

mistakes, but given the complexity of the problem, it is rather remarkable that we ever get it right. How do we perform these analyses so easily, so often, and (at least sometimes) so well? Social psychologists have been asking this question for more than half a century and the accumulated wisdom on the topic is substantial (see Gilbert, 1998). One small piece of the answer – the small piece reviewed here – is that we readily believe the claims that actions make about actors, and only later, with time and attention, consider the possibility that these claims were unfounded. There is still much to be learned about this small piece, and even more about the larger question to which it is a partial reply. But if we can chart the stellar universe by attending carefully to cold pinpricks of distant light, then perhaps with just a bit more time and effort we can map the inner world of the social perceiver as well.

## 10. Mental Contamination and the Debiasing Problem

*Timothy D. Wilson, David B. Centerbar, and Nancy Brekke*

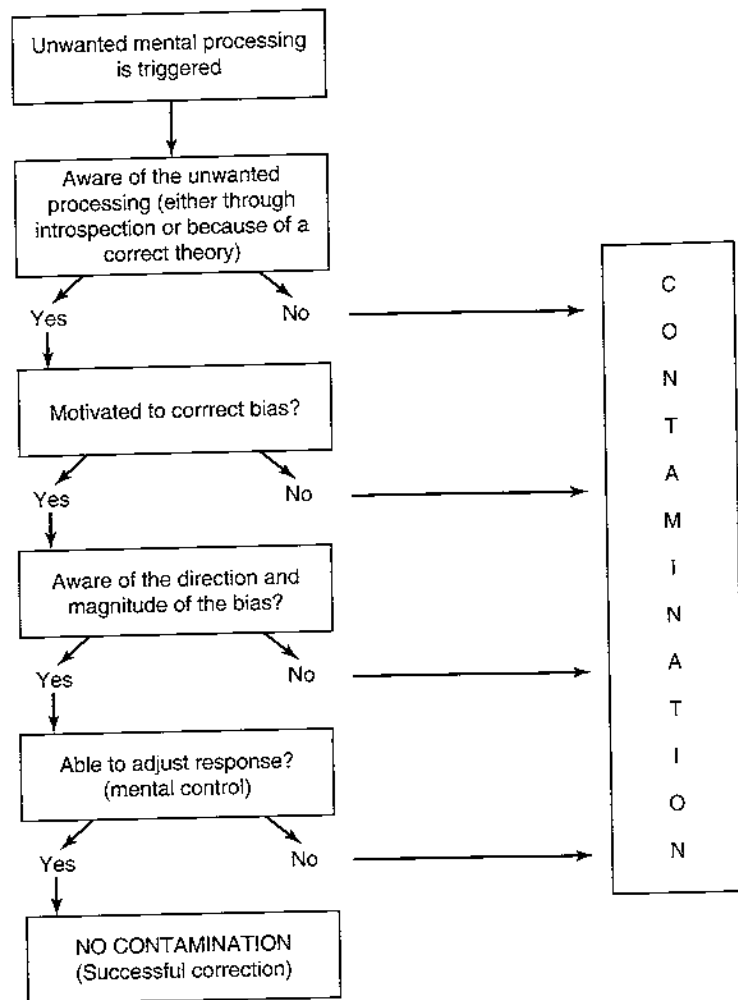
People do not have access to, or control over the recipe of their impressions, feelings, and attitudes. There exists a vast, adaptive unconscious that does much of the work of sizing up and evaluating the world (Nisbett & Wilson, 1977b; Wilson, 2001); people often know the final product with no awareness of exactly how it was formed. Consider Ms. Green, a partner in a prestigious law firm, who is interviewing candidates for the position of associate in her firm. When she interviews Mr. Jones, a young African-American attorney, she has an immediate negative impression, finding him to be arrogant and lacking the kind of brilliance she looks for in new associates. Ms. Green decides that her impression of Mr. Jones was accurate and at a meeting of the partners, argues against hiring him. She wonders, however, whether her negative evaluation was influenced by Mr. Jones' race.

Ms. Green may have suffered from mental contamination, which Wilson and Brekke (1994) defined as unconscious or uncontrollable mental processing that results in unwanted judgments, emotions, or behavior. Her dilemma illustrates the difficulty of knowing whether one's responses are contaminated, and if so, how to correct them. In this chapter, we review and update the arguments made by Wilson and Brekke (1994) concerning mental contamination, and also address several questions about the ways in which people try to protect their minds from unwanted influences.

### THE WILSON AND BREKKE MODEL

The Wilson and Brekke model of contamination is illustrated in Fig. 10.1. Contamination begins when an unwanted mental process is triggered, such as a negative evaluation of a person based on his or her race. There are many other examples of unwanted influences on people's judgments and emotions, such as the effects of advertising (e.g., buying a product based on exaggerated claims about its value), halo effects (e.g., a professor giving a student a bad grade because he/she does not like the student), and misattribution (e.g., confusing romantic attraction and fear).

In order to avoid contamination, people must first detect that it exists. This is often quite difficult, because people have poor access to the processes by which they form their judgments (Nisbett & Wilson, 1977b). Ms. Green could



**Figure 10.1.** The process of mental contamination and debiasing (from Wilson & Brekke, 1984).

not observe directly whether her negative evaluation of Mr. Jones was tainted by prejudice. Unlike many examples of physical contamination, such as spoiled milk, there are no overt signs that a final judgment is tainted. In the words of Wilson and Brekke (1994), "Human judgments – even very bad ones – do not smell" (p. 121).

There may sometimes be indirect clues that a judgment is contaminated. For example, an unexpected reaction to a stimulus (e.g., disliking a highly touted job candidate) might trigger thoughts about bias (e.g., "Maybe I'm racist"). Furthermore, a potential contaminant might be so salient that people cannot fail to wonder if it biased their judgment (Stapel, Martin, & Schwarz, 1998). More commonly, we have no direct way of knowing whether our judgment is biased.

As noted by Wilson and Brekke (1994), Petty and Wegener (1993), and Wegener and Petty (1997), people often are at the mercy of their lay theories about bias. An unexpected reaction, for example, is likely to be attributed to bias only if people have an applicable theory about the cause of the bias, such as racism. If the job candidate were a white male from Des Moines, it is unlikely that white male supervisors would conclude that they were prejudiced against white males or Iowans, even if their evaluation were unexpectedly negative.

Once people decide that mental contamination has occurred, how easily can they correct for it? As seen in Fig. 10.1, several conditions must be satisfied. First, people must be motivated to correct the error. Ms. Green may know that her impression was biased and prefer that it not be biased, but believe that it is not worth the time and energy to try to correct it. Second, even if motivated to correct the error, people must be aware of the direction and magnitude of the bias. Ms. Green may know that her impression of Mr. Jones is unfairly negative and want to avoid this bias, but have no idea of the extent of the bias. Should she change her evaluation from "Should not be hired" to "Barely acceptable" or to "Best applicant I've seen in years"? Finally, people must have sufficient control over their responses to be able to correct the bias. Ms. Green might strongly suspect that biased processing has led to a lower impression of Mr. Jones, but find it very hard to disregard this impression at the partners' meeting.

The main causes of mental contamination, then, are (1) a lack of awareness of one's mental processes (e.g., the extent to which a negative evaluation of a job candidate was due to race); (2) a lack of control over one's mental processes (e.g., an inability to prevent a person's race from influencing an impression of him or her); (3) inaccurate theories about unwanted influences on judgments (e.g., a failure to appreciate how a person's race can influence impressions nonconsciously); and (4) inadequate motivation to correct for contamination (e.g., an insufficient desire to avoid a racist response; see Fig. 10.1). Numerous biases can be viewed as examples of mental contamination, as long as the individual would prefer not to exhibit the bias. Examples include automatic stereotyping and prejudice; unwanted effects of accessible information, whereby a judgment is influenced by irrelevant information that happens to have been primed; unwanted effects of persuasive communications, such as advertising; and misattribution, whereby a person's emotions and judgments are based on erroneous inferences about the causes of their arousal (see Wilson & Brekke, 1994, for other examples).

In this chapter, we discuss the consequences of unwanted judgments and ways they might be avoided, with an emphasis on issues that have arisen since Wilson and Brekke's (1994) article was published.

#### **ARE ALL CORRECTION EFFORTS DELIBERATIVE AND THEORY-DRIVEN? IMPLICIT ADJUSTMENT VERSUS DEBIASING**

There has been some controversy over the manner in which people attempt to correct their psychological responses. Some theorists, such as Wilson and

Brekke (1994) and Wegener and Petty (1997), stress the importance of deliberative applications of lay theories. Wegener and Petty's (1997) flexible correction model, for example, argues that correction is driven by the conscious belief that bias has occurred and a naive theory about the direction and magnitude of the bias. Other theorists, such as Martin (1986) and Martin and Stapel (1998), argue that correction can also be a spontaneous and automatic response to the nature of the situation and stimulus.

We suggest that both camps in this debate are correct because there are different levels and types of correction. We agree with Martin and Stapel (1998), who state that a great deal of mental processing occurs at an automatic and implicit level, and that some of this processing involves a correction for information deemed inappropriate or irrelevant to the judgment at hand. In fact, a number of such automatic correction processes have been documented in the social cognition literature, such as research on self-attribution, in which people observe their behavior and, under some circumstances, infer that this behavior reflects their internal states (e.g., Bem, 1972). The attribution about the cause of one's behavior or physiological responses sometimes involves a correction process, whereby people discount the role of one potential cause if another plausible one is present. These corrections occur automatically and nonconsciously.

Consider a classic misattribution study by Dutton and Aron (1974), in which male participants encountered an attractive woman while crossing a flimsy footbridge spanning a 200-ft gorge. The participants had to decide how much of their arousal was due to fear from the bridge versus attraction to the woman. This decision can be characterized as a correction process; people had to partial out the effects of one plausible cause from another. As it happened, people who encountered the woman on the bridge misattributed some of their arousal to attraction to the woman.

It seems clear that this correction process did not occur consciously and deliberately. The men did not stand on the bridge scratching their heads thinking, "Hmm, is she beautiful or am I scared? Or a little of both? I'd say I'm feeling 37% fear and 63% attraction." Rather, their phenomenal experience was attraction to the woman with no conscious recognition of the processes that produced it. Although the issue of how aware people are of attribution processes was seldom discussed in the original literature on this topic, Nisbett and Wilson (1977b) later documented that participants in these studies rarely are able to verbalize the attribution processes hypothesized to have occurred.

The broader point is that mental correction occurs on a continuum of implicit and nonconscious adjustments of a response to explicit and deliberative adjustment, and researchers should keep in mind that when they use the term *correction*, they are sometimes referring to quite different mental processes. To help clear up this confusion, researchers should use different terms for the different ends of this continuum. We suggest that *implicit adjustment* be used to

refer to rapid and nonconscious correction discussed by attribution theorists and Martin (1986), and *debiasing* (Fischhoff, 1982) be used to refer to deliberative and theory-driven correction discussed by Wilson and Brekke (1994) and Wegener and Petty (1997). In the interest of promoting these terms, we use them as defined here in the remainder of this chapter.

As noted by Wilson and Brekke (1994), the very fact that implicit adjustment is so common is a major source of bias. They discussed two subcategories of contamination: (1) unwanted consequences of automatic processing, whereby people process information automatically in undesired ways; and (2) source confusion, whereby people confuse two or more sources of a judgment, memory, or feeling. These types of contamination can involve implicit adjustment, such as the kind of misattribution errors just discussed.

The existence of implicit adjustment puts people in a difficult position to know whether a particular response is biased because they do not know whether a phenomenal judgment has already been "corrected." Suppose that Michael finds Donald to be slightly aggressive. How accurate is this impression? Was it biased by irrelevant information, such as the fact that Michael had just read a newspaper article about spousal abuse, which primed the construct of aggression? There are three possibilities:

1. The newspaper article did not prime aggression, thus Michael's impression of Donald was unbiased.
2. The newspaper article primed the construct of aggression and produced assimilation, such that Michael finds Donald to be *more* aggressive than he would have if he had not read the newspaper article.
3. The newspaper article primed the construct of aggression, but Michael already corrected for this fact nonconsciously. This resulted in a contrast effect, such that Michael finds Donald to be *less* aggressive than he would have if he had not read the newspaper article.

Obviously, the extent to which Michael attempts to consciously adjust his impression depends on which of these processes he believes has occurred. Because he has no direct access to this process, he is at the mercy of his naive theories, which might well be wrong.

#### WHAT TRIGGERS DELIBERATIVE DEBIASING EFFORTS?

The previous discussion indicates that there are two stages of the debiasing process: The decision that bias has occurred and attempts to correct it. Naive theories can come into play at both stages. Because people do not have good access to their cognitive processes, they must rely on their theories when deciding whether a judgment is biased. Once this decision is made, people must again rely on their theories to inform them how to correct for the bias.

As noted by Martin and Stapel (1998), researchers have paid more attention to the second stage of the process (debiasing) than to the initial, bias-detection

stage. In most of the research in this area, people are specifically instructed to avoid bias, which begs the question of when and how they invoke debiasing processes on their own (for exceptions, see Petty & Wegener, 1993; Stapel, Martin, & Schwarz, 1998).

How likely are people, in the absence of blatant warnings, to invoke a theory that their judgments are biased? Although this question is difficult to answer, our hunch is similar to Martin and Stapel's (1998) that people's default response is to assume that their judgments are uncontaminated. One source of evidence for this conclusion comes from studies showing that people are more willing to attribute bias to other people's judgments than to their own (Kruger & Gilovich, 1999; Pronin, Puccio, & Ross, ch. 36, this volume; Wilson & Brekke, 1994; Wilson, Gilbert, & Wheatley, 1998).

The reason for people's faith in their own judgments can be traced to both motivational and cognitive factors. On the motivational side, it can be disquieting to conclude that one's own judgments are biased, and functional to overestimate the validity of these judgments (Taylor & Brown, 1988). On the cognitive side, the immediacy and inescapability of a phenomenal judgment probably contributes to its perceived validity. The phenomenal experience of one's own judgment has a compellingness that is absent from the knowledge that someone else feels differently, which probably contributes to the greater faith in one's own judgment.

Although the question of how often people detect bias and try to correct for it in everyday life is difficult to answer, we are not optimistic. If there were any group of people who would be particularly sensitive to mental contamination, one would think it would be research psychologists familiar with the extensive literature on biases in social cognition and the difficulty of avoiding these biases. We offer two anecdotes suggesting that psychologists are no more likely than others to take steps to avoid unwanted influences. The first is a colleague who said that when he grades his students' papers and exams, he does so with full knowledge of the students' identity and past performance, even though he is aware of studies showing that halo effects can color one's impressions of papers and exams. "I don't want to be biased by how much I like the students or how well they did on a previous test," he said, "so I simply keep a close watch on myself to make sure this doesn't happen." Although this colleague is aware of research showing that people do not have direct access to how much they are biased by halo effects (Nisbett & Wilson, 1977a), he believes that he is an exception.

The second anecdote concerns the validity of the interview in predicting job performance of a job candidate. Despite evidence for the extremely low validity of the interview (e.g., Hunter & Hunter, 1984), the members of most departments of psychology are as confident in their impressions of a candidate after meeting him or her for half an hour as anyone else would be, and are as willing to let this impression override information known to be more valid (e.g., letters of recommendation, publication record). We confess that we often feel the same way. A strong impression of a person is so compelling and *odorless*, in

Wilson and Brekke's terms, that it is difficult to acknowledge that it might be tainted by unwanted influences, despite the fact that we are quite willing to see other people's impressions as rather smelly.

### WHEN PEOPLE TRY TO CORRECT, HOW SUCCESSFUL ARE THEY?

Even in the rare instances in which people believe that their judgments are biased, they may not successfully debias these judgments. In fact, their corrected judgments might be worse than their uncorrected ones.

The dilemma people face is similar to the predicament of a hiker who arrives at the intersection of several trails. There is a sign pointing toward one of the paths indicating that it leads to the hiker's destination. However, the hiker has been told that a mischievous boy often moves the sign so that it points in the wrong direction, and she has just seen a boy running away from the sign, laughing gleefully. She is pretty sure that the sign is pointing to the wrong path (i.e., that it is "contaminated"), but how should she correct for this fact? Given that the sign is pointing to the trail on the far left, should she take the one on the far right? Or should she assume that the boy only had time to move the sign a little, and therefore take the middle trail? Or that he is a clever boy who tried to convince her that he moved the sign, when in fact he did not? Just as people are at the mercy of their theories when deciding whether a response is biased, so are they at the mercy of their theories when deciding how to correct for this bias.

Unfortunately, people's theories about debiasing do not appear to be particularly accurate. People who are exposed to contaminating information and engage in debiasing rarely end up with judgments similar to people who were not exposed to the contaminant. Three kinds of errors have been found: *insufficient correction* (debiasing in the direction of accuracy that does not go far enough), *unnecessary correction* (debiasing when there was no bias to start with), and *overcorrection* (too much debiasing, such that judgments end up biased in the opposite direction).

A number of studies found evidence for insufficient correction, which is the best of the three types of errors (because the corrected judgment is more accurate than the uncorrected one). Petty and Wegener (1993, Study 4), for example, found that rating the desirability of vacations in very appealing locations, such as Hawaii, lowered people's ratings of vacations in American cities such as Indianapolis and Kansas City. Giving people a subtle cue that their initial ratings might bias their later ones led to some, but not complete, debiasing. People in the subtle cue condition formed judgments that were still biased, but not as biased as people who did not receive the subtle cue. These results suggest that some debiasing is better than no debiasing.

However, this same study found evidence for unnecessary correction in another condition. Some participants received a more blatant cue that their ratings of the American cities might be biased. People in the control condition, who did not first rate vacations in Hawaii but received the blatant warning, engaged in

unnecessary correction. That is, they seem to have assumed that their evaluations of the American cities were biased when in fact they were not. They became more negatively disposed toward vacations in these cities, even though their evaluations had not been biased in the first place.

Stapel, Martin, and Schwarz (1998) found evidence for overcorrection. Similar to the Petty and Wegener (1993) study, people first rated the desirability of the weather in tropical locations and then rated the desirability of the weather in midwestern cities. In the absence of any warnings, this led to a contrast effect: Compared to a control group who did not first rate the tropical locations, people rated the weather in the midwestern cities as less desirable. When people were warned to avoid unwanted influences on their ratings, they overcorrected such that they showed an assimilation effect: Compared to the control group, they rated the weather in the midwestern cities as more desirable.

In sum, just because people attempt to correct a judgment they perceive to be biased is no guarantee that their result will be a more accurate judgment. Either unnecessary correction or overcorrection might occur, resulting in judgments that are more biased than they were before they were "corrected."

### A CLOSER LOOK AT LAY THEORIES OF MENTAL PROTECTION

Our discussion thus far has focused on biased judgments, such as beliefs about the weather in various locations or the merits of a job candidate. Wilson and Brekke (1994) defined *mental contamination* quite broadly, including cases in which any response is biased in an unwanted way (i.e., emotions as well as beliefs).

Wilson, Gilbert, and Wheatley (1998) examined the manner in which people manage emotions versus beliefs. They also took a closer look at the specific strategies people believe they can take to avoid unwanted influences on these responses. As seen in Fig. 10.2, there are five strategies people believe they can adopt at different points of the contamination process. The left side of this figure shows a time line of actual contamination, from people's anticipation that a state might become contaminated, to exposure to a contaminant, to unwanted changes in a psychological state. The right side of the figure shows the different strategies people think they can adopt at each time point to avoid or undo the contamination.

The first line of defense is what Gilbert (1993) termed *exposure control*, which is the decision whether to allow a potentially contaminating stimulus to enter our minds. If we are concerned that our liking for a student will bias the grade we give his or her paper, we can grade the papers blindly to avoid any bias. Exposure control, to the extent that it is feasible, is the most effective defense against contamination. A stimulus that never enters our minds cannot bias our judgments or feelings.

The second line of defense is *preparation*. If we know in advance that we will be exposed to potentially contaminating information, we can take steps in advance

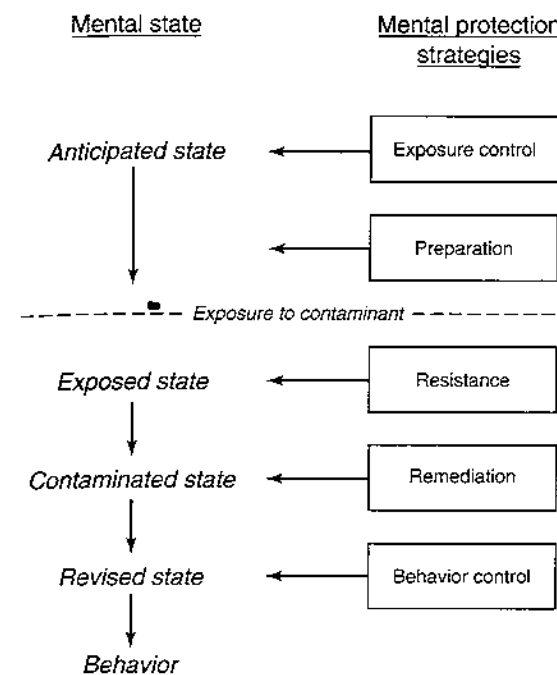


Figure 10.2. Lay beliefs about debiasing (from Wilson et al., 1998).

to blunt its impact. A person about to hear a speech from an untrustworthy source, for example, can try to strengthen her mental defenses by engaging in anticipatory counterarguing.

Once we are exposed to a contaminant – we hear the speech or know the author of the term paper we are grading – there are three steps people can take to try to avoid or undo contamination. The first of these steps is *resistance*, which occurs after a stimulus is encoded but before it has had an unwanted effect on our beliefs or emotions, resulting in what Wilson et al. (1998) called an *exposed state*. Resistance involves any mental operation that attempts to prevent an encoded stimulus from having an adverse effect, similar to an immunologic response that kills a virus after it enters the body but before it causes disease.

If resistance fails, people end up with a contaminated state. The next line of defense is *remediation*, defined as any mental operation that attempts to undo the damage done by a contaminant. People end up with a revised state, which, if remediation is completely successful, is the same as the original mental state. If remediation fails or is insufficient to correct the damage, then people's last line of defense is *behavior control*. This is the attempt to prevent one's contaminated state from influencing behavior. If our beliefs have changed in an unwanted way (e.g., a lowered evaluation of a job candidate due to his or her race), we can try to stop ourselves from acting on that belief.

We should stress that these stages are meant to capture people's beliefs about how to avoid or undo bias and are not necessarily a portrayal of effective strategies. In terms of effectiveness, it is clear that the earlier the defense, the better. The most effective defense is preventing exposure to contaminating information, and the least effective is trying to undo or ignore contamination once it has occurred.

People seem to appreciate this point when it comes to managing their affect and emotions. Much of our lives is spent arranging environments in ways that maximize our pleasure, with the recognition that once we are exposed to negative stimuli, the game is mostly lost. Given the choice of encountering an unpleasant and argumentative colleague who always manages to spoil our mood, or taking a longer route to our office that avoids the colleague, most of us choose the latter option. Just as it is better to avoid the flu by avoiding exposure to the virus, so is it better to maintain a good mood by avoiding contact with stimuli that will ruin it – a fact that people generally appreciate.

When it comes to avoiding contaminated beliefs, however, people seem to have more faith in the later defenses of resistance and remediation and are thus less likely to engage in exposure control. People seem to believe that there is little danger in being exposed to information that might bias their beliefs because they think they have the ability to resist or remediate any unwanted influences. As noted by Gilbert (1991), people assume that belief formation occurs according to a process outlined by Descartes: First we comprehend a proposition (e.g., "the moon is made of blue cheese"), then we freely decide whether to accept it as true (e.g., whether it fits with our belief system). Gilbert (1991) argues persuasively that belief formation actually operates according to a process described by Spinoza, whereby we initially accept as true every proposition we comprehend, and then decide whether to "unbelieve" it.

One implication of Gilbert's Spinozan model is that people have a misplaced belief in resistance (see Fig. 10.2). To be able to resist unwanted influences on our beliefs before they change our minds, we would have to be able to encode a message without it influencing us, placing it in a kind of mental holding pattern. We would then have to neutralize the message in some way, such as by thinking of counterarguments. There is no such thing as a mental holding pattern, Gilbert argues, because people initially believe all propositions. According to this view, people cannot encode something without believing it; thus, mental resistance is impossible.

People's faith in remediation is also misplaced, we suggest, because of the difficulty of recognizing that contamination has taken place. Even if we correctly guess that it has, we have to know how to correct for the bias. As we have seen, people often get it wrong, either failing to detect bias or failing to correct for it. As also seen, people appreciate that other people are not very good at avoiding biased influences on their beliefs, but have a misplaced faith in their own ability to control their beliefs and avoid unwanted influences.

Why are people's strategies for managing their beliefs less effective than their strategies for managing their emotions? One reason, we suggest, is the ease of detecting contamination in these different realms. As discussed earlier, there are seldom any phenomenological signs that a belief is biased in an unwanted way – contaminated judgments do not smell. It is much easier to tell whether our emotions are "contaminated" – we feel lousy. Consequently, we have developed better strategies for avoiding contaminated emotions, namely exposure control (avoiding exposure to possibly contaminating information).

## RECOMMENDATIONS

The lesson from our discussion so far should be clear: The best way to avoid biased judgments and emotions is exposure control, whereby we avoid stimuli that might influence our responses in unwanted ways. Although exposure control is the most effective prophylactic, Wilson and Brekke (1994) noted four problems with this strategy. First, we cannot control exposure to all possible contaminating information. When deciding which employees should be promoted, for example, we already know their gender, age, and race. Second, we do not always know in advance whether information will bias our judgments; therefore, excessive exposure control will result in the failure to receive information that is diagnostic and useful. Third, the excessive use of an exposure control strategy might cause people to examine only information that confirms their views, thereby fostering narrow-mindedness. Fourth, in order to avoid exposure to biasing information, people would have to be extremely vigilant, ready to shut their eyes and cover their ears whenever they suspected that potentially unwanted information was in the vicinity. Is such vigilance worth the effort it entails? 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 – people should not be in a state of constant nervousness, ever alert for contaminating information. However, we believe that mental errors can be "awfully solemn things" that are quite costly to ourselves and others. Some personal decisions, such as which brand of laundry detergent to buy or what to order in a restaurant, are relatively inconsequential. Other choices have much bigger consequences, such as whether to buy the \$100,000 house on a busy street or the \$150,000 house on the cul-de-sac, or whether to seek medical treatment for a mole on our forearm. For consequential decisions such as these, a dose of "excessive nervousness" might not be such a bad thing. Moreover, because our judgments often influence the lives of others, the costs of mental contamination can extend well beyond the personal realm. When we are in positions of power, such as professors assigning grades to students, or employers screening job applicants, we should be especially humble about the potential for bias.

Furthermore, the failure to appreciate the perils of mental contamination may lead people to design decision-making systems that are destined to produce biased judgments. The American legal system, for example, has overestimated people's ability to detect and correct for unwanted biases, such as jurors' ability to discount inadmissible evidence (Tanford, 1990). As a result, legal procedures often rely too much on defenses such as remediation and too little on preemptive strategies such as exposure control. At an institutional or societal level, the cumulative negative effects of such contaminated systems can be quite substantial.

### Strategies for Dealing with Mental Contamination in Everyday Life

Exposure control is sometimes possible and can be considered to be a preventive strategy. As professors, we might choose to grade student papers without knowing the identity of each student so as to avoid the halo effect. Similar strategies are occasionally feasible in the realm of employment. When many professional orchestras audition new musicians, they ask the musician to play behind a screen, so that the decision maker is unaware of the person's gender, age, race, and any other potentially biasing information.

What if exposure control is not possible or is not desirable? Although we remain pessimistic about people's abilities to detect and correct for contamination once it enters their minds, we offer three suggestions for strategies that can help.

*Examining the Test-Retest Reliability of One's Judgments.* When it comes to debiasing a judgment of a specific stimulus, it can be useful to observe one's reactions over time under different circumstances – essentially, to observe the test-retest reliability of the evaluation. Doing so may allow one to tease apart whether the evaluation is a reliable reaction to the stimulus or a spurious, contaminated reaction to something about the circumstances under which it is evaluated.

Consider an example: One of us adopts such an approach when grading students' final papers. Because this activity takes place over the course of several days, she knows that her judgments might be affected by order and contrast effects as well as various situational factors (e.g., whether she reads a paper at the office between appointments or at home in front of the fireplace with a glass of wine in hand and classical music in the background). Accordingly, after reading each paper, she writes a tentative grade and some comments on a separate sheet. Then, after reading all of the papers, she looks at each paper again and gives it another grade. If the two grades she assigns do not match, she suspects contamination and gives the paper further scrutiny (e.g., she compares it to other papers in the same grade ranges).

Such a strategy is not perfect, of course. Although it can help us detect contamination, it does not tell us how to correct for it. The professor who determines that her evaluations were biased by the setting must still decide whether the paper grades were artificially inflated by the fireplace, music, and wine, or whether

they were artificially deflated by the hectic, noisy office. Moreover, the test-retest strategy does not allow people to tease apart contaminants that are present at both times. For example, we have a friend who has a closet full of clothes she never wears that she ordered from a catalog. Each time she looks at the catalog she loves the way the clothes look, but she underestimates how much this is due to the fact that beautiful models are wearing them. When the clothes arrive, she is invariably disappointed at how they look on her.

*Examining the Covariation of One's Responses with Potential Contaminants.* It can be useful to examine our general patterns of response to different subgroups of stimuli. A faculty colleague of ours, for instance, once noted a suspicious trend in his evaluations of student papers: Heavy smokers tended to receive lower grades than nonsmokers. As it turned out, the papers that reeked of stale cigarette smoke gave him a headache as he read them, which then caused him to evaluate the work more negatively. The test-retest strategy would not have helped him here (because smokers' papers reliably reeked and reliably elicited headaches), yet by comparing his evaluations of smokers' and nonsmokers' work, he was able to identify a contaminant.

This strategy can be useful in the workplace. Recall Ms. Green from the prestigious law firm. If she notices that, in general, her reactions to African-American candidates are less favorable than her reactions to European-American candidates, this should give her reason to pause. Indeed, this is precisely the logic behind some affirmative action guidelines. Employers are required to keep records of group-related employment patterns (e.g., the percentages of women and minorities who apply, are interviewed, are hired) so that group-linked disparities can be more readily detected (Crosby & Cordova, 1996). The existence of group disparities does not prove that discrimination has occurred, but it should prompt further investigation by the employer. In other words, even though individual contaminated judgments are unscented, certain patterns of judgment do smell "fishy," thereby alerting us to the possibility of unwanted bias.

Once again, we must point out that this strategy is limited. Perceivers are not terribly adept at covariation assessment (Nisbett & Ross, 1980), and therefore may fail to detect suspicious patterns. When they do succeed, the patterns may be open to multiple interpretations and the implications of global trends for individual cases may be unclear.

*Considering the Opposite.* Some kinds of mental biases can be undone by asking people to "consider the opposite," whereby they imagine different outcomes than the ones they had been thinking about. Simply thinking about whether a specific hypothesis is true, for example, increases people's beliefs in its validity because the person focuses more on explanations as to why it could be true than why it could be false. Several studies have found that asking people to consider the possibility that the opposite hypothesis is true is sufficient to undo the bias of one-sided thinking (Lord, Lepper, & Preston, 1984).

For this strategy to work, however, people must be able to mentally simulate alternative outcomes or hypotheses (Hirt & Markman, 1995). In many everyday

cases of contamination, people may be so locked into one way of looking at something that it is difficult for them to imagine alternatives. This is especially likely to be true, we suspect, when people have contaminated affective reactions. The affect may anchor their judgments to the extent that it is difficult for them to imagine they could feel otherwise. Suppose Ms. Green tried to imagine how she would have evaluated Mr. Jones if he were White. Her negative feelings might anchor her thinking to the point where it is difficult to perform this mental simulation.

Furthermore, people do not always know what aspect of a situation to mentally undo. Consider the men in the Dutton and Aron (1974) experiment who were approached by the woman on the scary bridge. Presumably, they had little idea that they were misattributing fear to attraction to the woman. Even if it occurred to them to consider the opposite, they might not know which aspect of the situation to transform. Should they imagine that they had met the woman at a different time of day or season of the year? Without some idea about the source of contamination, people do not know which mental simulation to run.

### **Legal Implications: Should People Be Held Responsible for Mental Contamination?**

Given the difficulty of avoiding many forms of mental contamination, is it reasonable to expect people to try to avoid mental contamination? If a person in power acts on a contaminated judgment and harms another person as a result, should he or she be held liable for that harm? An interesting discussion of this issue has arisen among legal scholars in the area of employment discrimination.

Consider again our example of the law partner, Ms. Green, and the decision-making processes leading to her recommendation against hiring the young African-American attorney, Mr. Jones. Let us assume, for the sake of argument, that Ms. Green's negative decision was biased by Mr. Jones's race. Let us further suppose that Ms. Green is a well-meaning person who would be aghast to know that race influenced her decision not to hire Mr. Jones. If Mr. Jones were to file a racial discrimination suit, could Ms. Green or her firm be held legally liable under current antidiscrimination laws?

Because Ms. Green's decision was affected by Mr. Jones's race, she seems to be in violation of the law. Title VII of the Civil Rights Act (1964, amended 1994) prohibits discrimination in employment on the basis of race, color, sex, national origin, and religion. Krieger (1995) points out, however, that Title VII is based on an antiquated view of prejudice in which all discrimination is assumed to be conscious, deliberate, and intentional. In essence, antidiscrimination law was written to prevent blatant, conscious racism, and does not take into account the possibility that discrimination can be unintentional and nonconscious. More generally, the law assumes that people are fully aware of the mental processes leading to their decisions; in Krieger's words, that they possess "transparency of mind" (p. 1185).

Thus, under current law, it would be very difficult for Mr. Jones to prevail in a lawsuit against Ms. Green. Mr. Jones would be required to prove not only that Ms. Green had taken his race into consideration, but also that Ms. Green was motivated by deliberate discriminatory intent. The jury would be forced to choose between two extreme positions: a verdict of "no discrimination," which implies that race played no role in Ms. Green's decision, or the opposite verdict, which brands Ms. Green as a conscious discriminator who lied in court about her real reasons for not hiring Mr. Jones (Krieger, 1995). Clearly, the law is ill-equipped to deal with the kinds of bias we discuss here, in which discrimination may be unintended and nonconscious.

Krieger (1995) makes a number of recommendations to redress this lack of fit. For example, she suggests that a clear distinction in interpreting motivation and intent be made, such that unintentional forms of contamination and its consequences may be considered as legal evidence. She also argues that a two-tiered liability system be adopted such that there would exist a first tier for equitable relief for unintentional discrimination, and a second, higher tier providing compensatory and punitive damages for intentionally motivated discrimination (similar to the criminal justice system, in which there is a distinction between manslaughter and first-degree murder).

Other legal scholars appreciate the fact that nonconscious discrimination can exist, but are less sanguine about ways in which the legal system can be changed to prevent it. Wax (1999) suggests that in many cases, the unintentional and unconscious forms of bias are sporadic and unpredictable and that the financial and social costs of attempting to remedy such forms of discrimination would be prohibitive. Much of her argument hinges on the apparent difficulty of detecting unconscious discrimination and our lack of knowledge concerning how to prevent or combat it. As she puts it, "That we cannot know another mind is a problem that plagues discrimination law generally. The dilemma is even more acute when the other mind cannot know itself nor effectively control itself nor be effectively controlled by others" (Wax, 1999, p. 1226).

Wax argues that attaching liability to acts that are outside the realm of conscious control will have little deterrent effect and may actually produce undesirable outcomes. For example, employers may invest scarce resources in activities that protect them from lawsuits but do little to reduce the actual incidence of workplace discrimination.

### **CONCLUSION**

This legal debate underscores the challenges of dealing with mental contamination in the real world and points to some important directions for future research. It is clear, for example, that techniques for detecting individual cases of mental contamination are sorely needed. In applied settings, the question of whether contamination has actually occurred is critical, and perceivers are unable to make this determination on their own. In this regard, it is interesting

that a number of researchers have devised techniques to measure an individual's level of implicit racial prejudice (e.g., Dovidio, Kawakami, Johnson, & Johnson, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998). Might these measures index an individual's propensity to engage in unconscious discrimination? Might they also be used some day in the legal arena?

Such a day is clearly not yet here because these techniques are in their infancy and their reliability, validity, and controllability are still an open question. Quite possibly these tests will have the same status as the polygraph, a tool of some interest to psychological researchers but of little use in courts of law. It is not inconceivable, however, that some day experts will administer implicit tests to a defendant to determine his or her level of implicit prejudice, just as clinicians now administer psychological tests to assess a defendant's mental health. Whether such a day will be a welcome one depends, of course, on the construct and predictive validity of the tests.

Even if valid implicit tests are developed, it may be that the nature of the human cognitive system precludes complete decontamination. If so, the best hope may lie in educating people about the causes of mental contamination so that they are able to (1) identify, after the fact, those cases in which contamination has most likely occurred; and (2) when possible, structure their decision-making environments so as to reduce the likelihood of contamination in the first place. This strategy has been met with some success in the domain of eyewitness identification (see Wells, 1993, for a review). Researchers have determined, for example, that some lineup procedures and configurations produce more mistaken identifications than others. By using this information when devising lineups, the police are able to reduce the number of "contaminated" identifications that occur. Of course, not all factors that influence the accuracy of eyewitness identification are amenable to control. The police cannot manipulate the quality of the viewing conditions under which the witness saw the perpetrator or whether the suspect looks familiar because the witness saw him earlier that day at the bus stop. In such cases, educating people – jurors, the police, attorneys, and perhaps even the witnesses themselves – about potential contaminants can help them determine the appropriate weight to give to an individual eyewitness' identification.

In sum, although we are pessimistic about people's natural ability to willfully control and correct their judgments, we are by no means suggesting that reducing mental contamination is a lost cause. Researchers are making promising advances in the detection of non-conscious biases and may ultimately devise some effective debiasing strategies. The challenges of eliminating contamination are great, but so may be the personal and societal costs of ignoring the problem.

## 11. Sympathetic Magical Thinking: The Contagion and Similarity "Heuristics"

Paul Rozin and Carol Nemeroff

### THE LAWS OF SYMPATHETIC MAGIC

The laws of sympathetic magic constitute a small but central subset of what might be called *magical thinking*. They share with the larger category that they promote beliefs about the world that are generally contrary to current scientific beliefs. However, unlike most other examples of magic, the laws of sympathetic magic do not necessarily invoke a sense of human or animate agency as a device to account for events in the world (see Nemeroff & Rozin, 2000; Tambiah, 1990, for a discussion of magical thinking in a broader context). In contrast to the larger category of magical thinking, the laws of sympathetic magic may be more tractable to experimental study for three reasons: (1) they are clearly defined; (2) they are present in abundance as modes of thought among contemporary people in developed societies; and (3) they invoke principles (e.g., contact, resemblance) that are easy to manipulate in the laboratory.

Edwin Tylor (1879), James Frazer (1895), and Marcel Mauss (1902) proposed three laws of sympathetic magic that they took to be universal principles of thinking. The *law of contagion* holds that "once in contact, always in contact"; when objects make physical contact, essences may be permanently transferred. Thus, fingernail parings contain the "essence" of the person to whom they were previously attached, and foods carry the "essence" of those who prepared them. The *law of similarity* holds either that like causes like (causes resemble their effects) or appearance equals reality. A prototypical example of similarity is the voodoo practice of burning a representation of an enemy to cause the enemy harm; action on the image is believed to result in effects on the object it represents. The *law of opposites*, which is not discussed further in this chapter, is the opposite of the first reading of the law of similarity, and holds that causes are the opposite of their effects.

On some important dimensions, contagion is the opposite of similarity. Similarity bears a relation to the principle of generalization, and surely is manifested in nonhuman animals; "appearance equals reality" is a very useful heuristic.

Preparation of this paper was assisted by NIDA grant R21-DA10858-0 to Paul Rozin, and grant #H8-038 from the Arizona Department of Health Services funded through the Centers for Disease Control and Prevention to Carol Nemeroff.

BF  
447  
H48  
2002

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK  
40 West 20th Street, New York, NY 10011-4211, USA  
477 Williamstown, Port Melbourne, VIC 3207, Australia  
Ruiz de Alarcón 13, 28014, Madrid, Spain  
Dock House, The Waterfront, Cape Town 8001, South Africa  
<http://www.cambridge.org>

© Cambridge University Press 2002

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2002

Printed in the United States of America

Typeface Palatino 9.75/12.5 pt. System L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> [TB]

*A catalog record for this book is available from the British Library.*

*Library of Congress Cataloging in Publication data*

Heuristics and biases : the psychology of intuitive judgment / edited by Thomas  
Gilovich, Dale Griffin, Daniel Kahneman.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-79260-6 – ISBN 0-521-79679-2 (pbk.)

1. Judgment. 2. Reasoning (Psychology) 3. Critical thinking. I. Gilovich,  
Thomas. II. Griffin, Dale III. Kahneman, Daniel, 1934–

BF447 .H48 2002

153.4 – dc21

2001037860

ISBN 0 521 79260 6 hardback

ISBN 0 521 79679 2 paperback

*To the memory of Amos Tversky*