). Thus, avoiding mental contamination by stopping a sequence of thoughts or mental operations is unlikely to be a consistently successful strategy.

Once a mental process has run its course, resulting in a belief (e.g., "Hernandez's paper is of C quality") or feeling (e.g., sadness), it is very difficult to erase that belief or feeling. Beliefs and feelings change as people gain new information, of course. Reading a glowing letter of recommendation can change one's mind about a job applicant, and sadness can change to euphoria on the receipt of good news. It is very difficult, however, to erase a belief or feeling, in the absence of new information, by sheer will. One cannot look at a watch and say, "When the second hand reaches 12, I will believe that the wall is yellow instead of green and that the sun revolves around the earth." In William James's words, "We can *say* any of these things, but we are absolutely impotent to believe them" (1897, p. 5).

Furthermore, even if the original basis for a belief is discredited, the causal explanations people construct to support the belief persist, leading to belief perseverance (e.g., Anderson, 1989; Anderson, Lepper, & Ross, 1980; Ross, Lepper, & Hubbard, 1975). Thus, once one has formed a belief, it can be very difficult to erase. Rather than trying to "manage" beliefs through mental control, it might be better to manage them by controlling one's exposure to new information that confirms or disconfirms the belief (Price, 1954). As is shown later, however, it is difficult for people to act as self-censors, avoiding all information that challenges a cherished belief, and it is difficult for them to control the effects that new information has on their beliefs (Gilbert, 1991, 1993).

Perhaps it is easier to inhibit overt behavior than thoughts and feelings (Devine, 1989a; Fiske, 1989; Logan, 1989). We may not be able to avoid a stereotypical or prejudiced thought, but we can stop ourselves from acting on it. Unfortunately, however, even overt behavior can sometimes be difficult to control. People with compulsive disorders find it very difficult to stop performing ritual acts, and many people are not as successful at controlling their impulses (e.g., their urge to eat or drink) as they would like. Nonetheless, it is easier to inhibit overt action than it is to stop a thought or a mental process. We return to this point when we discuss strategies for avoiding mental contamination.

Summary. It should be possible to avoid or correct mental contamination if people successfully complete all of the steps depicted in Figure 1: awareness of the unwanted processing, the motivation to correct it, awareness of the direction and magnitude of the bias, and the ability to control one's responses sufficiently. We argue that it is difficult to satisfy these conditions, in part because of fundamental properties of human cognition: People are unaware of many of their cognitive processes, mental contamination often has no observable "symptoms," and people have limited control over their cognitive processes. These facts alone are cause for considerable pessimism about people's ability to avoid unwanted judgments. As discussed earlier, however, we need to consider lay beliefs about the mind as well, because these beliefs also influence the extent to which people will satisfy the conditions shown in Figure 1. For example, in the absence of direct introspective access to mental processes that lead to contamination, people could possess accurate theories about when these processes occur and how they bias their judgments. We turn now to a consideration of these lay beliefs about mental bias.

Causes of Mental Contamination: Lay Beliefs About the Mind

Are people concerned about mental contamination? An interesting question is whether people are at all concerned with mental contamination. As noted in a recent article about violence in movies and television, there is at least one segment of

[3] It is possible, of course, to try to estimate bias in everyday life by using the correlational method. For example, Professor X might realize that the average grade for the women in his class was higher than the average grade for men (e.g., that his grades were correlated with gender) and conclude that he was biased in favor of the women. There is ample evidence, however, that everyday covariations are difficult to detect and are often driven as much by theories as by actual data (e.g., Chapman & Chapman, 1967, 1969; Nisbett & Ross, 1980). Furthermore, even if one accurately perceives such a covariation, it is difficult to determine its cause. It is conceivable that the women in Professor X's class were far superior in ability to the men and would have achieved even higher grades had Professor X not been biased against them.

the population that is not: "It can be argued that the entertainment industry as a whole has probably given more thought to the pollution of rivers than it has to the pollution of minds" (Auletta, 1993, p. 46). We show here, however, that people are concerned with at least some forms of mental contamination. A second basic question is whether people should be concerned. Admittedly, mental contamination does not pose the same immediate threat to our well-being as do many physical contaminants. Experimental demonstrations of biased judgment, however, deliberately minimize the impact mental contamination has on participants because of ethical concerns. Extrapolating from many of these studies, it would seem that mental contamination can have very serious effects. Consider, for example, the many studies on priming effects that show that people's judgments of others are influenced by categories that happen to be accessible. It is not difficult to imagine how such an instance might occur in everyday life with very unfortunate consequences. For example, someone meets a stranger in a bar and, because the category of trust and kindness has recently been primed, decides to go home with the stranger. The stranger turns out to be anything but trustworthy. Or consider the way in which people make important personal decisions, such as whether to seek medical treatment, change jobs, or make a major purchase. It is quite likely that people are not immune to the biases we have discussed when making such decisions (see Gilovich, 1991, for many other examples of real-world, negative effects of biases in human reasoning).

Are people at all concerned with possible biases on their judgments and inferences? To find out, we (Wilson, Brekke, Etling, & Houston, 1992) polled a sample of college students, asking them to rate how concerned they were about a wide range of occurrences ranging from the mundane (e.g., being overcharged at the supermarket) to the serious (e.g., being the victim of a violent crime) to the ridiculous (e.g., being kidnapped by aliens). We intermixed several examples of mental contamination in the list, including both general instances (e.g., "the possibility that I might make errors of judgment when making important life decisions") and specific instances (e.g., "that I might misjudge someone's personality by failing to take his or her situation into account," that is, the well-known fundamental attribution error; Jones, 1990; Ross, 1977). The students rated each item on a scale ranging from *I never worry about this; I am not at all concerned about it* (0) to *I always worry about this; I am very concerned about it* (5).

The results indicated that students are concerned about the validity of their judgments. For example, the possibility that they might make errors of judgment when making life decisions ranked 5th out of 70 items, trailing close behind such obvious student concerns as getting good grades, getting a good job, and finding a spouse. There was a fair amount of concern with specific forms of contamination as well, particularly with cases in which other people's contamination would affect them. For example, students' concern that a teacher's grade might be biased by how much the teacher liked them ranked 16th, tied with "whether I'm smart." A concern that other people might not be able to get over a bad first impression of the students ranked 23rd, tied with being in a car accident. And researchers on the fundamental attribution error will be interested to hear that students are more concerned with this bias than they are with nuclear war.

Clearly, however, there are some forms of mental contamination that do not concern people, even though there is ample evidence in the literature for the existence of such effects. For example, the possibility that someone could bias the students' memory with leading questions ranked 64th, not far above being the victim of a terrorist attack or being kidnapped by aliens, despite ample evidence for the biasing effect of leading questions (Loftus, 1979). Also, the students were no more concerned that a distracting noise would influence their liking for an activity than they were about being overcharged in the supermarket, despite evidence for the biasing effect of distracting noises (Damrad-Frye & Laird, 1989). This raises the question of the extent to which people's beliefs are calibrated. How well do lay theories about mental contamination correspond to actual contaminants?

Are people's theories about mental contamination poorly calibrated? If people have poor access to cognitive processes that are biasing their responses, they can still gain indirect awareness of these processes by applying correct causal theories. Even if professors have no direct access to the process by which their liking for students biases their grading, they might hold the theory that such halo effects occur and thus take steps to avoid such biases. How accurate are people's theories about mental contamination? This is an extremely difficult question to answer, because there is no way to itemize the full range of actual contaminants for any given individual and no way to compare such a list with a complete inventory of the person's causal theories. Nonetheless, there are grounds for suspicion that people's theories are not very well calibrated. If people were familiar with all of the many shortcomings in human inference that have been documented, then it would not have taken social and cognitive psychologists so long to discover these biases.

As another indication of the inaccuracy of people's theories about contamination, it is instructive to examine the assumptions about mental bias that are part of legal proceedings. The rules of evidence and procedure could be considered codified versions of lay theories about judgment and decision making in that they specify conditions under which jurors are likely to form biased verdicts and procedures that are believed to prevent such biases. A considerable amount of research, however, suggests that many of these assumptions about bias are incorrect (for reviews, see Gerbasi, Zuckerman, & Reis, 1977; Kassin & Wrightsman, 1988; Wrightsman, 1991). For example, legal procedure is based on the assumption that jurors can easily discount testimony that they are told is inadmissable; there is considerable evidence that they cannot (e.g., Sue, Smith, & Caldwell, 1973; Thompson, Fong, & Rosenhan, 1981). Similarly, the legal system allows the use of surrogate witnesses (actors who read or act out the testimony of a witness who has already been questioned but is unable to appear in court). The assumption that jurors can keep their evaluations of the testimony separate from their impressions of the surrogate witness (e.g., his or her appearance, demeanor, and delivery) is incorrect (Kassin, 1983). Kassin and Wrightsman (1988) went so far as to argue that juries often function suboptimally because the legal system structures their task poorly as a result of misunderstandings of human cognition.

We obtained similar evidence for the inaccuracy of theories about mental bias in our previously mentioned survey in that some well-known cognitive biases, such as the effects of leading questions, were ranked very low, suggesting that people's theories do not cover some well-established biases. One problem with this finding, however, is that people could be unconcerned with a potential source of mental contamination for several reasons. First, they might not believe that they are ever influenced by such things as leading questions and distracting noises; that is, they do not know that such effects can occur. Second, people might recognize the power of the influences listed in our survey but be perfectly willing to be affected by them. That is, they may not want to avoid these influences because they believe that they are perfectly valid. Third, people might know that such phenomena occur and believe that they are contaminants but believe that the biases occur only rarely. For example, if people believe that leading questions can bias memory reports but that they are rarely in a situation where someone asks them leading questions, they are justified in being relatively unconcerned with this form of mental contamination. Fourth, people might believe that the contaminants are potentially influential but believe that they have effective strategies to counteract them. They might know that leading questions can bias memory but believe that they can take steps to resist this bias when they encounter it. Finally, people might believe that a contaminant exists, that they encounter it frequently, and that they cannot counteract it but believe that the consequences of the contamination are relatively unimportant. They might believe that distracting noises often influence their liking for an activity, for example, but believe that this effect is of a trivial magnitude.

We have recently conducted studies to demonstrate more clearly that people are concerned about consequential sources of bias on their judgments but that at least some of these concerns are misplaced (i.e., that people's theories about what contaminates them are poorly calibrated), with nontrivial consequences. Some of these studies involved surveys in which people were questioned about the effects that various contaminants, such as subliminal messages, false innuendo about a political candidate, and the knowledge of the gender of a job candidate, would have on them (we did not, of course, use the term *contaminant* in the actual surveys). People answered two questions about each piece of information: how much they would want to be influenced by it and how much they really would be influenced. By comparing people's answers to these two questions, we were able to assess the extent to which people believed each kind of information would have an unwanted influence on their decisions (Wilson et al., 1992).

We found that, for each of the kinds of information we described to subjects, they believed that the real influence of this information would be greater than the desired influence, supporting our hypothesis that people believe that their decisions can be biased in various ways. Averaging across the different kinds of information, the difference between the ratings of the real and desired influence was highly significant, although people believed that some of the information would bias them more than others. There is no way of estimating from this study, of course, how calibrated these beliefs are with the actual influences on people's judgments. On the basis of prior research, however, there are at least grounds for suspicion that people's beliefs are not very well calibrated. For example, one of the largest differences between ratings of desired and actual influence was for the effects of subliminal messages, yet there is little evidence that such messages can influence people's real-world decisions (Merikle, 1988; Moore, 1982; Pratkanis & Greenwald, 1988; see Zanot, Pincus, & Lamp, 1983, for similar results in a survey of randomly chosen adults in a metropolitan area). These results suggest that people's theories about mental contamination are incorrect in at least some instances.

This survey study, however, did not document that the inaccuracy of people's theories is consequential. To demonstrate that they can be, we recently conducted a laboratory study in which female college students were asked to make a decision of personal importance, namely, to evaluate two different brands of condoms and choose one to take home for their personal use (Wilson, Etling, & Houston, 1993). We made two kinds of information available to participants: (a) base-rate information about the brands that was reported in *Consumer Reports* magazine and (b) the reports of two students at their university who described their experiences with the two brands. Participants could choose to see neither, one, or both of these kinds of information, after which they rated their preference for the different brands and chose one to take home. All choices and ratings were private and anonymous to avoid self-presentational concerns.

Before discussing the results, it is important to note two things. First, the *Consumer Reports* data clearly indicated that one brand of condom was superior (Brand A), whereas both students reported that the other brand was superior (Brand B). For example, one of the students reported a vivid instance in which Brand A burst while she and her lover were having sex, resulting in a considerable amount of anxiety over whether she was pregnant. Second, it is necessary to demonstrate that our participants agreed with the literature on the use of base-rate information, in the sense that they would rather be influenced by the *Consumer Reports* data. That is, given that our definition of mental contamination was a subjective one, it was incumbent on us to show that in the eyes of our participants, the *Consumer Reports* data represented "desired" influence and the vivid student reports represented "undesired" influence.

To do so, we asked a separate group of participants to rate how much they would want to be influenced by the different kinds of information about the condoms. They did so by indicating which brand of condom they would prefer, knowing that *Consumer Reports* recommended Brand A and the students recommended Brand B.[4] Another group did the same thing, except that they rated how much they would want a close friend to be influenced by the different kinds of information. Very few of these latter subjects reported that they would want their friend to be influenced more by the student information than the *Consumer Reports* information: Only 3% of them stated that they

[4] We varied whether these subjects read brief descriptions of the *Consumer Reports* and student information (as described to the actual participants before they made their choice of which information to see) or the complete information from each source (as seen by actual subjects who choose to see the information). In both cases, control subjects were told which brand of condom was recommended by each source. As it happened, this manipulation did not affect people's ratings of which type of information they would prefer, thus, we collapsed across it.

would want their friend to choose the brand recommended by the students. Somewhat more indicated that they would prefer the influence of the student information for themselves (13%), although the percentage was still low (we return to a discussion of the self vs. other results shortly). Thus, it appears that, in this population, most people preferred to be influenced by the *Consumer Reports* data, which suggested that Brand A was superior.

How does this compare with what people actually did when offered both kinds of information? First, most people chose to see both kinds of information (77%), and very few chose to avoid the student information altogether (12%). This suggests that although people recognized the invalidity of the student information, they had faith in their ability to avoid being influenced by it in an undesirable way. Once people had both kinds of information, however, were they able to avoid the unwanted effects of the student reports? Not entirely: 31% preferred the condom recommended by the students after choosing both kinds of information. This percentage was significantly higher than the percentage of people in our control sample who stated that they would prefer that a close friend or that they themselves be influenced by the student information. That is, using people's peers as the standard of what is "correct" in this situation, a significant proportion of people did what was "incorrect."[5]

In summary, people's beliefs about mental contamination are not perfectly calibrated with those cognitive processes that can actually lead to contamination. Only a limited amount of research has examined the consequences of such faulty beliefs. Preliminary research from our laboratory, however, suggests that such beliefs can have unfortunate consequences. In our condom study (Wilson, Etling, & Houston, 1993), people's beliefs about the power (or lack thereof) of the student information and their faith that they could resist such effects led them to choose to see this information. A significant proportion of the participants, however, were unable to avoid the influence of the student information. Clearly, this is an area that would benefit from further empirical investigation.

Do people underestimate their own susceptibility? We suspect that a serious problem with lay beliefs about contamination is a lack of humility about one's susceptibility to it. We saw a hint of this in the condom study: People were more likely to report that they preferred to see the student information than to indicate that a close friend should see it. People seemed to believe that the information was more dangerous in the hands of someone else than in their own hands, possibly because they believed that they could control its effects more than could a close friend. But, as shown, at least some participants were wrong about their ability to control the impact of the student information. We also tested this self–other difference in our survey work mentioned earlier (Wilson et al., 1992). We gave some participants the same scenarios about the influence of possible contaminants and asked them how other people would be influenced by the information (e.g., innuendoes about a political candidate) rather than how they would be influenced. People believed that others would be significantly more biased by most of the information than they would be. Similarly, we found in a study of anchoring and adjustment that people believed that they were uninfluenced by the anchor value but that other participants in the study were influenced. Given that substantial anchoring effects were found, these results suggest that people were underestimating their own susceptibility to anchoring effects rather than overestimating other people's susceptibility (Wilson, Houston, Etling, & Brekke, 1993).

People's faith in their own lack of susceptibility to mental contamination may be an example of unrealistic optimism that has been documented elsewhere (Weinstein, 1980, 1982). Weinstein found that people believe that negative life events (e.g., getting divorced) are less likely to happen to them than to other people and that positive events (e.g., getting a high-paying job) are more likely to happen to them than to other people. Given that it is undesirable to be more influenced by a piece of information than one wants to be, the differences between self-ratings and other ratings in our studies may be a similar case of unrealistic optimism. If so, people may underestimate their own vulnerability to mental contamination and thus fail to take adequate steps to avoid it.

Do people overestimate their own mental control? Another possible reason people think they are less susceptible to bias than others is because of a faith in their ability to control their own thoughts and feelings. We have shown here that the actual amount of control people have is limited. It is important to consider people's beliefs about control because these beliefs will influence the strategies they use (or fail to use) to avoid mental contamination. To address this question, we (Wilson et al., 1992) conducted another version of our survey on mental contamination. The only difference was that the students answered different questions about the possible contaminants, choosing one of five options about how they would deal with each kind of information (e.g., innuendos). They could indicate that they (a) wanted the information to make up their minds, (b) wanted it but that it would not influence their decision, (c) did not want it because it might bias their decision, (d) did not want it because it would not influence them and would be bothersome or a waste of time, or (e) did not care whether they received the information or not.

The number of people who wanted the information (presumably out of curiosity) even though it would not influence their decision or did not care whether they received the information is of particular interest. These choices imply a belief in the ability to control one's mental processes, particularly if people believe that the information is potentially biasing. For example, consider the effects of gender on a hiring decision. Most people reported that they would not want gender to influence their decision but believed that it would influence them more than they would want it to. Nonetheless, 87% indicated that they would like to know the gender of a job candidate, presumably because they believe they are able to avoid any contaminating effects of such information. Only 5% stated that they would not want to know to avoid being biased. This faith in the ability to control

[5] When we gave people the choice of which information to see, there was, of course, a subject self-selection problem: Not everyone chose to see both kinds of information. To address this problem, we included other conditions in which people were randomly assigned to receive one or both kinds of information and found results very similar to those reported here. Among those people randomly assigned to receive both the *Consumer Reports* and student information, 28% preferred the condom recommended by the students.

one's judgments is probably misplaced. In our condom study, for example, once people had read the vivid student information about the condoms, more people were influenced by it than wanted to be.

Summary. There is little research on people's beliefs about mental contamination. The available evidence suggests that people are concerned about having biased judgments but that their theories are poorly calibrated with known contaminants, they underestimate their own susceptibility to bias, and they overestimate the extent to which they can control their judgments and feelings. These findings are cause for concern about the strategies people take to avoid mental contamination. They may not be very motivated to take preventive measures, and, when they do, they might try to avoid the wrong contaminants (e.g., subliminal influences instead of knowledge about the gender of a job candidate).

Types of Mental Contamination

To this point, we have reviewed the general causes of contamination, including both fundamental properties of human cognition and lay beliefs about the mind. Our discussion has been at a rather broad level, encompassing a wide variety of mental errors (e.g., halo effects, misattribution, and negative consequences of accessibility). One of our goals is to consider the best ways of avoiding mental contamination. To accomplish this, it is necessary to be more specific about the processes involved in different forms of contamination. We believe that it is fruitful to divide mental contamination into two subcategories: biases that are by-products of automatic processing and what we call source confusion, which is the case in which people confuse two or more sources of a memory, thought, feeling, or judgment. Although this categorization is not necessarily exhaustive, we believe that identifying these different kinds of contamination has important implications for how to avoid mental biases.

Unwanted Consequences of Automatic Processing

Over the millennia, humans have developed many highly useful and adaptive information-processing strategies. Many of these strategies involve unconscious, automatic processing that is highly functional. Indeed, it is difficult to imagine what it would be like to have to process all information about the environment through the narrow filter of consciousness, with no preattentive processing or automatic categorization of the environment. As with most procedures, however, these highly adaptive processes can be overused or can, under some circumstances, go awry, producing faulty output. Nonetheless, they are efficient and functional enough that their benefits outweigh their costs (Einhorn & Hogarth, 1981; Miller & Cantor, 1982). In Nisbett and Ross's (1980) words, "we suspect that many of people's failings will prove to be closely related to, or even an unavoidable cost of, their greatest strengths" (p. 14).

The kinds of strategies to which we refer are very different from the consciously learned rules we discussed in the section on rule knowledge and application (e.g., the law of large numbers). We refer now to strategies that may well be innate or at least are so overlearned as to become implicit, nonconscious, and, at least to some extent, uncontrollable. Many of these processes fit current definitions of automaticity. We recognize that it may be a mistake to consider automaticity as a single, unitary concept; as noted by Bargh (1989), automatic processing comes in different forms and flavors, with varying degrees of awareness, intentionality, and controllability. Nonetheless, we believe that it is useful to classify several types of human information processing as relatively automatic to distinguish them from the type of deliberate, consciously learned rules we discussed earlier. For example, the human predilection to categorize the physical and social world is probably unlearned and is certainly more automatic than is using the principles of sunk costs or regression to the mean.[6]

Furthermore, automatic processes are often accompanied by more conscious, controllable mental processes; in fact, several models of automatic processing, as shown later, specify the operation of controlled processing that operates in tandem with automatic processing. Our point is that there is a set of specifiable errors that result from this interplay between automatic and controlled processing. This class of errors can be considered a by-product of otherwise useful mental procedures. It can also be considered one type of mental contamination in that people end up with judgments, beliefs, or responses that they would agree are faulty as a result of the unwanted intrusion of automatic processing. That is, people's judgments, emotions, or behaviors go awry because they are "contaminated" by automatic mental processes and, in some cases, by the controlled processing triggered by automatic processes.

Again, we are not arguing that automatic processes always result in contamination. These processes are highly adaptive and are often quite functional. There are times, however, when they result in unwanted mental states. In this sense, this class of errors fits the scheme depicted in Figure 1. The unwanted processing in the first box includes those instances in which automatic processes have undesirable effects. Contamination results if people are unaware of these processes, if they are unmotivated to correct for them, if they are unaware of the direction in which they have biased their responses or the magnitude of this bias, or if they are unable to control their responses enough to correct for the bias. We have listed several examples of unwanted consequences of automatic processing in Appendix B. Some are extremely well documented, whereas others are relatively new and undeveloped. We provide some examples here, beginning with the human tendency to categorize the physical and social environment.

Categorization. People immediately place the things they encounter into preexisting knowledge structures or schemata. Indeed, it is difficult to "turn off" this mental proclivity, preventing ourselves from assuming, for example, that the object in front of us with a platform, straight back, and four legs is a chair or that the person wearing an expensive suit and talking into a mobile phone is an executive and not a janitor. Although

[6] An interesting question is whether such consciously learned rules as the principle of sunk costs can become automatic after prolonged use. It may be that they can, just as other well-learned rules become automatized with practice. We suggest that the process by which people learn and apply such inferential rules, however, is very different from the kinds of (possibly innate) automatic processes we are considering here, which are probably unlearned and more difficult to control.

there is some controversy over the exact nature of the categorization process, there is widespread agreement that humans are prone to quick categorization of their environment (Medin, 1989; Rosch, 1978; Smith & Medin, 1981).

The advantages of categorization are well documented, yet the drawbacks are equally clear (e.g., Brewer, 1989; Fiske & Neuberg, 1990). One particularly interesting case of the unwanted consequences of categorization is prejudice and stereotyping. According to many researchers, stereotypes of social groups are learned at an early age and are invoked automatically when people encounter members of that group (Allport, 1954; Billig, 1985; Brewer, 1989; Devine, 1989a, 1989b). Several researchers suggest that there are circumstances under which people find it difficult not to invoke these stereotypes and allow them to bias their judgments, even if they have the best of intentions (Devine, 1989b; Gilbert & Hixon, 1991; Wegner, 1994).

Clearly, for a nonprejudiced person, this effect would qualify as mental contamination: Knowledge of a negative stereotype is invoked automatically, tainting judgments in an unwanted way. Devine (1989a, 1989b) argued that nonprejudiced individuals are often able to countermand this automatic activation of stereotypes by consciously adjusting their responses in a nonprejudiced direction or at least attempting to do so. Wegner (1994), however, suggested that this is easier said than done. In fact, as is shown shortly, the very act of trying to suppress stereotypic responses can increase their frequency. In general, even when people try to adjust their responses when interacting with members of stereotyped groups, this adjustment process can be very difficult to get "just right" and is especially difficult when one's cognitive capacity is taxed.

Despite these problems with people's proclivity to categorize their environment automatically, we assume that the benefits of categorization outweigh the costs. The pernicious effects of stereotyping are certainly not to be underestimated, but we know of no one who argues that people should strive, at all costs, to rid their minds of the entire categorization process.

Knowledge accessibility. There has been a great deal of recent research on the determinants and consequences of accessibility, which can be defined as the "readiness with which a stored construct is retrieved from memory and/or is utilized in stimulus encoding" (Higgins & King, 1981, p. 71). A number of experiments have shown that when a construct in memory is accessible, either through recent or frequent contact with information relevant to that construct, people are more likely to use that construct in social judgment (Bargh, 1990; Higgins, 1989; Uleman, 1989; Wyer & Srull, 1989). One way of viewing accessibility is as a determinant of categorization; that is, people are likely to assimilate new experiences to those categories that are most accessible in memory. Accessibility effects occur automatically, without the intention of or control by the perceiver (Bargh, 1989). Several theorists have noted the benefits humans obtain by such a nonconscious, automatic activation system, such as the ability to resolve ambiguities by quickly applying one's working knowledge of the world (Higgins & Bargh, 1992).

Researchers in this area have been quick to note, however, the downside of such an automatic activation system. The initial investigations of priming effects demonstrated that social constructs can be activated by information irrelevant to the judgment at hand, thereby influencing people's judgments in arbitrary (and presumably undesirable) ways. Higgins et al. (1977), for example, showed that people assimilated their judgments of an ambiguous target person to a trait category that had been primed in an earlier, unrelated experiment. More recently, Higgins (1987) demonstrated that the chronic and temporary accessibility of certain discrepancies among people's "actual," "ideal," and "ought" selves can produce emotional distress (see Higgins & Bargh, 1992, for a discussion of other unwanted consequences of automatic knowledge activation).

Initial acceptance of propositions. As noted by Gilbert (1991, 1993), there is a long tradition in philosophy and psychology, originating with Descartes, that assumes that belief formation is a two-step process: First people comprehend a proposition (e.g., "Jason is dishonest") and then freely decide whether to accept it as true (e.g., whether it fits with other information they know about Jason). Thus, there is no danger to encountering potentially false information because people can always weed out the truth from the fiction, discarding those propositions that do not hold up under scrutiny. Gilbert (1991, 1993) argued persuasively, however, that human belief formation operates much more like a system advocated by Spinoza. According to this view, people initially accept as true every proposition they comprehend and then decide whether to "unbelieve" it or not. Thus, in the example just provided, people assume that Jason is dishonest as soon as they hear this proposition, reversing this opinion if it is inconsistent with the facts.

Under many circumstances, the Cartesian and Spinozan systems end up at the same state of belief (e.g., that Jason is honest because, on reflection, people know that there is no evidence that he is dishonest). Because the second, verification stage requires mental effort, however, there are conditions under which the two systems result in very different states of belief. If people are tired or otherwise occupied, they may never move beyond the first stage of the process. In the Cartesian system, the person would remain in a state of suspended belief (e.g., "Is Jason dishonest? I will reserve judgment until I have time to analyze the facts"). In the Spinozan system, the person remains in the initial stage of acceptance, believing the initial proposition. Gilbert has provided evidence, in several intriguing experiments, for the Spinozan view: When people's cognitive capacity is taxed, they have difficulty rejecting false propositions (see Gilbert, 1991, 1993).

Gilbert (1991) argued that the Spinozan procedure of initial acceptance is an adaptive one. It has its roots in the perceptual system, he suggested, wherein it is to people's great advantage to believe that what they perceive reflects reality (e.g., a car speeding in their direction is not a hallucination). In the realm of belief, there is a greater likelihood that new propositions are false (e.g., hearing that "New Blippo Detergent gets out even the toughest stains!"). But, even in the realm of belief, it can be highly efficient to initially believe what one hears. In Gilbert's (1991) words, "just as perceptual systems enable timely action by capitalizing on the fact that most percepts are faithful, cognitive systems may achieve similar efficiency by capitalizing on the fact that most propositions are true" (p. 116). The initial acceptance of propositions, then, fits nicely into our first class of mental contamination: This state of affairs is highly adaptive much of the time but can lead one astray under certain condi-

tions (e.g., when people are tired, preoccupied, or cognitively taxed, such that they do not have the resources to complete the second, "unacceptance" stage of the Spinozan process).

It should be noted that a number of biases in reasoning can be considered examples of the failure of this unacceptance process. Chief among them, perhaps, is the tendency to attribute people's behavior to their underlying dispositions and to underestimate the effects of situational variables on their behavior. This proclivity, called the correspondence bias by Jones (1990) and the fundamental attribution error by Ross (1977; Ross & Nisbett, 1991), has been demonstrated in dozens of experiments. Gilbert (1991) suggested that, in many instances, people initially accept as true the traits or attitudes implied by a person's behavior, even when there are clear situational constraints on the person's behavior. Adjusting one's attributions for the situation may require an effortful unacceptance that will be incomplete when people are cognitively busy. Interestingly, when people's goal is to assess the nature of the situation, this process may be reversed: People automatically categorize the situation and then adjust this inference to take into account an actor's dispositions (Krull, 1993).

Thought suppression. One final area of automatic processing is particularly intriguing because when it goes wrong, it goes very wrong. Wegner (1992, 1994) discussed the case of thought suppression, whereby people are motivated not to think about a particular topic (e.g., a lost love, an unpleasant encounter with one's boss, or a delectable piece of cheesecake calling from the refrigerator). According to Wegner's model, successful thought suppression depends on the interaction of two processes, one that is relatively unconscious and automatic and one that is relatively conscious and controlled. The first process, called the ironic monitoring process, is relatively automatic and searches for evidence that the unwanted thought is about to intrude on consciousness. Once the unwanted thought is detected, the second, more controlled process, called the intentional operating process, comes into play. This is the effortful, conscious attempt to distract oneself by finding something else to think about. These two processes operate in tandem like a feedback loop, whereby the ironic monitoring process finds an unwanted thought, initiating the operation of the intentional operating process, which attempts to think about something else. The process begins again the next time the ironic monitor detects further evidence of the unwanted thought.

Wegner (1992, 1994) noted that, under normal circumstances, this system can operate fairly successfully. In the short run, at least, people are able to reduce the number of times that an unwanted thought intrudes on consciousness. Although suppression can break down in the long run, resulting in rebound effects (the increased incidence of unwanted thoughts after suppression), it seems reasonable to view the operation of the automatic and controlled process in thought suppression as an adaptive process. It is sometimes very important to be able to put something out of our minds so that we can focus on the task at hand. If thoughts occurred to people in a willy-nilly fashion, with no ability to control, filter, and suppress them, their mental lives would be chaotic indeed.

As with our other examples of the unwanted effects of automatic processing, however, there are circumstances under which the thought suppression system breaks down, with unfortunate consequences. As with the Spinozan belief process, its success depends on a delicate balance between an automatic and controlled mental process. If people's ability to execute the controlled part of the process (the distracter search) breaks down, an ironic effect occurs: The ironic monitor continues to find instances of the unwanted thought, which then intrudes on consciousness unchecked by the conscious, controlled part of the process. Consequently, a state of hyperaccessibility occurs, whereby the unwanted thought occurs at a greater frequency than it would have if the thought suppression process had not been instigated (Wegner, 1992, 1994). Such a state of hyperaccessibility can lead to responses that most people would consider to be unwanted, fitting our definition of mental contamination. For example, Wegner, Erber, and Bowman (1993) demonstrated that asking people to try not to make sexist responses while under cognitive load (thus preventing the operation of the controlled distracter search) results in an increase in the number of sexist responses they emit.

To summarize, in this section we have argued that mental contamination can result when normally adaptive automatic processing strategies go awry or operate unchecked by conscious, controlled processes. We view these errors as the inevitable consequences of otherwise beneficial procedures, analogous to the unfortunate fact that some beneficial aspects of human physiology (e.g., genes granting immunity to malaria) have unfortunate by-products (e.g., a susceptibility to sickle-cell anemia). We should note an alternative to this way of considering the unwanted effects of automatic processing. It might be argued that the rules are themselves flawed and could be improved to eliminate the errors they sometimes cause. This view argues that we have focused on human judgment in only one thin slice of evolutionary time. Were we to view it a million years from now, we may well have developed superior procedures that avoid many of the pitfalls we now observe (see Gilbert, 1991, and Stich, 1990, for related discussions). This is an interesting issue about which to speculate, but one that is impossible to answer at present.

Our point is simply that there exists a class of errors that is a by-product of otherwise useful, automatic procedures. These errors differ in two important ways from our first category of mental errors (the failure to know or apply a conscious algorithm). First, they result from information-processing strategies that operate largely outside of awareness, are difficult to control, and, thus, are not easily "fixable" (Bargh, 1989). Second, even if one could control and tinker with these strategies, it is not clear that one would want to do so given that they are adaptive and beneficial much of the time. We discuss these issues in greater detail in the section on correcting mental errors. We first need to describe another way in which people's judgments can be contaminated.

Source Confusion

Mental contamination can also result from source confusion, which occurs when people confuse two or more sources of a memory, thought, feeling, or judgment. Our use of this term is very similar to Johnson, Hashtroudi, and Lindsay's (1993) term *source monitoring.* The concern in both cases is with people's ability to make accurate attributions about the causes of their

responses. Our view of this process is somewhat broader in that it encompasses a wide variety of cognitive and social phenomena. In everyday life, people's responses are almost always multidetermined; their feelings for a loved one spring from many sources, and their evaluation of a job candidate is based on more than one of the candidate's many attributes. A considerable amount of research suggests, however, that people are not skilled at dissecting a judgment (recognizing the exact contribution of everything that caused it). It is as if their minds were an inscrutable cauldron of mental activity. They recognize the thoughts and feelings that result but not the exact recipes that produced them (Nisbett & Wilson, 1977b; Wilson, Dunn, Kraft, & Lisle, 1989). For example, people cannot say with any confidence that "32% of the variance of my love for my spouse is based on her kindness, 11% on her sense of humor, 23% on her physical appearance, and the rest is pheromones" (Wilson & Hodges, 1992).

Because of this source confusion—the inability to recognize the exact contribution of all of the influences on one's judgments—it is difficult to take steps to avoid being affected by unwanted agents. As a result, people often end up with contaminated judgments. We have listed several examples of source confusion in Appendix B. The first is classic misattribution effects, wherein people confuse two or more causes of their arousal. People have been found to mistakenly believe that their arousal is caused by an insulting questionnaire instead of a drug (Schachter & Singer, 1962), by a placebo pill instead of painful electric shocks (Nisbett & Schachter, 1966), by an attractive experimenter instead of fear from crossing a scary bridge (Dutton & Aron, 1974), and by a sexually arousing film instead of physical exercise (Zillmann, 1978). Although the generalizability of misattribution effects has been questioned (Marshall & Zimbardo, 1979; Maslach, 1979; Reisenzein, 1983), the basic phenomenon of confusing one source of arousal with another is well established (Olson, 1990).

Recent research on the unconscious influence of memory can also be considered to be a case of source confusion. For example, Jacoby, Woloshyn, and Kelley (1989) asked participants to read a list of nonfamous names (e.g., "Sebastian Weisdorf"). Later in the experiment, participants viewed a list of names that included some of the nonfamous ones they had seen earlier and were asked to judge whether these names were those of famous people. If people remembered that a name on the second list had appeared on the first list, then they could correctly label it as nonfamous because they knew that the first list included only nonfamous names. If they did not remember seeing a name before, however, then the unconscious effects of memory might cause them to mistakenly label it as famous. That is, the name might seem familiar and thus be mistaken for a famous person, precisely because people do not consciously remember ever seeing it. Jacoby et al. (1989) found that many of the nonfamous names were mistakenly identified as famous. In our terms, source confusion occurred, in that people mistakenly believed that the familiarity of a name was caused by its famousness rather than from prior exposure to the name in the experiment.

Several other examples of source confusion are listed in Appendix B. Among them is the halo effect, an example of which we have used throughout this article (the case in which a college professor's grade is biased by his or her liking for a student). Halo effects are examples of source confusion in the sense that people become confused about the exact source of their judgment (e.g., the grade), namely, the extent to which it is based on an objective evaluation of a person's performance versus their liking for that person.

These examples fit our definition of mental contamination in that mental processes result in judgments, emotions, and other responses that people, at least at times, find undesirable. In terms of the scheme displayed in Figure 1, the unwanted processing in the first box is the misattribution that occurs when people confuse two or more sources of their responses. The conditions under which contamination occurs are then the same as outlined earlier: Once source confusion occurs, it can be corrected only if people are motivated to do so, are aware of the direction and magnitude of the bias, and have sufficient control over their responses to avoid the bias.

We should note that this category of mental contamination shares some features with the previous one (unwanted by-products of automatic processing). They occur, in large part, because of the same limitations of the human mind (e.g., the limits on awareness of mental processing and mental control) reviewed earlier. Indeed, some examples of mental contamination are difficult to classify, such as the effects of knowledge accessibility. Earlier we discussed this as an example of the unwanted effects of automatic processing, in that a construct is primed automatically and that construct taints later judgments. Such a process might also be considered a case of source confusion in that people are confused about the extent to which their judgment of a stimulus (e.g., another person) is influenced by Source 1 (objective attributes of the stimulus) versus Source 2 (a previously primed social category). In fact, all source confusion errors could be viewed as the unwanted consequences of automatic processing if one assumes that people's inability to tease apart different sources is the by-product of their general tendency to form judgments and evaluations outside of consciousness.

Despite this overlap, we believe that it is fruitful to consider our two categories of contamination separately for three reasons. First, in our examples of source confusion, it is more difficult to identify the functional mental process that causes contamination, unlike the errors we reviewed earlier (e.g., the unwanted consequences of automatic categorization). For example, the benefits of the processes leading to misattribution or halo effects are not clear, and if one were to design the perfect human, it seems doubtful that one would include the processes responsible for such biases. In contrast, one might well include automatic accessibility effects, given the benefits of having a working model of the world at one's fingertips (Higgins & Bargh, 1992). Second, unlike errors resulting from automatic processing, source confusion involves misattribution, wherein people make mistaken inferences about the causes of their responses. Third, the strategies to avoid or undo these biases are somewhat different, as shown later.

We would also like to emphasize that, although there are "pure" cases of each type of error in human reasoning, errors in everyday life can be overdetermined. Consider the example of Jane, who invests all of her retirement savings in an aggressive mutual fund that suffers severe losses over the next year. It is possible that her unfortunate decision was determined in part by all of the types of errors in reasoning we have documented.

First, her failure to appreciate the phenomenon of regression to the mean may have contributed to her decision (rule ignorance). That is, the fact that the mutual fund posted substantial gains the previous year may have made Jane overly confident that it would post similar gains the next year. Second, there may have been some unwanted consequences of automatic processing. Perhaps Jane was tired when she read the prospectus, making it difficult for her to "unaccept," in Gilbert's (1991) terms, some of the lofty claims made by the fund managers. Or the profitable performance of mutual funds may have been so accessible in Jane's memory that it was difficult for her to consider alternative scenarios. Finally, source confusion may have come into play, whereby a specific agent or mental state influenced Jane's decision in an unwanted way. Perhaps she was unduly influenced by a magazine advertisement for the fund or by a friend's comment about how wonderful it was (more so than she wanted to be).

Even though faulty judgments can be multidetermined, we believe that our classification scheme is useful for determining how to improve judgment. As shown earlier, Nisbett and his colleagues have had considerable success in teaching people better rules of inference; thus, teaching Jane about regression to the mean might make her wary about basing her judgment on the fund's performance during the previous year. But what about the different kinds of mental contamination that contributed to her decision, such as the tendency to accept propositions as true and the unwanted effects of the advertisement on her decision? We have touched on the issue of how to correct or avoid mental contamination throughout this article. We turn now to a more detailed discussion of this issue, which is, after all, the bottom line: How can people avoid contamination?

Avoiding and Correcting Mental Contamination

We outlined in Figure 1 the process by which mental contamination can be avoided or eliminated: having awareness of the contamination, the motivation to correct it, awareness of the direction and magnitude of the bias, and sufficient control over one's responses to accomplish the correction. We also reviewed several literatures suggesting that these conditions are difficult to meet. These literatures were concerned with people's general capabilities relevant to correction (e.g., their awareness of their mental processes). Many studies have examined this issue more directly by trying to increase awareness of bias, asking people to correct for their exposure to contaminants, or warning people that they are about to be contaminated. Even with such explicit attempts to warn or debias people, it is often very difficult to satisfy all of the conditions necessary to avoid contaminated responses.

Effects of Forewarning and Debiasing Manipulations

A number of studies have attempted to reduce biases in information processing and judgment by forewarning people about or drawing their attention to potentially biasing information and examining the extent to which they are able to avoid the unwanted effects of this information. These studies include those in the law and psychology area that examine jury decision making and those that forewarn people about biases on other kinds of judgments and decisions. In general, these studies have revealed a wide range of seemingly contradictory effects. Some studies have shown that an increase of people's awareness eliminates mental contamination; some have found that awareness causes people to adjust insufficiently, leading to undercorrection; some have indicated that awareness causes people to adjust their responses too much, leading to overcorrection; and some have shown that awareness does not cause people to adjust their responses.

Consider, for example, studies that have forewarned people about specific biases (e.g., the halo effect) and examined the extent to which people are then able to avoid the biases. On the face of it, this literature has produced inconsistent effects. Some studies have found that forewarning has little or no effect on people's responses (e.g., Fischhoff, 1977; Kurtz & Garfield, 1978; Wegner, Coulton, & Wenzlaff, 1985; Wetzel, Wilson, & Kort, 1981; Wilson, Houston, Etling, & Brekke, 1993). Other studies have shown that forewarning can be effective in eliminating biased judgments (e.g., Golding, Fowler, Long, & Latta, 1990; Schul, 1993; Weinberg & Baron, 1982). Most notably, a consistent finding in the persuasion literature is that people who are forewarned about the content of an attitudinal message or simply of the intent to persuade them show less attitude change than people who are not forewarned (e.g., Petty & Cacioppo, 1977, 1979; for a review, see Petty & Cacioppo, 1986).

These inconsistencies can be understood, we suggest, by examining which of the steps necessary to avoid contamination (shown in Figure 1) were satisfied in each study. That is, we believe that participants in the studies in which debiasing manipulations failed did not satisfy one or more of the steps necessary to avoid contamination. We consider these steps in turn, discussing how research on forewarning and debiasing illustrates the role of each one.

Increasing awareness of unwanted processing. Consider the first step, becoming aware of a potential bias or contaminant. Studies that attempt to make people aware of biasing information depend on subjects' believing the manipulation (i.e., accepting the fact that their judgments were biased, for example, by discredited testimony in a trial). Such awareness can occur in two ways: through introspection, whereby people discover the bias through an examination of their mental processes, and through the application of a correct causal theory. As argued earlier, the first route to awareness is often very difficult because of people's limited access to their mental processes. Thus, the success of attempts to increase people's awareness of bias should depend in part on the extent to which researchers can convince participants that their judgments are, in fact, open to bias.

One way of accomplishing this, of course, is to use an especially extreme manipulation. Consider, for example, research on the effects of discredited testimony wherein mock jurors hear testimony that is later shown to be inaccurate (e.g., an eyewitness is shown to have poor eyesight and admits to not having worn glasses at the time of the identification). A meta-analysis of this area found that although jurors discount discredited testimony, they fail to do so completely; they adjust their responses significantly, but they do not erase the effects of the discredited testimony entirely (Whitley, 1987). One reason for this might be that people were not sufficiently aware of the extent to which they were influenced by the discredited testimony. In the study

that, perhaps, used the strongest manipulation (the eyewitness ended up admitting the mistake and apologizing to the jury), Hatvany and Strack (1980) found that jurors did adjust more thoroughly for exposure to the invalid testimony. Interestingly, the jurors sometimes overcorrected their responses, indicating that awareness of a bias does not automatically eliminate the effect (we return to this point shortly).

Strong manipulations, however, may not be enough, because the extent to which people will believe forewarning manipulations depends on their theories about bias. The more such a manipulation corresponds to people's causal theories, the more they will believe it, increasing the likelihood that they will attempt to correct the bias. Evidence consistent with this hypothesis was found in a study conducted by Thompson et al. (1981). These researchers instructed jurors to ignore inadmissible evidence that was either proconviction or proacquittal. Interestingly, the jurors discounted only the types of evidence that were consistent with their causal theories about the impact of different kinds of testimony, namely, the proconviction evidence. Thompson et al. speculated that jurors who were exposed to proconviction inadmissible evidence believed that they were biased by it and therefore corrected for it, whereas jurors who were exposed to proacquittal evidence did not believe that they were biased by it—because of an incorrect causal theory—and therefore failed to correct for it.

Similar findings were obtained by Wyer and Budesheim (1987) outside of the courtroom context on a person-perception task. When subjects were instructed to ignore certain pieces of behavioral information, their ability to do so depended on whether the behaviors were positive or negative. When the behaviors in question were positive, people did not always discount them as instructed. When the behaviors were negative, instead, people generally corrected for them and sometimes even overcorrected. Wyer and Budesheim suggested that people tend to assume positive things about others, so when they receive positive information it is in accordance with expectations and is not seen as having a sizable impact. Negative information, in contrast, contradicts expectations and is perceived as particularly damaging. As a result, they argued, people adjust too little when instructed to discount positive information and sometimes adjust too much when instructed to discount negative information. Both lines of research underscore the importance of having the correct causal theory about the nature of the biasing influence.

The motivation to correct for bias. Even if people are aware of a potential bias—either through introspection or because of an accurate causal theory—they must be motivated to correct for it. The lack of motivation to correct may explain the results of another set of studies in the legal context: those that explicitly ask jurors to ignore legally irrelevant information (as opposed to studies that discredit testimony by showing jurors that it is faulty or invalid, as reviewed earlier). Numerous studies have investigated jurors' ability to follow such instructions, and the results are not especially encouraging. Research on the effects of judicial instructions to ignore inadmissible evidence, for example, indicates that jurors do not always discount such evidence completely (e.g., Caretta & Moreland, 1983; Sue et al., 1973; Thompson et al., 1981). Similar failures to follow judicial instructions are evident in research on pretrial publicity (e.g., Kramer, Kerr, & Carroll, 1990; Sue, Smith, & Pedroza, 1975) and joined trials, in which jurors are instructed to consider multiple charges against a defendant in the same trial but to render a verdict on each charge independently (e.g., Greene & Loftus, 1985; Horowitz, Bordens, & Feldman, 1980; Tanford & Penrod, 1982, 1984). Explicitly instructing jurors not to be biased appears to be rather ineffectual.

Why do jurors fail to correct for the effects of information they are instructed to discount? One possibility, as shown earlier, is that people may not possess the causal theory that they were, in fact, influenced by the biasing information. Even if they are aware of the potential influence of the information, however, they may simply not be motivated to resist its effects. Whereas discredited testimony (e.g., an eyewitness who is shown to have poor eyesight) is clearly invalid, inadmissible testimony (e.g., illegal wiretapping evidence) may be seen by jurors as valid and probative of guilt, even though it is legally inadmissible (Hatvany & Strack, 1980). Thus, jurors may choose to use the information despite judicial instructions to the contrary. Indeed, Wolf and Montgomery (1977) found that jurors who were given an especially strong judicial admonishment to ignore inadmissible testimony seemed to exhibit reactance and used the testimony even more than those who were not asked to ignore it. Similarly, outside the courtroom context, Golding has found (Golding et al., 1990; Golding & Hauselt, in press) that instructing people to disregard information about another person because it was confidential or mentioned inappropriately had no effect on people's judgments. People used the information despite the instructions not to, presumably because they viewed it as highly relevant and diagnostic.

These studies illustrate the effects of motivation on people's decisions about whether to use a piece of information. There is considerable evidence that motivation also influences the extent and type of information processing in which people engage, which, in turn, influences their susceptibility to bias (Borgida & Howard-Pitney, 1983; Chaiken, Liberman, & Eagly, 1989; Darley, Fleming, Hilton, & Swann, 1988; Devine, 1989b; Erber & Fiske, 1984; Flink & Park, 1991; Harkness, DeBono, & Borgida, 1985; Kruglanski, 1989a; Neuberg, 1989; Petty & Cacioppo, 1986; Tetlock, 1985). Kruglanski (1989a), for example, found that when people have a low need for closure (i.e., they are motivated to keep an open mind), they are less likely to rely on an initial hypothesis when forming a social judgment. Consequently, they are better able to avoid the biasing effects of stereotypes and first impressions by seeking out additional information. Thus, people's motivation to correct for a bias and, more generally, their motivation to form an accurate judgment are important determinants of the extent to which they will avoid mental contamination.

Awareness of the direction and magnitude of the bias. A number of studies indicate that even when people are aware that information can bias them and are motivated to resist the bias, they adjust their responses either too little or too much. The reason, we suggest, is that they are unaware of how much they have been biased and thus do not know how much to alter their responses. Consider, for example, the classic priming effect, whereby people's judgments shift in the direction of the primed category (e.g., if the category of "kindness" is accessible, people typically rate another person as more kind than they normally

would; Higgins et al., 1977; Srull & Wyer, 1989). Recent studies have shown that making people aware that the category has been primed by an arbitrary event causes them to adjust their responses (Lombardi, Higgins, & Bargh, 1987; Martin, 1986; Martin, Seta, & Crelia, 1990). Interestingly, however, increasing awareness does not make the priming effect disappear; it often reverses, resulting in a contrast effect (Lombardi et al., 1987; Martin, 1986; Martin et al., 1990). For example, if people realize that kind thoughts are accessible for arbitrary reasons, they end up rating the target person as less kind than they normally would.

There is some debate as to why this contrast effect occurs, that is, whether it reflects a change in interpretation of the prime as a proper standard for judgment (Lombardi et al., 1987) or whether it reflects an actual awareness of the biasing effects of the prime (Martin, 1986). Either way, it is clear that even if people are aware of the prime and try to adjust for it, it can be very difficult to adjust "just right," presumably because people are unaware of how much their judgments are influenced by the prime (Higgins & Bargh, 1992). Interestingly, Petty and Wegener (1993) demonstrated that, under some conditions, people recognize that contrast effects occur and try to adjust their responses in the opposite direction but, again, adjust too much (in this case, in the direction of assimilation).

Basic research on impression formation illustrates that the nature of the cognitive system places limits on people's ability to tease apart the effects of invalid information and to correct for any unwanted effects. Schul and Burnstein (1985), for example, found that once people were exposed to invalid information, the success with which they followed instructions to discount it depended on the manner in which the information was originally encoded (see also Golding et al., 1990). When what they called discrete encoding (in which every piece of information is separate and understandable without reference to the other pieces of information in the stimulus set) was used, people followed discounting instructions much more effectively than when integrative encoding (in which individual pieces of information in the stimulus set are elaborated and related to one another) was used. Presumably, integrative encoding interferes with a person's ability to go back later and extract only subsets of information. In our terms, source confusion is more likely to occur when information is encoded integratively, making it difficult to tease apart the effects of specific pieces of information.

In a similar vein, Wyer and his colleagues (Wyer & Budesheim, 1987; Wyer & Unverzagt, 1985) have demonstrated that success in following instructions to disregard information depends on the timing of the instructions, the nature of the information to be disregarded, and the point in the stimulus set at which that information is presented. Srull and Wyer (1989) argued that such effects are consequences of the manner in which person memory operates and is organized. In other words, as we argued earlier, the cognitive system itself places limits on the extent to which people can correct for exposure to information after the fact.

Further evidence for the necessity of knowing how much to correct one's judgments comes from studies on the effects of forewarning people about specific biases, such as the halo effect. As noted earlier, these studies have produced conflicting results. We suggest that these contradictory findings can be understood by considering the last two stages depicted in Figure 1: people's awareness of the direction and amount of bias and their ability to control their responses. It is unlikely that the first two conditions necessary to avoid contamination are involved. People recognize in these studies that their mental processes might be biased; indeed, the whole point of the manipulations is to warn people about such bias. And manipulation checks in several of these studies have shown that people believed the forewarning manipulations, thereby satisfying the first condition (awareness of the unwanted mental processing).

Motivation to avoid bias may enhance the impact of forewarning, as indicated by the fact that, in persuasion studies, subjects who are more personally involved with an attitude issue show the most resistance to persuasive appeals after forewarning (Chen, Reardon, Rea, & Moore, 1992; Petty & Cacioppo, 1979). If one assumes that personally involved subjects are the most motivated to avoid having their attitudes changed, this suggests that motivation to avoid mental contamination could play a role in the effectiveness of forewarning.[7] However, motivation is not sufficient to make forewarning effective, as indicated by forewarning studies in which people were explicitly motivated to avoid the bias but in which forewarning had no effect (e.g., Wetzel et al., 1981).

We suggest that one major difference between those studies in which forewarning worked and those in which it did not involves people's awareness of the direction and magnitude of the bias. In most of the studies that failed to find a corrective effect of forewarning, people had difficulty teasing apart two or more influences on their responses; that is, they were faced with clear cases of source confusion, making it difficult to determine the extent to which they were biased by the unwanted source of information. In the Wetzel et al. (1981) halo effect study, for example, people had difficulty determining how much of a professor's physical appearance was due to their objective appraisal of his physical attributes versus their liking for the professor, even when they were forewarned about the halo effect. Thus, they had no way of knowing how much to adjust for the bias. Moreover, it seems likely that even if people did have an intuitive notion of how much to correct for the halo effect, they would have no idea how to prevent the bias from occurring. Lay theories of the mind are unlikely to contain information regarding how to avoid a halo effect when forming an impression of someone.

In contrast, when forewarning is effective (e.g., in the literature on persuasion), the source of bias is much easier to distinguish from other influences; that is, people are able to code the persuasive appeal as a discrete unit that they know is a source of the potential influence on their attitudes. Furthermore, people have knowledge of their attitudes in the absence of the persuasive appeal. This puts them in a better position to assess how much impact the contaminant is having on them. In addition, the process by which persuasion occurs is at least partially conscious, so people have some knowledge of how effective it is.

[7] It is also possible that only the personally involved subjects actually fit our definition of mental contamination. Presumably, personally involved subjects view attempts to change their attitudes as undesirable, whereas less involved subjects may or may not mind being influenced.

We suggest that the preceding conditions are unlikely to hold in other areas of mental contamination in which invalid information often comes intertwined with valid information, the process by which one is contaminated is poorly understood or operates largely outside of awareness, or people find it difficult to know how they would have felt in the absence of the contamination. Because it is more difficult in other areas to recognize the amount of contamination that has occurred, forewarning people about this bias is less effective.

The ability to control one's response. There is another potential reason why forewarning works better in the persuasion literature: People have a clear-cut strategy with which to counteract persuasive messages, that of generating counterarguments. That is, people can generate arguments that negate the effects of the persuasive appeal, giving them some control over their responses (i.e., their attitudes). Consistent with this notion, research shows that forewarning produces resistance to persuasion through counterarguing; people bolster their initial opinions by generating arguments in support of them (Petty & Cacioppo, 1986). Furthermore, it is worth noting that forewarning is not effective when people are unable to generate counterarguments (e.g., if forewarned subjects are distracted [Chen et al., 1992], are not given sufficient time before presentation of the persuasive appeal to generate counterarguments [Petty & Cacioppo, 1977], or have no counterarguments available, as in the case of cultural truisms [McGuire & Papageorgis, 1962]).

In contrast, there is no known strategy that people can adopt, in other forewarning studies, to control their responses to the unwanted influence. Even if people know that they are susceptible to a halo effect, for example, it is unclear what strategy they could adopt to avoid its unwanted influence. In this sense, people have less control over their responses to the contaminant.[8]

There are ways in which people can control their responses other than adopting specific strategies to ward off the influence (e.g., counterarguing). As mentioned earlier, some responses (e.g., overt behavior) are inherently easier to control than others (e.g., affective reactions). Consequently, forewarning should work better when people are asked to avoid a bias on a response that is relatively easy to control (Clore & Parrott, 1991; Strack, 1992). Schul and Manzury (1990) found precisely these results in a study of debiasing in the courtroom. They presented mock jurors with evidence damaging to the defendant in an assault and battery case but then told the jurors to ignore this information because the judge had ruled it to be inadmissible. Interestingly, participants were unable to ignore the discredited information when forming an impression of the defendant: They had a more negative impression than people in a control condition who did not hear the damaging testimony. Participants were, however, able to ignore the damaging testimony on a response that was easier to control: their verdict in the case. Participants were no more likely to convict the defendant than were control subjects. That is, presumably it was easier to control one's conscious, dichotomous decision about whether to vote guilty or not guilty than to control one's relatively automatic affective reaction, making the former judgment more immune from bias.

As a final example of the importance of being able to control one's response, consider a series of studies by Kubovy (1977; Kubovy & Psotka, 1976) on spontaneous number generation. Kubovy found that the frequency with which people report the first number that comes to mind can be increased with subtle primes. For example, when people are asked to give "the first *one*-digit [italics added] number that comes to mind" (Kubovy, 1977, p. 360), the frequency with which they give the number one increases. If, however, the question is asked in a way that makes people more aware of the biasing nature of the prime, people are able to resist its influence. For example, when people were asked to give "the first number that comes to mind between 0 and 9 . . . *like one* [italics added]" (Kubovy, 1977, p. 361), there was no increase in the frequency with which they gave one as their answer. Why? We suggest that the number one probably did come to mind first for many subjects in this latter condition, but because they recognized the "biasing" role of the experimenter's prompt and could easily control their overt response, they reported another number that seemed more spontaneous. When people have less control over their responses—such as their affective reactions—they are less able to avoid being influenced by a contaminant, even when they try (e.g., Wetzel et al., 1981).

Unresolved Issues

There is considerable evidence, then, that forewarning and debiasing manipulations are most likely to work when they satisfy each of the conditions illustrated in Figure 1: They make people aware of the unwanted processing, they motivate people to resist it, and people are aware of the direction and magnitude of the bias and have sufficient control over their responses to correct for it. There are, however, some unresolved issues about exactly when and how mental contamination can be avoided. In the preceding discussion about the effects of forewarning, we assumed that correction occurs through a fairly conscious process involving awareness of the bias and deliberate attempts to reduce it. We have discussed in some detail when such conscious, deliberate debiasing attempts will be successful.

There may be times, however, when correction occurs less consciously. Consider a study conducted by Schwarz and Clore (1983) in which people's judgments of life satisfaction and happiness were biased by their current moods. In one study, people gave higher ratings of life satisfaction on sunny days than rainy days, presumably because the effects of the weather on their mood biased their judgments of life satisfaction. However, in some conditions, the researchers eliminated this bias by making the weather more salient (e.g., by asking people about the weather immediately before assessing their life satisfaction). When the weather was salient, people were able to "partial out" its effects on their satisfaction judgments. Similarly, numerous

[8] An interesting question is whether there are any strategies that would succeed in negating a halo effect. One possibility is that because the halo effect operates automatically and nonconsciously, there are no strategies that people can use to prevent it from occurring. Another possibility is that the correct strategy has not been discovered. For example, is there something analogous to counterarguing, such as generating negative thoughts about a person, that will successfully avoid a positive halo effect? Our point is simply that some responses are more controllable than others, making them less susceptible to unwanted influence.

studies on the misattribution of emotion have found that when a nonemotional explanation for one's arousal is salient (e.g., it is clear that the arousal is due to a drug or to physical exercise), people do not misattribute their arousal to an emotional source (e.g., Schachter & Singer, 1962; Zillmann, 1978).

These studies differ from the forewarning ones we reviewed earlier in that the adjustment process probably occurred less consciously and deliberately (Clore & Parrott, 1991). That is, it is unlikely that participants in misattribution studies deliberately reflect about the causes of their responses; instead, the attribution process probably occurs quickly and nonconsciously, at least at times (Nisbett & Wilson, 1977b). To fully understand when mental contamination will occur, then, it may be necessary to distinguish between influences and adjustments at what Strack (1992) referred to as the "judgment phase," in which people form a specific judgment (at least partly nonconsciously), and the "representativeness check and correction" stage, in which people go back and check the validity of their judgments, correcting for any perceived biases. The processes we have been discussing (i.e., those illustrated in Figure 1) refer to the latter stage, in which people make deliberate attempts to adjust for possible biases on their judgments. It is important to acknowledge, however, that nonconscious processing might also influence whether contamination occurs.

The extent to which attribution processes occur at the prior, nonconscious stage versus the latter, conscious, correction phase is controversial. Although we suspect that many attribution processes are quick, implicit, and nonconscious (Nisbett & Wilson, 1977b; Wilson, 1985), others suggest that these processes—such as the correction for the weather performed by Schwarz and Clore's (1983) subjects—occur more consciously and deliberately (Strack, 1992). A fruitful area for future research would be to examine this question more closely by, for example, determining whether placing people under cognitive load eliminates the correction found by Schwarz and Clore (1983). If not, it would seem that people are able to partial out potential contaminants at an initial judgment stage, as opposed to a later correction stage, if their attention is drawn to these contaminants. Such a finding would raise several intriguing questions, such as the extent to which people's causal theories can operate at a nonconscious level, influencing automatic attribution processes.

At this point, such questions remain highly speculative. They do suggest, however, that the kinds of correction processes we have discussed (e.g., those illustrated in Figure 1) may need to be expanded to include processes that are less conscious and deliberate and that succeed in preventing contamination before it has had a chance to occur.

Correcting Unwanted Consequences of Automatic Processing Versus Source Confusion

The issue just discussed—the extent to which people can automatically adjust for contaminants before contamination occurs—is relevant to our distinction between two types of mental contamination: the unwanted consequences of otherwise functional automatic processing versus source confusion. As noted earlier, both types involve nonconscious, automatic processing. Source confusion is distinct, however, in that it always involves misattribution (mistaken inferences about the causes of one's responses) and typically does not involve cognitive processes that are otherwise functional. That is, whereas the functional advantage of many of the automatic processes we have discussed is clear (e.g., categorization), the inability to detect the relative contributions of different sources to one's judgments and feelings is not. Consequently, there may be differences in how people can best avoid source confusion versus the unwanted by-products of functional automatic processes.

If people are able to make nonconscious, automatic attributional adjustments, as we have just suggested, it may be possible, under some conditions, to avoid source confusion before it occurs. For example, in the classic Schachter and Singer (1962) experiment, subjects who were told that the placebo caused side effects of arousal did not misattribute this arousal to an emotional source, and this process probably occurred at an implicit level. People did not have to go back after the fact and adjust their responses; misattribution was "headed off at the pass."

Even if this (admittedly speculative) nonconscious attributional process does not succeed in avoiding source confusion, there are other ways that it can be avoided or undone. We have discussed those cases wherein the unwanted source of influence is blatant and easy to counteract, as in the case of a counterattitudinal persuasive communication. Here, people have well-developed strategies (e.g., counterarguing) to negate the unwanted effects of the contaminant.

In contrast, it may be more difficult to avoid the undesirable by-products of otherwise functional automatic processes. By definition, these processes are adaptive and useful, such as the processes of categorization and belief formation. Thus, it would not be desirable to try to disrupt these processes altogether. Instead, it may be most useful for people to develop accurate theories about the operation of such processes so that they can initiate controlled processing to counteract them when necessary. For example, if nonprejudiced people automatically categorize a member of a stigmatized group in a negative way, as Devine (1989a) argued they do, those who are aware that such a process occurs will be more likely to try to correct it. Indeed, the perceiver who is extremely well informed about automatic processing and the ability of controlled processing to counteract it should be more likely to avoid cognitive busyness, thereby allowing sufficient mental capacity for the controlled correction processes to occur (Gilbert, 1991).

Nonetheless, as argued earlier, there is reason for pessimism about the ability of such correction processes to completely correct for the unwanted effects of automatic processing. Similarly, there are many examples of instances in which people are unable to avoid source confusion, even when forewarned. It can simply be too difficult to know the extent of the bias and to control one's responses sufficiently to correct for the bias. If the unwanted information is external to oneself, however, there is another option that we have not yet discussed: avoiding exposure to this information.

Exposure Control

A final strategy for avoiding mental contamination is to make sure that it never has the opportunity to occur by avoiding contaminants that might bias one's judgments. This technique,

termed *exposure control* by Gilbert (1993), is often imposed on people by authorities. For example, the rules of evidence dictate that juries may not hear certain information so that they will not be biased, and some professional journals use a masked review procedure whereby reviewers are unaware of the identity of the manuscript's author. We know of very few studies that have examined whether individuals use such a strategy as a way of controlling their thoughts and feelings to avoid mental contamination.

It is clear that people sometimes use exposure control for other purposes. One of the original predictions of dissonance theory was that people avoid exposure to information that casts doubt on a choice they have just made (Festinger, 1957). Although initial research failed to confirm this hypothesis (Freedman & Sears, 1965), later research has shown that it holds under certain conditions (Frey, 1986). Research on self-verification theory also shows that under some conditions people avoid information that contradicts their views of themselves, even if the information is flattering (Swann, 1990).

The important point for our purposes concerns people's motivation to avoid exposure to information. According to both dissonance and self-verification theories, the motive is to avoid psychological discomfort, namely the unpleasantness of discovering that one's choice was the wrong one or the loss of predictability and control that results from a challenge to the self-concept. (According to self-verification theory, there is another motive as well: the pragmatic concern of keeping social interactions on an even and predictable keel.) Our concern is with a different motive: Do people ever avoid information as a strategy to avoid mental contamination? That is, independent of people's need to reduce dissonance or self-verify, do they limit their exposure because they believe that they cannot help but be influenced by the information in unwanted ways?

We suspect that the answer to this question depends on the kind of response people are trying to protect: "hot" affective responses versus "cold" inferences about the social world. We have been discussing mental contamination in its broadest sense, including in our definition unwanted judgments, inferences, emotions, and actions. To understand people's use of exposure control as a strategy to avoid contamination, however, we need to consider people's theories about how controllable these different kinds of responses are. Most people, we suspect, believe that their affective reactions (e.g., moods and emotions) are difficult to control and are thus willing to avoid information that might have unwanted effects on these affective states. If people are in a good mood, for example, they are, at least under some conditions, likely to avoid activities that are likely to dampen their spirits (Isen, 1987; Wegner & Erber, 1993). When it comes to judgments and inferences, most people assume that they can regulate their acceptance of new facts. If a piece of new information turns out to be invalid or of questionable validity, they can simply dismiss it from further consideration, or so they think.

As shown by Gilbert (1991, 1993), this faith in one's ability to control beliefs is misplaced in that people sometimes cannot help believing false propositions. This may be especially true as time goes by, because sometimes people remember a fact longer than they do accompanying information about its validity (or lack thereof). For example, when people hear a persuasive communication and learn that the source of the message has little credibility, they are, not surprisingly, unpersuaded by the message. Over time, however, the content of the message remains in memory longer than the information about the credibility of the source. As a result, a "sleeper effect" occurs whereby people are influenced more by the communication at a later time—when they have forgotten that the source was not credible—than they were at the time they heard the communication (Pratkanis, Greenwald, Leippe, & Baumgardner, 1988).

Thus, if people do not want to be influenced by a persuasive communication that is not credible, they would be better off not hearing it at all. People seem not to appreciate this fact, however, thereby exposing themselves to information that will influence them in unwanted ways. Consider, for example, a study conducted by Wegner, Klein, and Dimitri (1992). These investigators asked subjects to select information about a target person that was known to be true, false, or uncertain to help them predict the target person's behavior. Subjects liberally selected all three types of information, confident that they could keep track of the truth value of each fact. Interestingly, however, this confidence was misplaced: As time passed, people had difficulty remembering which facts were true, causing them to form questionable inferences about the target person. Because of the fallibility of their memories, people would have been better off using exposure control, selecting only information known to be true.

As with the other ways of avoiding mental contamination we have discussed, however, exposure control has some important limitations. First, it is likely to work only if the information that biases people's judgments is external and easily controllable. This might be the case with potential contaminants such as advertising or invalid information about another person, but it is not the case with other kinds of contamination. Many of the examples of contamination listed in Appendix B, for example, involve internal contaminants, such as moods and memories, that are very difficult to avoid.

Second, even if the contaminant is external, it is sometimes difficult to know in advance whether it is potentially biasing and thus should be avoided (Gilbert, 1993). People may occasionally know that information is likely to be suspect (e.g., when they see a headline in a supermarket tabloid arguing that drinking milk causes cancer) and thus know that they should read no further. More often, however, people do not know the validity of information before they receive it; information rarely comes with warning labels. A friend might discuss the latest studies on milk consumption and cancer only to mention, when questioned, that his source is a supermarket tabloid. If people decide to avoid all information that is potentially untrue or invalid, they will miss out on a good deal of true, valid information as well.

Third, an exposure control strategy might cause people to examine only information that confirms their views, thereby fostering narrow-mindedness, stifling creativity, and inhibiting social change. Fourth, to avoid exposure to unwanted information, people would have to be ever vigilant, ready to shut their eyes and cover their ears whenever they suspected that potentially unwanted information was in the vicinity. As noted by William James (1897),

> Our errors are surely not such awfully solemn things. In a world

> where we are so certain to incur them in spite of all our caution, a certain lightness of heart seems healthier than this excessive nervousness on their behalf. (p. 19)

We agree with James that people should not be in a state of constant, excessive nervousness, unable to carry out the most mundane task for fear of forming a biased judgment. On the other hand, we believe that sometimes mental errors are, in fact, "awfully solemn things" that are quite costly to ourselves and others. One possible solution to this dilemma would be to allow others to control our exposure to certain kinds of mental contaminants, as is now done with such procedures as the masked review of manuscripts. The institutionalization of exposure control, however, raises some thorny issues about freedom of expression. If one accepts the premise that exposure control is an effective means of avoiding mental contamination, does that mean that governments should allow the censorship of some kinds of information to protect the welfare of their citizens? Research on mental contamination is, in some ways, on a collision course with the First Amendment right to freedom of expression.

We do not profess to have "the answer" to this dilemma and can conclude only by pointing to some of the difficult issues it raises. Consider, for example, the powerful effects of advertising. There is evidence that advertising has a large impact on people's attitudes and behavior (e.g., Abraham & Lodish, 1990; Liebert & Sprafkin, 1988; Ryan, 1991), and most people would prefer not to have their attitudes and behavior shaped by advertising. An effective way of avoiding this form of mental contamination is with exposure control, whereby people's exposure to advertising is limited. But whose job is it to implement such control? A third party such as the federal government, thereby involving a considerable amount of censorship, or individuals themselves? Leaving exposure control to a third party would be the most effective means of eliminating unwanted effects of advertising but would involve massive censorship. Leaving it up to the informed individual to decide what to see would probably be insufficient to eliminate all unwanted effects of advertising. The invention of the remote control has made such exposure control easier, but it would be nearly impossible to eliminate all advertising from people's lives. And sitting on the couch poised with the remote control, ever ready to protect our delicate minds from unwanted advertising, sounds dangerously close to the excessive nervousness that concerned James.

Perhaps the best strategy is to teach individuals when it is important to engage in exposure control. For example, the next time we go to the ice cream parlor, it hardly seems worth it to worry that our decision about which flavor of ice cream to buy might be biased or contaminated. The costs of making a decision we regret are not worth the time and effort that would go into making sure that our judgments are uncontaminated (e.g., making sure that an earlier discussion with a friend about the *Rocky Horror Picture Show* did not inadvertently prime rocky road ice cream or that the order in which the flavors are listed does not somehow bias our choice). When making important decisions, however—such as which job to accept or whether to seek medical treatment—a dose of nervousness may be worthwhile if it prevents decisions with severe negative consequences.

Comparison With Recent Approaches to Correction Processes

There have been a number of recent attempts to describe how people correct biased judgments for contextual factors, and it is useful to consider the similarity and differences between these approaches and ours. One difference concerns the breadth of the models. Many other approaches, such as Martin's set/reset model (Martin, 1986; Martin & Achee, 1992), Schwarz and Bless's (1992) inclusion/exclusion model, and Petty and Wegener's (1993) flexible correction model, are concerned with the conditions under which assimilation and contrast effects occur in social judgment. That is, each of these models attempts to explain the conditions under which priming a social category will lead to assimilation (judgments in the direction of the prime) versus contrast (judgments away from the prime). We are concerned with correction more broadly, namely instances in which people believe that their response has been biased inappropriately in any fashion, be it assimilation or contrast effects, halo effects, the unwanted effects of a persuasive communication, or the unwanted application of a stereotype (in short, all examples of the unwanted consequences of automatic processing and source confusion; see Appendix B).

We believe that some of the principles of models of assimilation and contrast can be applied to this broader class of cases in which people try to correct their judgments for any unwanted influence. Specifically, there are two chief areas of agreement between these models and our broader model of mental correction. The first is the assumption that successful correction depends on accurate lay beliefs about mental contamination. Most models of assimilation and contrast assume, at least implicitly, that if people are to correct for the unwanted effects of context on judgment, they must have an accurate belief about the direction and magnitude of these unwanted effects. This assumption is stated most explicitly in Petty and Wegener's (1993) flexible correction model, in their hypothesis (and supporting data) that the direction of people's correction depends on their naive theories of assimilation and contrast. It is also part of Strack's (1992) discussion of "representativeness checks," in which people are said to check their judgments to determine whether they are appropriate and unbiased. People's lay beliefs about bias are an integral part of our model of mental correction, as seen in the second (awareness of the unwanted processing) and fourth (awareness of the direction and magnitude of the bias) boxes in Figure 1.

Second, we share with other approaches an emphasis on people's ability and motivation to correct for unwanted influences (see the fifth box in Figure 1). A number of other approaches have similarly noted that judgments will be biased if people do not have the cognitive capacity or the motivation to adjust their responses (e.g., Gilbert, 1991; Kruglanski, 1989a; Martin, 1986; Petty & Cacioppo, 1986).

The chief differences between our approach and other models of correction are largely matters of emphasis. First, as already noted, our model is broader, encompassing all instances of unwanted influence rather than only assimilation and contrast effects in social judgment. Second, we stress people's lay beliefs more explicitly, arguing that these beliefs have been largely neglected by researchers and are a key determinant of whether

people correct successfully. Third, we have included in our model a means of avoiding bias other than trying to correct a judgment after exposure to a stimulus: exposure control, whereby people may decide to avoid the stimulus altogether. Fourth, we are, perhaps, more pessimistic in our assessment of how successful correction is likely to be. We emphasize the basic nature of the cognitive system, more than other approaches, that limits people's ability to correct (e.g., a lack of awareness of mental processes and the difficulty of controlling such processes).

Summary and Conclusions

We have argued that there are two distinctive classes of mental errors: those that stem from the failure to know or apply a useful inferential rule and those that are the result of mental contamination (in which a person has an unwanted judgment, emotion, or behavior because of mental processing that is unconscious or uncontrollable). There is some cause for optimism concerning people's ability to avoid the first kind of error in that recent research has shown that some inferential rules can be taught successfully. We are more pessimistic about people's ability to avoid mental contamination because of the nature of human cognition (e.g., imperfect access to mental processes) and lay beliefs about the mind (e.g., people's faith that they can control their beliefs better than can other people and the poor calibration of theories about contamination).

One interesting question that is beyond our scope concerns the cultural relativity of beliefs about contamination. Other cultures may have very different conceptions of bias and error and, consequently, take different steps to avoid mental contamination. The correspondence between the mental processes that produce bias and cultural beliefs about these processes may well differ across cultures and would make a fascinating topic for future research. At least in Western countries such as the United States, there appears to be a lack of such correspondence, making people more susceptible to unwanted bias in many of their responses.

Clearly, many of the topics we have discussed are in their infancy. More work is needed to address questions such as the following: Is it possible to educate people about sources of mental contamination, thereby improving the accuracy of lay theories? If so, will people be sufficiently armed to avoid consequential mental errors? One implication of our discussion is that forewarned is not always forearmed; even with a better understanding of mental contamination, people might not be able to overcome it. If so, can people successfully use exposure control to avoid bias, or is it simply too difficult to filter out information that is potentially biasing? Are there other, better ways of reducing mental contamination? Research on questions such as these should illuminate how people can best avoid unwanted influences on their judgments, emotions, and behaviors.

References

Abraham, M. M., & Lodish, L. M. (1990). Getting the most out of advertising and promotion. *Harvard Business Review, 68,* 50–58.

Allport, G. W. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Anderson, C. A. (1989). Causal reasoning and belief perseverance. In D. W. Schumann (Ed.), *Proceedings of the Society for Consumer Psychology* (pp. 115–120). Knoxville: University of Tennessee.

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.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35,* 124–140.

Auletta, K. (1993, May 17). Annals of communication: What won't they do? *New Yorker,* pp. 45–53.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A. (1990). Auto-motives: Preconscious determinants of social interaction. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition* (Vol. 2, pp. 93–130). New York: Guilford Press.

Baron, J. (in press). Nonconsequentialist decisions. *Behavioral and Brain Sciences.*

Billig, M. (1985). Prejudice, categorization, and particularization: From a perceptual to a rhetorical approach. *European Journal of Social Psychology, 15,* 79–103.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Borgida, E., & Howard-Pitney, B. (1983). Personal involvement and the robustness of perceptual salience effects. *Journal of Personality and Social Psychology, 45,* 560–570.

Bornstein, R. F., & D'Agostino, P. R. (1990). *A perceptual fluency/attributional model of the mere exposure effect.* Unpublished manuscript, Gettysburg College, Gettysburg, PA.

Brewer, M. B. (1989). A dual process model of impression formation. In R. S. Wyer & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Caretta, T. R., & Moreland, R. L. (1983). The direct and indirect effects of inadmissible evidence. *Journal of Applied Social Psychology, 13,* 291–309.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

Chapman, L., & Chapman, J. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72,* 193–204.

Chapman, L., & Chapman, J. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 249–260.

Chen, H. C., Reardon, R., Rea, C., & Moore, D. J. (1992). Forewarning of content and involvement: Consequences for persuasion and resistance to persuasion. *Journal of Experimental Social Psychology, 28,* 523–541.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328.

Clore, G. L., & Parrott, W. G. (1991). Moods and their vicissitudes: Thoughts and feelings as information. In J. Forgas (Ed.), *Emotion and social judgment* (pp. 107–123). Elmsford, NY: Pergamon Press.

Cohen, L. J. (1977). *The probable and the provable.* New York: Oxford University Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317–331.

Conee, E., & Feldman, R. (1983). Discussion: Stich and Nisbett on justifying inference rules. *Philosophy of Science, 50,* 326–331.

Damrad-Frye, R., & Laird, J. D. (1989). The experience of boredom: The role of the self-perception of attention. *Journal of Personality and Social Psychology, 57,* 315–320.

Darley, J. M., Fleming, J. H., Hilton, J. L., & Swann, W. B., Jr. (1988). Dispelling negative expectancies: The impact of interaction goals and target characteristics on the expectancy confirmation process. *Journal of Experimental Social Psychology, 24,* 19–36.

Devine, P. G. (1989a). Automatic and controlled processes in prejudice: The roles of stereotypes and personal beliefs. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 181–212). Hillsdale, NJ: Erlbaum.

Devine, P. G. (1989b). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology, 30,* 510–517.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* 52–88.

Erber, R., & Fiske, S. T. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology, 38,* 889–906.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fischhoff, B. (1977). The perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 349–358.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge University Press.

Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 253–283). New York: Guilford Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906–911.

Flink, C., & Park, B. (1991). Increasing consensus in trait judgments through outcome dependency. *Journal of Experimental Social Psychology, 27,* 453–467.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18,* 253–292.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General, 120,* 34–45.

Freedman, J. L., & Sears, D. O. (1965). Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 57–97). San Diego, CA: Academic Press.

Frey, D. (1986). Recent research on selective exposure to information. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 41–80). San Diego, CA: Academic Press.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–91.

Gerbasi, K. C., Zuckerman, M., & Reis, H. T. (1977). Justice needs a new blindfold: A review of mock jury research. *Psychological Bulletin, 84,* 323–345.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107–119.

Gilbert, D. T. (1993). The assent of man: Mental representation and the control of belief. In D. M. Wegner, & J. W. Pennebaker (Eds.), *The handbook of mental control* (pp. 57–87). Englewood Cliffs, NJ: Prentice Hall.

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.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reasoning in everyday life.* New York: Free Press.

Golding, J. M., Fowler, S. B., Long, D. L., & Latta, H. (1990). Instructions to disregard potentially useful information: The effects of pragmatics on evaluative judgments and recall. *Journal of Memory and Language, 29,* 212–227.

Golding, J. M., & Hauselt, J. (in press). When instructions to forget become instructions to remember. *Personality and Social Psychology Bulletin.*

Goodman, N. (1965). *Fact, fiction, and forecast* (2nd ed.). New York: Bobbs-Merrill.

Greene, E., & Loftus, E. F. (1985). When crimes are joined at trial: Institutionalized prejudice? *Law and Human Behavior, 9,* 193–208.

Harkness, A. R., DeBono, K. G., & Borgida, E. (1985). Personal involvement and strategies for making contingency judgments: A stake in the dating game makes a difference. *Journal of Personality and Social Psychology, 49,* 22–32.

Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193–208). Cambridge, England: Cambridge University Press.

Hatvany, N., & Strack, F. (1980). The impact of a discredited key witness. *Journal of Applied Social Psychology, 10,* 490–509.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94,* 319–340.

Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 75–123). New York: Guilford Press.

Higgins, E. T., & Bargh, J. A. (1992). Unconscious sources of subjectivity and suffering: Is consciousness the solution? In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 67–103). Hillsdale, NJ: Erlbaum.

Higgins, E. T., & King, G. (1981). Accessibility of social constructs: Information-processing consequences of individual and contextual variability. In N. Cantor & J. F. Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 69–121). Hillsdale, NJ: Erlbaum.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13,* 141–154.

Hornstein, H. A., LaKind, E., Frankel, G., & Manne, S. (1975). Effects of knowledge about remote social events on prosocial behavior, social conception, and mood. *Journal of Personality and Social Psychology, 32,* 1038–1046.

Horowitz, A., Bordens, K. S., & Feldman, M. S. (1980). A comparison of verdicts obtained in severed and joined trials. *Journal of Applied Social Psychology, 10,* 444–456.

Isen, A. M. (1987). Positive affect, cognitive processes, and social behav-

ior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 203–253). San Diego, CA: Academic Press.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin, 13,* 314–336.

Jacoby, L. L., Lindsay, S. D., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist, 47,* 802–809.

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118,* 115–125.

James, W. (1897). *The will to believe and other essays in popular philosophy.* New York: Longmans, Green.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Jones, E. E. (1990). *Interpersonal perception.* San Francisco: Freeman.

Kahneman, D., Slovic, P, & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1983). Can irrationality be intelligently discussed? *Behavioral and Brain Sciences, 6,* 509–510.

Kassin, S. M. (1983). Deposition testimony and the surrogate witness: Evidence for a "messenger effect" in persuasion. *Personality and Social Psychology Bulletin, 9,* 281–288.

Kassin, S. M., & Wrightsman, L. S. (1988). *The American jury on trial: Psychological perspectives.* New York: Hemisphere.

Keysar, B., Ginzel, L. E., & Bazerman, M. H. (1992). *States of affairs and states of mind: The curse of knowledge of beliefs.* Unpublished manuscript, University of Chicago.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237,* 1445–1452.

Kramer, G. P., Kerr, N. L., & Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior, 14,* 409–438.

Kruglanski, A. W. (1989a). *Lay epistemics and human knowledge: Cognitive and motivational bases.* New York: Plenum.

Kruglanski, A. W. (1989b). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin, 106,* 395–409.

Krull, D. S. (1993). Does the grist change the mill? The effect of the perceiver's inferential goal on the process of social inference. *Personality and Social Psychology Bulletin, 19,* 340–348.

Kubovy, M. (1977). Response availability and the apparent spontaneity of numerical choices. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 359–364.

Kubovy, M., & Psotka, J. (1976). The predominance of seven and the apparent spontaneity of numerical choices. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 291–294.

Kurtz, R. M., & Garfield, S. L. (1978). Illusory correlation: A further exploration of Chapman's paradigm. *Journal of Consulting and Clinical Psychology, 46,* 1009–1015.

Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29,* 299–304.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science, 1,* 362–370.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning. *American Psychologist, 43,* 431–442.

Lewicki, P. (1986). *Nonconscious social information processing.* San Diego, CA: Academic Press.

Liebert, R. M., & Sprafkin, J. (1988). *The early window* (3rd ed.). Elmsford, NY: Pergamon Press.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D. (1989). Automaticity and cognitive control. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 52–74). New York: Guilford Press.

Lombardi, W. J., Higgins, E. T., & Bargh, J. A. (1987). The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task. *Personality and Social Psychology Bulletin, 13,* 411–429.

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica, 64,* 167–185.

Marshall, G. D., & Zimbardo, P. G. (1979). Affective consequences of inadequately explained physiological arousal. *Journal of Personality and Social Psychology, 37,* 970–988.

Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology, 51,* 493–504.

Martin, L. L., & Achee, J. W. (1992). Beyond accessibility: The role of processing objectives in judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 195–216). Hillsdale, NJ: Erlbaum.

Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology, 59,* 27–37.

Maslach, C. (1979). Negative emotional biasing of unexplained arousal. *Journal of Personality and Social Psychology, 37,* 953–969.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90,* 215–238.

McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly, 26,* 24–34.

Medin, D. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469–1481.

Merikle, P. M. (1988). Subliminal auditory messages: An evaluation. *Psychology and Marketing, 5,* 297–316.

Miller, G. A., & Cantor, N. (1982). [Review of *Human inference: Strategies and shortcomings of social judgment*]. *Social Cognition, 1,* 83–93.

Mishan, E. J. (1976). *Cost-benefit analysis.* New York: Praeger.

Moore, T. E. (1982). Subliminal advertising: What you see is what you get. *Journal of Marketing, 46,* 38–47.

Morgan, J. N., & Duncan, G. J. (1982). *Making your choices count: Economic principles for everyday decisions.* Ann Arbor: University of Michigan Press.

Nelson, T. O. (Ed.). (1992). *Metacognition: Core readings.* Needham Heights, MA: Allyn & Bacon.

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology, 56,* 374–386.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238,* 625–631.

Nisbett, R. E., Krantz, D. H., Jepson, D., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Nisbett, R. E., & Ross, L. D. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R. E., & Schachter, S. (1966). Cognitive manipulation of pain. *Journal of Experimental Social Psychology, 2,* 227–236.

Nisbett, R. E., & Wilson, T. D. (1977a). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35,* 250–256.

Nisbett, R. E., & Wilson, T. D. (1977b). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Olson, J. M. (1990). Self-inference processes in emotion. In J. M. Olson & M. P. Zanna (Eds.), *Self-inference processes: The Ontario Symposium* (Vol. 6, pp. 17–41). Hillsdale, NJ: Erlbaum.

Petty, R. E., & Cacioppo, J. T. (1977). Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology, 35,* 645–655.

Petty, R. E., & Cacioppo, J. T. (1979). Effects of forewarning of persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin, 5,* 173–176.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology, 29,* 137–165.

Posner, M. I., & Rothbart, M. K. (1989). Intentional chapters on unintended thoughts. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 450–469). New York: Guilford Press.

Pratkanis, A. R., & Greenwald, A. G. (1988). Recent perspectives on unconscious processing: Still no marketing applications. *Psychology and Marketing, 5,* 337–352.

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead. Long live the sleeper effect. *Journal of Personality and Social Psychology, 54,* 203–218.

Price, H. H. (1954). Belief and will. *Aristotelian Society, 28*(Suppl.), 1–26.

Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin, 94,* 239–264.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). San Diego, CA: Academic Press.

Ross, L., Greene, D., & House, P. (1977). The false consensus phenomenon: An attributional bias in self-perception and social-perception processes. *Journal of Experimental Social Psychology, 13,* 279–301.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880–892.

Ross, L., & Nisbett, R. E., (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.

Rozin, P. (1990). Social and moral aspects of food and eating. In I. Rock (Ed.), *The legacy of Solomon Asch: Essays in cognition and social psychology* (pp. 97–110). Hillsdale, NJ: Erlbaum.

Rozin, P., & Nemeroff, C. (1990). The laws of sympathetic magic: A psychological analysis of similarity and contagion. In J. W. Stigler, R. A. Shweder, & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human judgment* (pp. 205–232). Cambridge, England: Cambridge University Press.

Ryan, B., Jr. (1991). *It works! How investment spending in advertising pays off.* New York: American Association of Advertising Agencies.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotion. *Psychological Review, 69,* 379–399.

Schul, Y. (1993). When warning succeeds: The effects of warning on success in ignoring invalid information. *Journal of Experimental Social Psychology, 29,* 42–62.

Schul, Y., & Burnstein, E. (1985). When discounting fails: Conditions under which individuals use discredited information in making a judgment. *Journal of Personality and Social Psychology, 49,* 894–903.

Schul, Y., & Manzury, F. (1990). The effects of type of encoding and strength of discounting appeal on the success of ignoring an invalid testimony. *European Journal of Social Psychology, 20,* 337–349.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Hillsdale, NJ: Erlbaum.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 189–286). Hillsdale, NJ: Erlbaum.

Smith, E., & Medin, D. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E. R., & Miller, F. D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review, 85,* 355–362.

Srull, T. K., & Wyer, R. S., Jr. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37,* 1660–1672.

Srull, T. K., & Wyer, R. S., Jr. (1989). Person memory and judgment. *Psychological Review, 96,* 58–83.

Stich, S. P. (1990). *The fragmentation of reason.* Cambridge, MA: Bradford.

Stich, S. P., & Nisbett, R. E. (1980). Justification and the psychology of human reasoning. *Philosophy of Science, 47,* 188–202.

Strack, F. (1992). Sensation- and information-based social judgments: Toward a general model. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Sue, S., Smith, R. E., & Caldwell, C. (1973). Effects of inadmissible evidence on the decisions of simulated jurors: A moral dilemma. *Journal of Applied Social Psychology, 3,* 345–353.

Sue, S., Smith, R. E., & Pedroza, G. (1975). Authoritarianism, pretrial publicity, and awareness of bias in simulated jurors. *Psychological Reports, 37,* 1299–1302.

Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 457–477.

Swann, W. B., Jr. (1990). To be adored or to be known? The interplay of self-enhancement and self-verification. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 408–448). New York: Guilford Press.

Tanford, S., & Penrod, S. (1982). Biases in trials involving defendants charged with multiple offenses. *Journal of Applied Social Psychology, 12,* 453–480.

Tanford, S., & Penrod, S. (1984). Social inference processes in juror judgments of multiple offense trials. *Journal of Personality and Social Psychology, 47,* 749–765.

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48,* 227–236.

Thompson, W. C., Fong, G. T., & Rosenhan, D. L. (1981). Inadmissible evidence and juror verdicts. *Journal of Personality and Social Psychology, 40,* 453–463.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Uleman, J. S. (1989). A framework for thinking intentionally about unintended thought. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 425–449). Hillsdale, NJ: Erlbaum.

Wegner, D. M. (1989). *White bears and other unwanted thoughts.* New York: Viking Press.

Wegner, D. M. (1992). You can't always think what you want: Problems in the suppression of unwanted thoughts. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 193–225). San Diego, CA: Academic Press.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101,* 34–52.

Wegner, D. M., Coulton, G. F., & Wenzlaff, R. (1985). The transparency of denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology, 49,* 338–346.

Wegner, D. M., & Erber, R. (1993). Social foundations of mental control. In D. M. Wegner & J. W. Pennebaker (Eds.), *The handbook of mental control* (pp. 36–56). Englewood Cliffs, NJ: Prentice Hall.

Wegner, D. M., Erber, R., & Bowman, R. E. (1993). *On trying not to be sexist.* Unpublished manuscript, University of Virginia, Charlottesville.

Wegner, D. M., Klein, S. B., & Dimitri, S. (1992, August). *From here to uncertainty: Selective exposure to misleading information.* Paper presented at the 100th Annual Convention of the American Psychological Association, Washington, DC.

Wegner, D. M., & Pennebaker, J. W. (Eds.). (1993). *The handbook of mental control.* Englewood Cliffs, NJ: Prentice Hall.

Wegner, D. M., Wenzlaff, R., Kerker, R. M., & Beattie, A. E. (1981). Incrimination through innuendo: Can media questions become public answers? *Journal of Personality and Social Psychology, 40,* 822–832.

Weinberg, H. I., & Baron, R. S. (1982). The discredible eyewitness. *Personality and Social Psychology Bulletin, 8,* 60–67.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820.

Weinstein, N. D. (1982). Unrealistic optimism about susceptibility to health problems. *Journal of Behavioral Medicine, 5,* 441–460.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43,* 337–375.

Wertheimer, M. (1959). *Productive thinking* (rev. ed.). New York: Harper & Row.

Wetzel, C. G., Wilson, T. D., & Kort, J. (1981). The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology, 17,* 427–439.

Whitley, B. E. (1987). The effects of discredited eyewitness testimony: A meta-analysis. *Journal of Social Psychology, 127,* 209–214.

Wicklund, R. A., Reuter, T., & Schiffmann, R. (1988). Acting on ideas: Appropriation to one's self. *Basic and Applied Social Psychology, 9,* 13–31.

Wilson, T. D. (1985). Strangers to ourselves: The origins and accuracy of beliefs about one's own mental states. In J. H. Harvey & G. Weary (Eds.), *Attribution in contemporary psychology* (pp. 9–36). San Diego, CA: Academic Press.

Wilson, T. D., Brekke, N., Etling, K., & Houston, C. (1992). [People's beliefs about and preferences for biasing information]. Unpublished raw data.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). San Diego, CA: Academic Press.

Wilson, T. D., Etling, K., & Houston, C. (1993). [The consequences of erroneous beliefs about unwanted influences]. Unpublished raw data.

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Hillsdale, NJ: Erlbaum.

Wilson, T. D., Houston, C., Etling, K., & Brekke, N. (1993). *Origins and limits of anchoring effects.* Unpublished manuscript, University of Virginia, Charlottesville.

Wilson, T. D., & Stone, J. I. (1985). Limitations of self-knowledge: More on telling more than we can know. In P. Shaver (Ed.), *Review of personality and social psychology* (Vol. 6, pp. 167–183). Beverly Hills, CA: Sage.

Wolf, S., & Montgomery, D. A. (1977). Effects of inadmissible evidence and level of judicial admonishment on the judgments of mock jurors. *Journal of Applied Social Psychology, 7,* 205–219.

Wrightsman, L. S. (1991). *Psychology and the legal system* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Wyer, R. S., Jr., & Budesheim, T. L. (1987). Person memory and judgments: The impact of information that one is told to disregard. *Journal of Personality and Social Psychology, 53,* 14–29.

Wyer, R. S., Jr., & Srull, T. K. (1989). *Memory and cognition in its social context.* Hillsdale, NJ: Erlbaum.

Wyer, R. S., Jr., & Unverzagt, W. H. (1985). The effects of instructions to disregard information on its subsequent recall and use in making judgments. *Journal of Personality and Social Psychology, 48,* 533–549.

Zanot, E. J., Pincus, J. D., & Lamp, E. J. (1983). Public perceptions of subliminal advertising. *Journal of Advertising, 12,* 39–45.

Zillmann, D. (1978). Attribution and misattribution of excitatory reactions. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2, pp. 335–368). Hillsdale, NJ: Erlbaum.

(Appendixes follow on next page)

Appendix A

Failure of Rule Knowledge and Application

Definition

The failure to know or apply a useful rule or algorithm relevant to human reasoning. These rules, when known, can be applied consciously and deliberately.

Examples

1. The law of large numbers (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991)
2. Regression to the mean (Nisbett, Krantz, Jepson, & Kunda, 1983)
3. Cost–benefit rules (e.g., sunk costs; Larrick, Morgan, & Nisbett, 1990)
4. Causal and contractual rules (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986)
5. Bayes's theorem (Lopes, 1987)

Appendix B

Different Types of Mental Contamination

Unwanted Consequences of Automatic Processing

1. Categorization (e.g., Devine, 1989a): unwanted effects of automatically categorizing the social environment (e.g., prejudice and stereotyping)
2. Accessibility (e.g., Higgins, 1989; Wyer & Srull, 1989): unwanted effects of information that is accessible because of recent or frequent exposure (e.g., instances in which arbitrary information primes a social category, thereby influencing one's impression of another person)
3. Initial acceptance of propositions (Gilbert, 1991): unwanted effects of the tendency to initially believe propositions (e.g., when people are cognitively taxed, they have difficulty "unaccepting" false propositions)
4. Ironic effects of thought suppression (Wegner, 1992, 1994): unwanted effects of ironic monitoring processes (e.g., when people are cognitively taxed, the very thought they are trying to suppress becomes hyperaccessible)

Source Confusion

1. Misattribution of arousal/emotion (confusing emotion and arousal; Schachter & Singer, 1962; Zillmann, 1978)
2. Implicit memory effects (e.g., confusing familiarity and fame; Jacoby, Woloshyn, & Kelley, 1989)
3. Mere exposure (confusing familiarity and liking; Bornstein & D'Agostino, 1990)
4. Halo effects (confusing liking and objective judgments; Nisbett & Wilson, 1977a)
5. Effects of mood on judgment (not recognizing the contribution of mood to one's judgments; Schwarz & Clore, 1983)
6. False consensus (not recognizing the contribution of one's own construal of a situation to estimates of population frequency; Ross, Greene, & House, 1977)
7. Effects of media denials (not recognizing contribution of headline to one's negative impressions of a person; Wegner, Wenzlaff, Kerker, & Beattie, 1981)
8. Effects of extraneous events (noise) on boredom (Damrad-Frye & Laird, 1989)
9. Effects of cues on spontaneous number generation (Kubovy, 1977)
10. Appropriation of others' ideas (Wicklund, Reuter, & Schiffmann, 1988)
11. "Curse of knowledge" effect (inability to discount privileged information; Keysar, Ginzel, & Bazerman, 1992)
12. Persuasive messages (advertising: believing that one's own attitudes are not due to persuasive messages; sleeper effects; Pratkanis, Greenwald, Leippe, & Baumgardner, 1988)
13. Leading question effects (confusion about source of memory; Loftus, 1979)

Received October 14, 1992
Revision received November 23, 1993
Accepted December 2, 1993 ■