18

The Cognitive Monster

THE CASE AGAINST THE CONTROLLABILITY OF AUTOMATIC STEREOTYPE EFFECTS

JOHN A. BARGH

This chapter provides a more or less opinionated history of the standard dual-process model of stereotyping effects on judgment and behavior. It focuses particularly on the fluctuations over the past 30 years in the relative power ascribed to the automatic influences of stereotypes versus the conscious, intentional attempts to control them. The major theme is that the evidence of controllability is weaker and more problematic than we would like to believe.

The Fable of the Cognitive Monster

Once upon a time, in the land of Social Psychology, there lived the Cognitive Miser. This creature was the object of much sympathy and compassion from the good people of Social Psychology, for it was afflicted with the curse of Limited Processing Resources, and therefore could do naught but give scant attention and time to most of the world around it. All agreed that it was necessary and wise for the creature to depend on simplifying modes of thought, in order to conserve its constrained mental capacity for when it was most needed. The people watched and noted how the Cognitive Miser never learned about them as separate individuals, but only reacted to each of them on the basis of their superficial aspects and the roles they played in daily life. It was unfortunate, but a reasonable strategy for the Miser to pursue, given its limits. And so the people in the land of Social Psychology were (reasonably) content.

But then the people became terribly afraid and anxious. For lo! the Cognitive Miser had become transformed, by the magic of Further Research, into the Cognitive Monster. No longer did the creature use simplifying categories and stereotypes by choice or strategy; their use had become an addiction—uncontrollable, not a matter of choice at all—and the creature's Will was powerless to do anything else.

"We must do something!" cried the people of Social Psychology. "We must slay the Monster!" And so their heroes came forth. They rode to the nearby friendly lands of Awareness and Motivation, and raised a formidable army. Then, cleverly using the very weapons of Further Research that had created the Monster, they turned them against the dreaded foe. The soldiers of Awareness shone their bright lights on the Monster, and thus aided, the people of Motivation lashed and tethered the beast. Victory followed upon victory. The Monster was greatly diminished in power and scope, and the people were no longer afraid. Indeed, soon came the day when they laughed and jeered at the Monster, tethered in chains in the village square.

Unfortunately, monster stories rarely end on such a happy note, as anyone familiar with the tales of King Kong or Frankenstein knows. The chained monster, it turns out in each of these stories, is only temporarily under control. And so, when the populace becomes complacent and lets its guard down, the monster bursts its bonds and rampages again, causing more mayhem than before.

Those classic monster stories were written as allegories for the very real monsters that individuals and societies face (be they dictators or unchecked scientific progress). They were also intended as warnings of the need for constant vigilance against such menaces. My theme in this chapter is similar. It is that in many ways, the field of social cognition has become overly optimistic about the "cognitive monster" of automatic stereotype activation. I contend here that, contrary to what our research is actually showing, the conclusions drawn from the data have overestimated the degree to which automatically activated stereotypes can be controlled through good intentions and effortful thought—and thereby have underestimated the extent to which stereotypes continue today to cause problems in social relations.

META-ASSUMPTIONS ABOUT STEREOTYPE CONTROLLABILITY: A BRIEF HISTORY

"The Fable of the Cognitive Monster" is intended to illustrate the assumptional shifts in social cognition that have taken place since the 1960s concerning the controllability (vs. automaticity) of social perception and judgment. The early attribution models broadly assumed an active and effortful search after meaning, with conscious and deliberate scrutiny of the co-occurrence of effects with their possible causes (e.g., Jones & Davis, 1965; Kelley, 1967; Weiner, 1974). Individuals were assumed to be largely in executive control of their perceptual and judgmental faculties.

The Cognitive Miser

However, the 1970s saw a reaction to this meta-assumption of effortful processing, as Langer (e.g., 1978; Langer & Abelson, 1972) and Taylor and Fiske (e.g., 1978) voiced doubts as to whether people are always so thoughtful and in control. Instead, they argued, people are often "mindless" in their behavior and choice making, following stored scripts based on the routines of social interactions. These authors also described people as lacking the mental and attentional capacity to engage in effortful thought on a moment-to-moment basis. Therefore, people are forced into using mental resources in a sparing or "miserly" fashion—relying on simplifying tactics such as heuristic decision rules and stereotypes (see also Dawes, 1976; Kahneman & Tversky, 1973).

The reliance on simple decision rules and on pigeonholing of individuals into stock characters or categories was viewed mainly as a matter of strategic necessity, or even as an adaptive way of dealing with our mental shortcomings as human beings. A miser, after all, is one who is intentionally and deliberately stingy when doling out money, and who jealously guards existing funds. Because the world is filled with unexpected and potentially dangerous events, and these draw heavily on our limited attention (e.g., Fiske, 1980; Pratto & John, 1993), it would seem to be a reasonable strategy to use attention in this miserly fashion.

As an illustration of how such minimal and noneffortful information processing was viewed at the time as a matter of strategic choice, consider the classic Langer, Blank, and Chanowitz (1978) experiments on mindless behavior. These studies demonstrated how participants would react mindlessly in routine situations, evidently not paying much attention to the content (only to the form) of the social interaction. But at the same time, the experiments also showed that participants quickly became mindful, and engaged in effortful processing, when behaving mindlessly would have had important costs to them. For instance, when a confederate trying to cut ahead in the line to use the copying machine promised the participant that this would cause only a short delay, the quality of the excuse given did not matter to the participant's behavior; however, the quality of the reason *did* matter when the promised delay was longer and thus a real inconvenience.

The Cognitive Monster

It soon became apparent, however, that much of the documented reliance on cognitive short-

cuts was not so much a matter of strategic choice as of automatic, unintended processes operating in person perception and social judgment. Automaticity was first raised as a possibility in the closing pages of Taylor and Fiske (1978), and in the 1980s it was applied to and demonstrated for nearly all social-psychological phenomena: trait attributions (e.g., Gilbert, 1989; Gilbert, Pelham, & Krull, 1988; Winter & Uleman, 1984), an attitude's effect on behavior (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), self-judgments (Bargh & Tota, 1988; Paulhus, Graf, & Van Selst, 1989), interpretation of another's behavior (Bargh & Thein, 1985), and of course stereotyping (Brewer, 1988; Deaux & Lewis, 1984; Dovidio, Evans, & Tyler, 1986). The mere perception of easily discernible group features (e.g., skin color, gender, and age-related characteristics) was sufficient in these latter studies to cause the activation of the stereotype associated with the group, which then was shown to influence judgments of a group member in an unintended fashion, outside of a perceiver's awareness (see reviews in Bargh, 1989; Brewer, 1988).

These latter demonstrations were what raised the specter of the cognitive monster of automatic stereotyping. If it were indeed the case, as research appeared to indicate, that stereotyping occurs without an individual's awareness or intention, then the implications for society—specifically, the hope that prejudice and discrimination could eventually be eradicated—were tremendous, as well as tremendously depressing. Most ominously, how could anyone be held responsible, legally or otherwise, for discriminatory or prejudicial behavior when psychological science had shown such effects to occur unintentionally? The legal profession has a term for such a dilemma: the "parade of horribles," in which the appropriate legal cure for an existing unfair or unconstitutional situation itself opens a Pandora's box of still worse evils. Did social psychologists *really* want to go on record as saying that stereotyping and prejudice are uncontrollable?

The "Nightmare Scenario"

More than one social psychologist lost sleep over the implications of these demonstrations of automatic stereotyping, but it was Fiske (1989) who most eloquently captured the essence of the dilemma:

> An absence of intent ultimately implies an absence of responsibility for the effects of categorization. . . . It has led me to have the following nightmare: After testifying for the plaintiff in a case of egregious and demonstrable discrimination, a cognitive social psychologist faces the cross-examining attorney. The hostile attorney, who looms taller than Goliath, says, "Tell us, Professor, do people intend to discriminate?" The cognitive social psychologist hedges about not having any hard data with regard to discrimination, being an expert mainly in stereotyping. When pressed, the psychologist admits that stereotypic cognitions are presumed to underlie discriminatory behavior. Pressed still further, the psychologist reluctantly mumbles that, indeed, a common interpretation of the cognitive approach is that people do not stereotype intentionally, whereupon the cross-examining attorney says in a tone of triumph, "No further questions, Your Honor." The plaintiff is led shaking from the courtroom. . . . (Fiske, 1989, p. 265)

Faced with this possibility, Fiske argued that the mind does not work by cognition alone. Although the cognitive miser may shy away from using such effortful control processes, Fiske argued that properly motivated individuals can "make the hard choice" and overcome the influences of automatically activated stereotypes. In her view, this is theoretically possible because of the general ability of conscious or control processes to dominate and inhibit the influence of automatic processes (e.g., Posner & Snyder, 1975); however, it will depend both on the person's being aware of the nonconscious influence in the first place, and then on his or her having the motivation and also the ability (i.e., enough time and attentional resources) to engage in the control process.

There were those who were skeptical about how often all of these conditions can be met in the real world (e.g., Bargh, 1989; Hilton & Darley, 1991), but at the time this was almost beside the point. As the evidence appeared to show that stereotyping and prejudice are inevitable, or at least highly likely, a way out had to be found; the alternative was to give up and go home.

And so, faced with the cognitive monster, the battle cry of social cognition researchers became (to borrow the words of Isaac Bashevis Singer): "We've got to believe in free will—we have no choice!"

A Critique of "Fiske's Dilemma"

Bringing motivation back into stereotyping research transformed what had been seen as a cognitive inevitability (e.g., Billig, 1985; Hamilton, 1979) into a matter of personal choice and goal setting. It was part of a larger trend within social cognition that started around the mid-1980s (e.g., Neuberg & Fiske, 1987; Sorrentino & Higgins, 1986). In another widely cited paper, also appearing in 1989, Devine divided the phenomenon of automatic stereotyping into two distinct components: stereotype activation and stereotype application, with only the first stage being automatic and inevitable. The application and use of the activated stereotype in person perception and judgment were argued to be under motivational control.

The dilemma posed by Fiske (1989) was between the pursuit of scientific insight into how much control an individual does have over prejudicial judgments and behavior on the one hand, and the implications of that knowledge for society on the other. The possibility was clearly and dramatically sketched of a "parade of horribles" if society and the legal system adopted the apparent findings of psychological research as to the unintentional nature of stereotyping. If people cannot help stereotyping, then they cannot be held personally responsible for their actions, and so cannot be sanctioned for any prejudicial actions.

Yet, as participants in a scientific enterprise, we enter upon a very slippery slope when we attempt to tailor our conclusions to fit what we believe to be good for society. We should not be guided, as a field, by a motivation to demonstrate and conclude from our research findings that automatic stereotypes are in fact controllable. No matter how well-meaning and virtuous this intention may be, it can do nothing but compromise our objectivity as psychological scientists.

Societal versus Scientific Concepts of Personal Responsibility

Whether or not personal responsibility over stereotyping and prejudice exists in fact, there is no doubt that society has a powerful interest in presuming that it does. The notion of individual responsibility and culpability for one's own actions is the bedrock of every legal system, and the "parade of horribles" that would follow from abandonment of this notion would indeed be calamitous. This is one reason why, even if psychologists were someday to prove beyond a shadow of a doubt that free will does not exist, society would nonetheless continue to hold people personally responsible for their actions (see Koestler, 1967, Ch. 1). It would cease to function otherwise.

In a recent essay, Prinz (1996) distinguishes between the *scientific status* and the *moral function* of the concept of "free will." The latter sense of the concept is described as "a social construction aimed toward the societal control of actions." According to this analysis, psychologists focus on the causal explanation of actions, whereas society is much more interested in their evaluation and moral justification than in their causal explanation. But the important point is that the societal and the scientific versions of the concept of personal responsibility can coexist independently of each other. The social utility of the idea of individual responsibility exists, the scientific evidence regarding personal control notwithstanding. Therefore, society's use of the notion of free will is something that can never be dictated to by psychological research findings. At the same time, psychological research into the degree of personal control over stereotyping and prejudice is of enormous theoretical and practical significance. *And so, by the same logic, the conclusions of this research should not be dictated by the agenda of society.*

Prinz (1996) frames the question thus:

> How should we view this? Where is the hen and where is the egg? Should we consider free will to be a basic fact of our psychological make-up that, quite arbitrarily, has desirable social side effects—or as the product of a social construction aimed toward the societal control of actions? Do the psychological facts precede the moral function—or should we at last realize that the moral function elicits psychological facts?

The Monster in Chains

Fiske's (1989) analysis of the interplay between cognitive factors (stereotype activation) and motivational factors (effortful control) in

prejudice gave free will and intentionality a chance again. The pendulum was swinging back to a meta-assumption of strategic control over stereotyping, as it had been assumed in the era of the cognitive miser. Note, however, that there was something of a strategic retreat regarding the role of free will in person perception: Whereas the initial deployment and use of categories and stereotypes had been thought to be under the miser's strategic control, now only the subsequent influence of the already activated category or stereotype was believed to be controllable.

Two Stages of Stereotyping

That same year, as noted above, Devine (1989) published a similar analysis of distinct activation and application stages of stereotyping. With the perspective of time, what may well have contributed to the impact and importance of Devine's paper was that it presented, simultaneously, the scariest version of the cognitive monster yet imagined, followed by reasons why and evidence as to how this worst-case scenario was not really the intractable problem we had assumed. This was indeed welcome news: If even the most threatening version of stereotype automaticity could be shown to be ultimately controllable, as Devine argued, then the monster could be fought, and perhaps further diminished in power by additional discoveries.

Devine (1989) found that all of her participants, regardless of their expressed or explicit level of prejudice against African Americans, could accurately report the content of the culture's stereotype of that group. Next—and most chillingly—all of the participants showed evidence of having this stereotype automatically activated by subliminally presented stereotypic features. In her Experiment 2, Devine (1989) showed effects of the primed stereotype on judgments of a target person's degree of hostility, even though in her design the primes did not include "hostile" or a synonym for it. Therefore, the African American stereotype had become active automatically and had affected opinions about a person's hostility in the absence of any information in the experimental situation concerning hostility—exactly the "going beyond the information given" and filling-in-the-blanks function long associated with stereotyping—yet the experimental participants never intended to stereotype, and low-prejudice individuals showed the effect as much as did high-prejudice individuals.

Therefore, even if a person expressed egalitarian and nonprejudiced beliefs concerning African Americans, he or she nonetheless seemed just as vulnerable to automatic stereotype influences as were prejudiced individuals. Where in these data was there any reason for optimism that a person's values, and egalitarian motives, and good intentions could prevent cultural stereotypes from influencing judgments of minority group members?

Yet there was room for hope. It came from Devine's (1989) theoretical analysis of stereotype automaticity into two components: the activation stage and the application stage. In her view, the activation stage is completely automatic (assuming a stereotype that permeates the culture), in that one cannot help having the stereotype activated by relevant group features. This is what Experiment 2 showed. Experiment 3, on the other hand, concerned the application stage. Assuming an activated stereotype, can an individual consciously control its application to a target individual? In Experiment 3, the participants' level of expressed racism did relate to how stereotypic and prejudicial were thoughts listed about the category "blacks": High-prejudice participants' thoughts were more negative and stereotypic than those of low-prejudice participants.

Further Restrictions on Automatic Stereotype Influences

With the momentum swinging back in the direction of conscious control and away from demonstrations of automaticity, other limits to the phenomenon of automatic stereotyping were documented in short order. Devine (1989) had argued that even though stereotypes become active automatically, their influence on judgment is under motivational control. Neuberg (1989, 1994) showed that the self-fulfilling prophecy effects of negative stereotypic expectancies could be controlled if participants had a conscious goal to form an accurate impression of the target person (see also Fiske & Neuberg, 1990). Gilbert and Hixon (1991) argued that even the stereotype

activation stage may not be automatic under all conditions, because it did not happen when participants were busy with a secondary task (holding a series of digits in memory) while a member of a stereotyped group was present. Blair and Banaji (1996) concluded that the initial stage of automatic activation can be prevented through counterstereotypic expectancies. And Jussim (e.g., 1990) argued that if all else fails, and stereotypes are nonconsciously activated and their influence not then controlled, this influence should nonetheless be benign because stereotypes are accurate descriptions of the social group (see also Lee, Jussim, & McCauley, 1995).

Perhaps I'm just a pessimist, but I don't buy any of this. It would be *nice* if stereotypes were found not to be activated automatically. It would be *nice* if, failing that and stereotypes were found to be automatically activated, then it was found that an individual could prevent this activation by having a conscious, counterstereotypic expectancy. It would be *nice* if, even if automatic activation could not be shown to be prevented in this (or any other) way, individuals were found to be indeed cognizant of the possibility of being nonconsciously influenced, and when aware of that influence, to have the motivation and the time to effortfully control it. And it would be *nice* if, even if all these propositions failed and stereotypes were shown to be automatically activated and to affect perceptions of and behavior toward a member of a minority group, this influence was still found to be benign because the group stereotype was a demonstrably accurate portrait of the target individual.

All of this would indeed be nice—if it were true. But the relevant research evidence largely contradicts this rosy picture.

WHAT DOES THE EVIDENCE REALLY SAY ABOUT STEREOTYPE CONTROLLABILITY?

In this section, I first provide a critical appraisal of the evidence as to (1) whether automatic stereotype activation is likely, and, if so, whether such activation is unconditional or depends on the perceiver's current processing goal; (2) whether automatic activation can be eliminated by counterstereotypic expectancies; and (3) whether the influence of a stereotype, once activated, on judgments and behavior is likely to be controlled by the perceiver. Following this review, new evidence supporting a dual-process model of self-fulfilling prophecy effects is described.

Are Stereotypes Always Activated?

The Role of Attentional Resources

Gilbert and Hixon (1991) showed that a load on attentional resources—induced by giving participants a secondary task to perform—disrupted the otherwise automatic activation of the Asian American stereotype. Whereas participants made more stereotypic completions of word stems (e.g., "shy" for "s_ y") when an Asian American woman presented those stems in a video presentation, compared to baseline levels, this effect disappeared if they had to hold a series of digits in memory simultaneously. Thus, at least some stereotypes did not become active upon the mere presence of group features, if the perceiver had been "cognitively busy" at the time.

One question concerning this finding is its generalizability to other stereotypes. Perhaps those for women, African Americans, and the elderly are "stronger" stereotypes (at least in U.S. society), and consequently are more efficiently activated and not prevented from activation by a shortage of attention. Recently, however, Spencer, Fein, Wolfe, Fong, and Dunn (1998) have replicated the Gilbert and Hixon (1991) finding for the African American stereotype, so it does seem that when attention is divided, stereotypes are less likely to become activated.

However, several other studies, beginning with Devine (1989, Experiment 2), show that the African American stereotype is so efficient that it can become activated even with subliminal presentation of group features (see Bargh, Chen, & Burrows, 1996, Experiment 3; Chen & Bargh, 1997; Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995, Experiment 1), that is, with no conscious attentional processing needed. It therefore seems odd that a shortage of attentional resources would knock out the stereotype activation effect in the Gilbert and Hixon paradigm. This is a

clue that the particular manipulation employed may have blocked stereotype activation for a reason other than, or in addition to, the shortage of processing resources it produced.

Macrae, Bodenhausen, Milne, Thorn, and Castelli (1997) have proposed one such alternative explanation for the Gilbert and Hixon (1991) findings. They presented faces of female undergraduates as well as pictures of common household objects to their participants. Some participants were instructed to detect, by means of a key press, whether or not a white dot appeared on each photograph. Others were told merely to press a key upon the presentation of each stimulus photograph on the screen. A third group was directed to process the stimuli in a semantic manner, deciding whether a given photograph was of an animate or an inanimate object. On each trial, following the response to the photograph, a participant was presented a word string and was to indicate, as quickly as possible, whether it was a word or a nonword. Some of the word strings were stereotypic of women, while others were counterstereotypic.

The results showed that on trials on which a photograph of a woman had just been presented, but not on the other trials, responses were faster to the stereotypic word strings than to the counterstereotypic ones in this lexical-decision task. However, as predicted, this effect held only for those participants who had processed the photographs in a semantic fashion; that is, responses were based on the content and meaning of the photograph. Those who had searched each photo for the presence of a white dot, or who had merely pressed a button to indicate the presentation of any stimulus, did not show any evidence of stereotype activation.

This study is reminiscent of earlier work by Uleman and Moskowitz (1994), who studied the goal dependence of spontaneous trait inferences by varying the participants' processing goals during presentation of trait-implying sentences. Across three experiments, Uleman and Moskowitz found that behavioral stimuli automatically activated the trait concept they implied (e.g., "considerate" for "The deliveryman slows down and motions the pedestrians to cross"), as long as the participants processed the behavioral description for meaning in some way—either by indicating (via key press) the gender of the actor, by forming an impression of the actor, by deciding whether they themselves would engage in the given behavior, or by deciding whether they were similar or not to the actor. But in other conditions in which the meaning of the sentence was not relevant to the judgment task, such as detecting the appearance of specific letter combinations, spontaneous trait inferences were reduced or eliminated.

Taken together, these studies suggest that as long as a perceiver is dealing with a target individual as a social being—that is, whenever the perceiver is making judgments about or forming impressions of the target—trait concepts and stereotypes relevant to that target individual will become active automatically. However, if the perceiver is instead dealing with the other person not as a person at all, but as a device that turns cards, or as a stimulus that may or may not contain a white spot, then these social concepts will not be activated.

Beyond the hopeful implication that dermatologists are unlikely to stereotype their patients, what is the "real-world" relevance of studies involving such presemantic processing goals? The results seem to suggest that when we are dealing with minority group members not as people, but as stimuli or devices, the group stereotype will not become active; however, as long as we *are* dealing with them as people (even if it is only to memorize their faces), the stereotype will become active automatically. This may not be a trivial reduction of the scope of automatic stereotyping effects, because we often encounter people, especially functionaries, in whom we have no interest as social beings.

The Role of Processing Goals

In addition, examining the generality of automatic stereotype activation across a variety of processing goals, as Macrae et al. (1997) did, may well turn out to be a promising line of attack on the cognitive monster. The telling evidence is yet to come, however. For instance, people may have many interpersonal goals during social interaction other

than impression formation or evaluation (Chaiken, Liberman, & Eagly, 1989; Gollwitzer & Moskowitz, 1996; Hilton & Darley, 1991; Jones & Thibaut, 1958), and some or most of these well may override, prevent, or suppress stereotype activation. Examples include social comparison, having a good time, competing, self-enhancement, self-presentation, having a smooth rather than awkward interaction with the other person, and information seeking, as well as the pursuit of personal goals that involve the other's help or assistance (e.g., a salesperson's advice in making a purchase; negotiation with that same salesperson over the price; working with superiors, subordinates, or equals on a shared task). These are the kind of goals that seem quite relevant to the question of how widespread (vs. goal-dependent) automatic stereotype activation actually is in social life.

One recent study underscores the importance of goals and motivations in automatic stereotyping effects. Spencer et al. (1998) showed that even under the original Gilbert–Hixon conditions that knocked out stereotype activation, the stereotype was nonetheless automatically activated if a participant had just suffered a blow to his or her self-esteem. The moral here is that the impact of motivation on stereotyping is a double-edged sword. Just as motivation *not* to stereotype can possibly overcome the automatic influence, as Devine (1989) and Fiske (1989) argued, so too can there be motivations to *use* stereotypes that can overcome conditions that otherwise successfully block them.

Can Stereotype Activation Be Eliminated by Expectancies?

Blair and Banaji (1996) have approached the cognitive monster from a different direction. Devine (1989) had distinguished between stereotype activation and stereotype application, and argued that only the latter is potentially controllable by the appropriate motivation (i.e., to be egalitarian and nonprejudiced). Blair and Banaji (1996) have gone still further and argued that motivational control is possible even over the first stage of stereotype activation.

In their Experiment 1, participants were faster to classify names as male or female when those names were immediately preceded (i.e., with a stimulus onset asynchrony [SOA] of 350 milliseconds [ms]) by stereotypic trait and nontrait primes consistent with the gender of the target name (e.g., "aggressive–Mike," "petite–Carol") than when the primes were consistent with the opposite gender stereotype (e.g., "flowers–Tom," "briefcase–Susan"). Previous studies had consistently shown that an SOA of 350 ms or less is too brief for an effect of expectancy or intention on responses (e.g., Fazio et al., 1986; Neely, 1977; see review in Neely, 1991), and so the results of Blair and Banaji's (1996) Experiment 1 were consistent with the hypothesis of automatic activation of gender stereotypes.

In Experiment 3, again with a 350-ms SOA between prime and target on each trial, half of the participants were presented with stereotype-inconsistent primes on most of the trials. That is, male names were usually preceded by primes consistent with the female stereotype, and female names were usually preceded by male stereotypic terms. Participants were explicitly informed, in fact, to expect this combination. The remaining participants were presented most of the time with stereotype-congruent prime–target combinations, and were told to expect that combination.

The startling apparent outcome of the experiment was that for the participants told to expect stereotype-inconsistent prime–target combinations, there was no longer any significant response time advantage for the stereotype-consistent primes, as there had been in Experiment 1 with no expectancies operating. Response times when the stereotypic trait primes matched the target names' gender did not differ significantly from the times when the primes corresponded to the opposite-gender stereotype. The authors conclude that their results "suggest that stereotype activation may not be unconditional and stereotypic cues need not result in a stereotypic response," and moreover that the findings "support proposals . . . that perceivers can control and even eliminate such effects" (Blair & Banaji, 1996, p. 1159).[1]

The present Figure 18.1, however, tells a somewhat different story. It presents the mean response latencies from Experiment 1, from which the authors conclude that the automatic stereotype effect did occur, and from

the counterstereotype expectancy (350-ms SOA) condition of Experiment 3, from which they conclude that the automatic stereotype effect did not occur—that is, that the counterstereotypic expectancy successfully controlled and eliminated the automatic stereotype. It is quite evident that the two patterns are nearly identical (if anything, the effect in Experiment 3 was stronger). If one considers the amount of response time facilitation produced by the stereotype-consistent primes compared to the stereotype-inconsistent primes (the difference between the average of the F-F and M-M means and the average of the F-M and M-F means), this facilitation effect was 8 ms in the no-expectancy condition (Experiment 1), but 13 ms under the counterexpectancy conditions (Experiment 3) that were said to have eliminated the effect.

The authors' conclusion that the stereotype activation effect was moderated by counterstereotypic expectations was based on the statistical significance of the Prime × Target interaction in Experiment 1 but not in Experiment 3. However, it should be noted that the Experiment 3 participants were the same people as those who took part in Experiment 1, and that they did so after a 5-minute break (half went into the stereotype expectancy condition and the other half into the counterstereotype expectancy condition of Experiment 3; Blair & Banaji, 1996, p. 1150). Therefore, it would be possible to test whether the expectancy manipulation moderated the size of the Prime × Target interaction, because the same participants experienced both no expectancy in Experiment 1 and the counterstereotypic expectancy in Experiment 3. Blair and Banaji (1996) did not report such a test in the article; I. Blair (personal communication, November 8, 1997) reports that the authors did not consider such a comparison to be valid "because participants reported that the priming tasks made them fatigued, and stereotype effects are more likely when

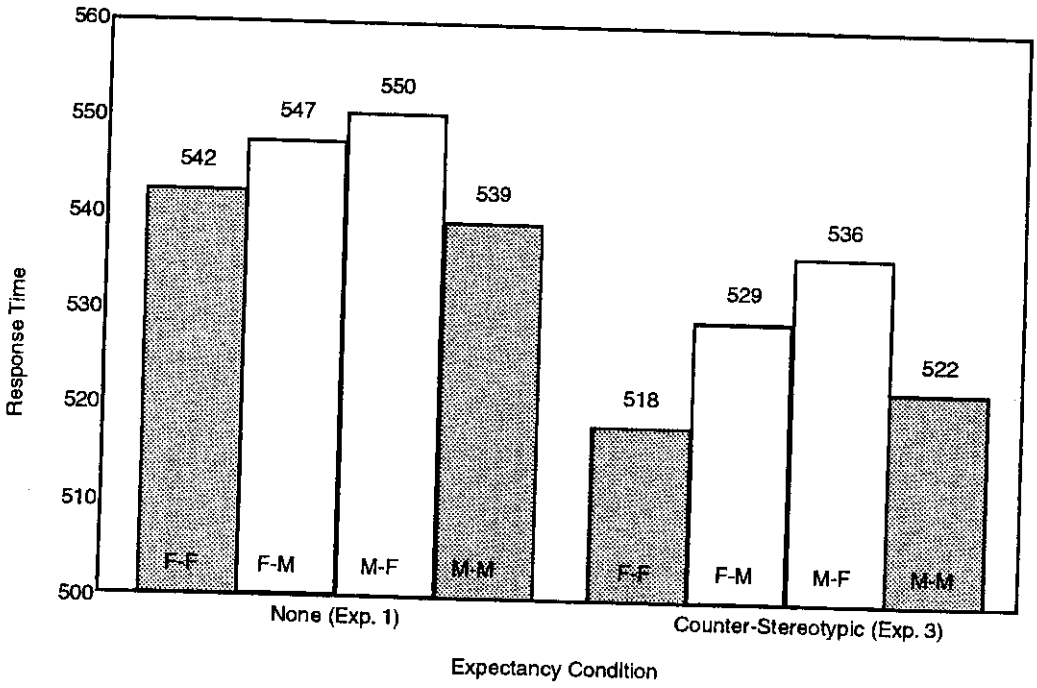


FIGURE 18.1. Mean response latencies to classify common names as male or female, by prime (male vs. female stereotypic trait) and expectancy for prime–target combinations (none vs. counterstereotypic). The first letter on each bar denotes the gender of the prime, and the second denotes the gender of the target stimulus. Shaded bars represent conditions in which prime and target matched in gender. The data are from Blair and Banaji (1996, Experiments 1 and 3, 350-ms SOA conditions).

people are tired." Nonetheless, it appears unlikely, given the results shown in the present Figure 18.1, that such a within-subjects test would reveal that the counterstereotype expectancy conditions had attenuated the effect of prime on target in any way.[2]

Nevertheless, these findings have been cited widely in support of the opposite conclusion: as "demonstrating that perceivers' expectations can impede stereotype activation" (Macrae et al., 1997, p. 474); as showing "that conscious efforts to suppress stereotypically biased reactions can inhibit even the immediate activation of normally automatic associations" (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997, p. 536); and as demonstrating "that, if perceivers have counterstereotypical replacement thoughts, stereotypes may be inhibited . . . [which is] particularly encouraging in that the effects of the counterstereotypic expectations were revealed within milliseconds, which does not allow enough time for controlled processing" (Monteith, Sherman, & Devine, 1998, p. 71).[3] My own assessment differs; to me, the Blair and Banaji (1996) results are actually the best evidence to date that automatic stereotype activation is *impervious* to cognitive control. That is, even when participants were told that on most trials the gender of the prime name would signal a stereotypic trait associated with the opposite gender, and then encountered a majority of such trials, the automatic stereotype activation effect was unchanged.

Alas, the story gets even bleaker for the possibility of controlling stereotypes through expectancies. The Blair and Banaji (1996) Experiment 3 did find a strikingly powerful effect of expectancies on the size of the automatic stereotype effect—but it was the effect of stereotype-*consistent* expectations, which dramatically *increased* the automatic stereotype effect (see their Figure 4, p. 1152). When participants were notified that stereotype-consistent pairings between prime and target would occur on the majority of the trials, and then experienced these mostly stereotype-consistent trials, the conscious stereotype-consistent expectancy made the automatic stereotype effect even stronger than the automatic effect was by itself, with no expectancy operating, in Experiment 1. When no expectancy was operating (Experiment 1), the average stereotype facilitation effect on response latencies was 8 m; by contrast, when participants in Experiment 3 held stereotype-consistent expectancies, the average automatic stereotype effect increased to 96 milliseconds, or *12 times greater.*

The reality is that when people have expectancies about stereotyped groups, those expectancies are for stereotypic—not counterstereotypic—behavior (e.g., Jones, 1990; Neuberg, 1989, 1994). Thus, Blair and Banaji's (1996) findings that expectations either leave the automatic stereotype effect alone if they are inconsistent with the stereotype, or substantially enhance it if they are consistent with the stereotype, are hardly good news about the chances of reducing automatic stereotyping through expectations.

In sum, with enemies like expectancies, the cognitive monster doesn't need friends.

Is Successful Control Following Automatic Activation Likely?

The ability to control a stereotype (given motivation to do so) depends heavily on one's awareness of the possibility of unconscious prejudicial influence, but also on one's theory of how that unconscious prejudice may be manifested and expressed (see Wilson & Brekke, 1994). Devine's (1989) Experiment 3 instructed participants to write down the characteristics of the typical African American, and here people who were not prejudiced (according to an explicit measure of prejudice, the Modern Racism Scale) produced more positive and less stereotypic descriptions. But stereotypic assumptions and beliefs can emerge and be expressed in ways about which a person with egalitarian motives has no theory concerning the unfelt influence. Examples of this phenomenon include the linguistic intergroup bias (Semin & Fiedler, 1988; Maass, Salvi, Arcuri, & Semin, 1989; von Hippel, Sekaquaptewa, & Vargas, 1997), in which the same behavior is described in more abstract, pansituational terms when it confirms a stereotype (e.g., "Ramon was violent") than when it is performed by a member of a nonstereotyped group (e.g., "Bill pushed the reporter away from his car"), or the tendency when completing behavior stem sentences to explain the reasons for counterstereotypic but not stereotype-consistent behavior

(von Hippel, Sekaquaptewa, & Vargas, 1995). Stereotype effects emerge in the "tacit inferences" people make when interpreting the meaning of others' behavior; for instance, upon hearing "Some felt that the politician's statements were untrue," people assume that the politician was lying, but do not make this assumption if the actor was a physicist instead (Dunning & Sherman, 1997). Again, these implicit effects are just as likely for those who score low on explicit measures of stereotypic beliefs as for those who score high (Dunning & Sherman, 1997, Study 5).[4]

Stereotype effects appear as well in judgments of fame (Banaji & Greenwald, 1995), in larger priming effects when the primed trait category is stereotype-relevant (Banaji, Hardin, & Rothman, 1993), and also in response latencies to evaluate good and bad adjectives (Dovidio et al., 1997; Fazio et al., 1995). Here again, they are independent of explicit, expressed measures of prejudice. In other words, when a participant does not realize how his or her response can be a sign of stereotyping and prejudice, he or she manifests the stereotype. Even well-meaning attempts at controlling prejudicial influences require knowing what those influences are (Wilson & Brekke, 1994), and for those that are less obvious than overtly describing characteristics of minority groups, the experimental evidence taken as a whole provides scant indications of control being exerted. And if not in psychology experiments, where people know that they are being measured and their behavior scrutinized, why should we expect such control to be exercised elsewhere?

Even if an individual is aware of possibly being prejudiced, and is motivated to engage in control over the stereotype in question, such control attempts often backfire. It is an "ironic" effect of mental control attempts that suppressed thoughts often bounce back, becoming even more accessible than before, when the person's guard is down and there is a letup in control (Wegner, 1994). Consistent with this principle, Macrae, Bodenhausen, Milne, and Jetten (1994) found that participants who had been suppressing stereotypic thoughts about others subsequently responded more negatively to a stereotyped target, compared to those who had not been attempting to control such thoughts. And Wegner, Erber, Bowman, and Shelton (1997) showed that ongoing control attempts (not to be sexist) backfired under conditions of mental load (time pressure): Participants trying to control their stereotypic assumptions about females actually made more sexist remarks under such conditions than did participants not attempting to control their assumptions. Though attempts at suppression do not always produce such rebound effects (see Monteith et al., 1998), the point is that even in the unlikely event that both of the necessary conditions—awareness of nonconscious stereotype operation and motivation to do something about it—are in place, stereotypic judgments and behavior can nonetheless occur.

Do Behavioral-Confirmation Effects No Longer Exist?

Another avenue of attack on the enormous problem posed by automatic stereotyping has been in the area of self-fulfilling prophecies or behavioral confirmation of stereotypes (e.g., Jussim, 1986; Merton, 1948; Rosenthal & Jacobsen, 1968; Snyder, Tanke, & Berscheid, 1977). The standard model of such effects (for reviews, see Darley & Fazio, 1980; Hamilton, Sherman, & Ruvolo, 1990; Jones, 1990; Olson, Roese, & Zanna, 1996; Snyder, 1984) assumes that activated stereotypes generate negative expectancies concerning the behavior of the minority group member, causing the perceiver to behave toward the other person in such a way as to produce the very stereotype-consistent behavior he or she expects. For instance, a teacher assuming a lack of intelligence or promise from a pupil may spend less time with him or her and in other ways communicate those assumptions, causing the pupil's performance to suffer. This mechanism by which stereotypes perpetuate themselves has also been called into question over the past 10 years.

Accuracy Motivation

Neuberg (1989) manipulated his participants' motivation to produce accurate impressions of a target person by offering a monetary prize for the most accurate judgment, and this

did create more positive behavior toward the stereotyped-group member and consequently reduced stereotypic judgments. The behavior in question had to do with the questions asked by participants while interviewing the target person; these were less expectancy-confirming than otherwise (e.g., Snyder & Swann, 1978).

But this is a type of behavior (i.e., verbal) that is under people's cognitive control to a great extent; other forms of behavior in naturalistic social interactions are less controllable (e.g., nonverbal) and provide more possibility for "leakage" from stereotypic assumptions. Fazio and Dunton (1997) conclude that when "the behavior is not easily controllable, the behavior should be less influenced by motivational concerns and more directly and singly influenced by any automatically activated evaluations" (p. 469), and they point to nonverbal behaviors as a good example of such difficult-to-control behaviors (see Dovidio, Brigham, Johnson, & Gaertner, 1996).

Accuracy motivation as a method of reducing stereotyping is not promising for other reasons as well. Even when perceivers are accuracy-motivated, their effortful or systematic processing of person information is influenced by their stereotypes (Chaiken & Maheswaran, 1994; Chen & Chaiken, Chapter 4, this volume; Trope & Alfieri, 1997); for example, stereotypes affect how the target behavior is identified or encoded initially, before the accuracy-driven systematic processing begins (see also Darley & Gross, 1983; Dunning & Sherman, 1997; von Hippel et al., 1995). Furthermore, several recent reviews of expectancy and goal effects have concluded that it is relatively rare in naturally occurring situations for an individual to have the goal of forming an accurate impression (e.g., Hilton & Darley, 1991; Olson et al., 1996).

A Methodological Critique of Behavioral-Confirmation Studies

Jussim and his colleagues (e.g., Jussim, 1990; Madon, Jussim, & Eccles, 1997, p. 792) have also questioned the existence of behavioral-confirmation effects, on both methodological and theoretical grounds. They point out that experimental demonstrations of self-fulfilling prophecy effects typically induce false expectancies in participants, so that such studies do not address the likelihood that such false beliefs develop on their own, without such interventions on the part of the experimenter. According to Jussim (1990, p. 30), "Despite some grandiose claims to the contrary, there is currently no evidence that naturally occurring expectations lead to huge self-fulfilling prophecy effects or perceptual biases." Jussim's argument is that if the stereotype is *accurate* (Lee et al., 1995), the expectancies it generates in the perceiver are not false, and thus are not factors contributing to the stereotype-consistent behavior of the target person.

A Nonconscious Route to Behavioral-Confirmation Effects

A recent study (Chen & Bargh, 1997), however, demonstrated that self-fulfilling prophecy effects could be produced experimentally without giving the participants any expectancies at all. Instead, a second, nonconscious route from stereotype activation to the production of confirming behavior in the stereotyped-group member was documented, distinct from the standard route through consciously held expectancies (see Figure 18.2). Building on the long-standing hypothesis in psychology of a direct, "express" link between perception and behavior (see next paragraph), Chen and I argued that automatic stereotype activation can produce a tendency for a perceiver to be the first to act in a stereotype-consistent manner within an interaction with a stereotyped-group member. This may well produce similar behavior in response by the target. However, without awareness of the effect of the stereotype on his or her own behavior, the perceiver will be highly likely to interpret the response of the other as just another confirming instance of the stereotype (cf. Jones & Nisbett, 1971).

The idea that mental activity can directly affect behavior without an intervening act of will was championed by William James (1890) as the principle of "ideomotor action." James argued that merely thinking about a behavior increases its probability of occurrence because of the impulsive nature of consciousness, which does not require an act of "express fiat" or will prior to each behavioral response to the environment. James in

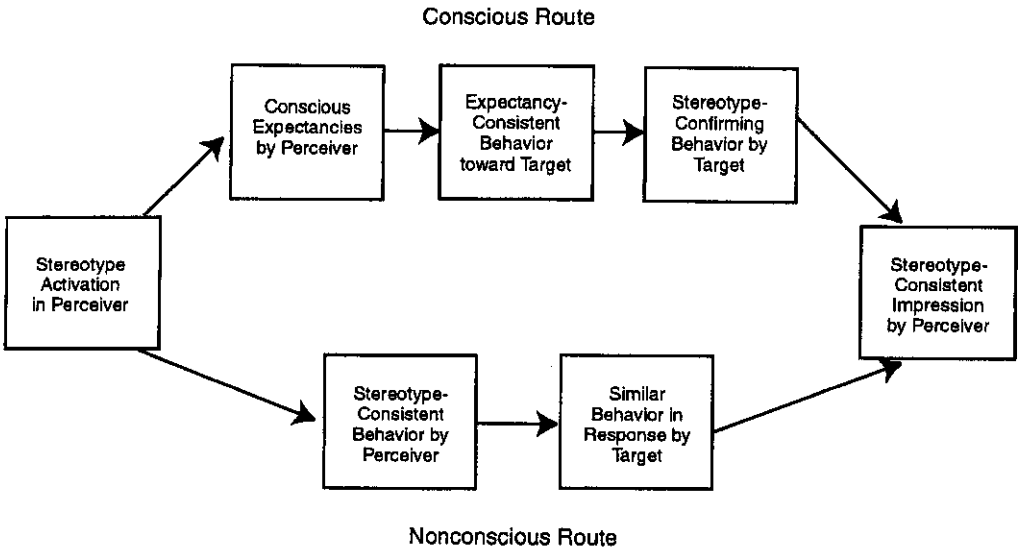


FIGURE 18.2. A dual-process model of behavioral confirmation of stereotypes.

fact argued that such cases of intervention by the will are the exception and not the rule. Later, the Gestalt psychologists endorsed the notion, and Piaget (e.g., 1948) considered it an indispensable mechanism of imitative learning. More recently, Berkowitz (e.g., 1984) and Carver, Ganellen, Froming, and Chambers (1983) argued that an express link between perception and behavior produces passive and unintended media effects on behavior, as well as interpersonal effects.

We (Bargh et al., 1996) combined the idea of a passive perception–behavior link with the evidence of automatic stereotype activation. If stereotypes can be activated automatically, without awareness or intention on the part of the perceiver, then this perceptual activity should activate related behavioral tendencies. That is, we assumed (in line with the common-coding model of Prinz, 1990) that the trait concepts activated in the course of stereotype activation contain within them knowledge not only of how to detect such trait-like behavior in others (e.g., what it means to act honestly or aggressively, etc.) but also of how to produce the same behavior oneself.

In Experiment 2 of the Bargh et al. (1996) study, the stereotype of the elderly was primed through a standard method, the "scrambled-sentence test" (see Bargh & Chartrand, in press; Srull & Wyer, 1979). Words related to the stereotype (e.g., "conservative," "grey," "bingo") were embedded for some participants in a sentence construction task, while the remaining participants received a version of this task with neutral primes. Then participants were thanked for their help and left the experimental room, believing that the experiment was over. However, a confederate posing as the next participant was waiting outside in the hall, and surreptitiously timed how long it took the participants to walk down the hallway on the way to the elevator. The prediction was that because slowness and weakness were components of the elderly stereotype, participants whose stereotype of the elderly had been activated without their awareness would walk more slowly down the hall than the other participants would. Because neither slowness nor weakness (or synonyms of either) appeared as priming stimuli, such an effect could only occur through activation of the elderly stereotype. The prediction was confirmed in two experiments.

The Bargh et al. (1996) Experiment 3 focused on the African American stereotype. First, some participants were subliminally primed with photographs of young black

male faces, and others with young white male faces (all participants were white). Then a mild provocation was staged, in which the experimenter requested the participant to do the fairly boring experimental task again because of a computer error in saving the data. A hidden video camera recorded each participant's reaction to this request, and judges who were unaware of the experimental hypotheses rated these tapes later for the degree of hostility shown by the participants. Hostility is a trait component of the African American stereotype (e.g., Devine, 1989). Again as predicted, those whose stereotype of African Americans had been primed subliminally showed greater hostility in response to the provocation than did the other participants, consistent with the hypothesized behavioral effects of the activated stereotype. These stereotype-priming effects on behavior have subsequently been replicated and extended for different stereotypes (e.g., "professor," "soccer hooligan") by Dijksterhuis and van Knippenberg (1998).

The implications of these findings for self-fulfilling prophecies are evident. It has long been established that there are nonverbal modes through which stereotypic affect and beliefs are expressed (e.g., Dovidio et al., 1996; Word, Zanna, & Cooper, 1974), and that nonverbal expressions of affect are difficult to control (e.g., DePaulo, 1992; Ekman & Friesen, 1969; Fazio & Dunton, 1997). Nonverbal expressions of hostility (tone of voice, facial expression, aversion of eye contact, etc.) are also perceived without any difficulty by their target. Thus, it is easily conceivable that activation of the concept of hostility through one's perception of an African American may quickly and nonconsciously cause such nonverbal manifestations in the perceiver, as well as possibly more overt expressions in verbal or behavioral content. The target person may understandably respond in kind with hostility, and this should be perceived by the stereotyper—lacking awareness of his or her own role in producing the interpretation—as occurring without any provocation. Thus, the stereotyper should attribute the hostility to the dispositional characteristics of the target person.

We (Chen & Bargh, 1997) designed a study to test this implication. In essence, the design was a merging of two paradigms—those of automatic stereotype activation and of behavioral confirmation (as described by Snyder et al., 1977). Pairs of participants played a game of "Catch Phrase"—similar to the old television game show *Password*—in which one player attempts to get his or her partner to guess each of several target words. The game readily creates mild frustration and sometimes even anger, as one's partner often cannot guess the word, despite what the clue giver often believes to be (biased by his or her knowledge of the word) excellent and obvious hints. In our experimental version of this game, players were in different rooms and communicated via headphones and microphones. We recorded the two players' voices on separate channels of a stereo tape recorder.

Prior to playing the game together, the individual members of a pair performed the priming task alone. This was the same task as in the Bargh et al. (1996) Experiment 3. One member of the pair in each session was randomly assigned the role of "perceiver," and the other the role of "target." During this task, half of the perceiver participants were presented subliminally with black faces, and the other perceiver participants with white faces. All of the target participants were presented with white faces; that is, for none of the target participants was the African American stereotype automatically activated.

Following the game, each participant rated his or her partner on a variety of traits, including hostility. We also had judges who were unaware of the experimental hypotheses listen to the audiotapes in a random order and rate each participant on hostility, as well as on stereotype-irrelevant trait dimensions. The audiotape ratings showed that, as predicted, both the perceiver and the target participants in the stereotype-primed condition were rated as being more hostile than their counterparts in the no-priming condition. Thus, the subliminal priming manipulation increased the hostility of the perceiver participants, and this in turn increased the hostility of the target participants. A path analysis confirmed that the greater the hostility shown by a perceiver participant, the greater was the hostility evidenced by the target participant. Finally, the signature of the behavioral-confirmation effect was also observed: Perceiver participants primed with subliminal black faces rated their interaction partners as

being more hostile, compared to nonprimed perceivers' ratings of their partners. That this effect was driven specifically by the activated African American stereotype was further shown by the fact it held only for ratings of hostility and not for the overall negativity of the ratings on stereotype-irrelevant trait dimensions.

This demonstration of self-fulfilling prophecy effects is free from the objections raised concerning the phenomenon by Jussim, Neuberg, and others. The study differed from previous experimental tests in that we did not give our participants any false expectancies; we merely showed them young black male faces outside of awareness. That this presentation evidently activated the stereotype of African Americans, and thus the concept of hostility, had to be because this concept was stored in the memories of the participants—it was not present in their environment.

Furthermore, as to the contention that stereotypes are benign influences on social perception and behavior because they are accurate, note that all of the participants in the Chen and Bargh (1997) study were white and not black. Thus there was no way that the behavior of the target persons could have been actually, or "accurately," more hostile (i.e., because of any presumed "actual" greater hostility of African Americans) to then initiate the greater hostility found in the perceiver–target pairs where the perceivers had been primed with black faces. If the targets had been black, the argument could possibly have been leveled by those who take the position of stereotype accuracy that the targets did behave first with greater hostility because the stereotype is accurate and black people are more hostile. This cannot be raised as an objection to the Chen and Bargh (1997) behavioral-confirmation effects.

In short, the results newly document an insidious way in which stereotypes can be confirmed—unconsciously, because the effect does not depend on consciously held expectancies and their guidance of behavior toward the target person. Although it may be possible for a person with the motivation to behave in a fair and egalitarian manner to be aware of and then control negative, consciously operating stereotypic expectancies, it is less clear how even such motivated individuals will have the opportunity to control the immediate, nonconscious effect of the stereotype on their own behavior (see Fazio & Dunton, 1997, for a similar argument regarding the importance of opportunity conditions).

But perhaps this conclusion is premature, given the currently available evidence. Studies such as ours (Chen & Bargh, 1997) and that of Devine (1989) deliberately did not have participants interact with or read about a member of the focal stereotyped group (i.e., the target person was not African American), in order to rule out conscious or intentional sources of the stereotyping effect. But this also eliminated a potential cue in the situation that might have triggered stereotype suppression or control strategies for at least some participants. In other words, if our white participants had interacted with a black target, perhaps they would have been alerted to the possibility of unintended prejudicial behavior and would have exercised greater care and guidance over their behavior.

Further research is needed on this point, but merely to show that such experimental effects vary with black versus white targets (if indeed they do) is not enough; the question is not how people behave in a psychology experiment setting, in which they are more likely to be on their guard, but how they naturally react to members of stereotyped versus nonstereotyped groups in real life. My feeling is that the Chen and Bargh (1997) findings do generalize to actual encounters between black and white Americans, because these strategic "guards" are often down in both cases, and we are dealing with immediate behavioral reactions for which there exists no lay "theory" of influence (see Wilson & Brekke, 1994). Nonetheless, a well-designed study camouflaging the role that race of target person plays as an independent variable could give us some valuable answers.

IMPLICATIONS FOR ERADICATING PREJUDICE

Some clear prescriptions can be based on the foregoing review. One is not to count too much on a person's ability to control the impact of an automatically activated stereotype. Once it is activated, the horse has left the barn, and shutting the barn door at that point does no good. To be able to control these ef-

fects, the person first has to accept the idea of being influenced in ways of which he or she is not aware, and people generally have a difficult time accepting this possibility. The research described above shows that control is especially unlikely for nonobvious manifestations of prejudice, and that stereotypes leak out into nonverbal behavior and evaluations even for those who consciously endorse nonprejudicial values and beliefs.

Should a person accept the possibility of unconscious influence, then he or she must have the motivation and ability to control it. By "ability" here is meant such things as having enough time and attention to engage in the effortful, "hard" choice of individuating information processing. More than that, an individual needs to have a relatively accurate "theory" about the nature of the unfelt influence in order to correct for it, but generally these lay theories are faulty (see Wilson & Brekke, 1994). The odds that all of these necessary conditions will be met in a given situation therefore become vanishingly small.

Can People Break the Habit?

If a "cure" for automatic stereotyping through conscious correction processes is not possible, what are we left with? Clearly, one implication is that efforts would be better spent in "prevention"—that is, finding a way to stop stereotypes from being activated automatically in the first place. Lepore and Brown (1997) and Fazio et al. (1995) have shown that automatic stereotype activation does not occur for everyone, despite a stereotype's permeation of a culture. Although all individuals appear to possess knowledge of the stereotype, there may be individual differences in whether that stereotype is activated upon activation of the group representation. Lepore and Brown (1997) showed this by distinguishing between the representation of the social group (the individual's own knowledge base about the group) and the cultural stereotype associated with that group (see also Pratto & Bargh, 1991). The group representation is automatically activated by distinguishing group features (e.g., skin color), but the group representation may or may not be automatically associated with the group stereotype.

Lepore and Brown (1997) argued that Devine (1989, Experiment 2) showed automatic stereotype activation for all participants, regardless of level of overt racism, because the priming stimuli used were directly relevant to the stereotype (and therefore bypassed the group representation). In an experiment in which the priming stimuli were relevant only to the group representation and not to the stereotype, Lepore and Brown (1997) found automatic stereotype activation to be a function of a participant's degree of overt racism. This result is reminiscent of Gilbert and Hixon's (1991) and Macrae et al.'s (1997) findings of no effect of stereotypes under memory load, despite participants' later ability to recall the race of the target person—except that there was no manipulation of cognitive load in Lepore and Brown's (1997) study. Such dissociations between group representation and stereotype activation suggest that it is possible for individuals not to automatically apply stereotypic conceptions that they do possess, by virtue of the pervasiveness of the beliefs within the culture.

Yet the fact of individual differences in the application of a cultural stereotype is not in itself evidence that stereotypes can be controlled once they become so strong as to be automatically applied. The individual differences in application may exist because the automatic pathway or association has never been developed by some people in the first place. Though this is potentially good news about the degree of pervasiveness of stereotypes, it does not directly address the theme of the present chapter, which is the degree to which automatic stereotype effects can be controlled once they are in place.

Several studies do bear directly on this question. Monteith (1993) provided a potential method for breaking or retraining the stereotype habit. After rating the job suitability of a gay job applicant, participants were informed that their ratings were lower than those given by participants who read and evaluated the identical application, except that the applicant was described as heterosexual. Thus admonished for the prejudice they had shown, those participants who had earlier expressed values for being nonprejudicial reacted to this evident discrepancy between their behavior and their self-concept with an increased effort to be nonprejudicial in subsequent parts of the experimental session (see

also Brunstein & Gollwitzer, 1996; Wicklund & Gollwitzer, 1982). Those who did not express such values were not affected by the apparent act of prejudice they had committed.

Therefore, if people can be made aware of committing acts of prejudice, their motivation to be egalitarian (or not to be) may well come into play. But will this motivated engagement in nonstereotypic thought about the target group have any effect on changing the automaticity of the stereotype activation? A recent set of studies by Kawakami (1997) provides some evidence on this score. She found that the link between a group representation and a stereotype could be broken, at least temporarily, by retraining a different set of automatic beliefs about the group. Across a series of 240 trials, participants were instructed to respond by pressing a key labeled "No" whenever a photograph of a skinhead (or elderly person) was presented along with a stereotypic term associated with that group, and to respond "Yes" whenever a term not associated with that group was presented with the photograph. This associative training of different content to the skinhead or elderly group representation was successful in eliminating automatic stereotype activation, as measured by a subsequent Stroop task: The usual effect of greater response interference by stereotype-consistent than by stereotype-inconsistent stimuli was eliminated. Kawakami's procedure of training a different response to the stereotyped group is reminiscent of the approach taken in the cognitive therapy of emotional disorders (Beck, 1976; Beck, Rush, Shaw, & Emery, 1979).

Can an Automatic "Good" Defeat an Automatic "Evil"?

It is too early to tell, but perhaps the greatest hope for stereotype control at the intraindividual level lies in the battle between two different forms of preconscious, automatic processing. If Kawakami's experimental situation can produce at least temporary inhibition or reduction of the automatic stereotyping effect, then perhaps a well-intentioned individual, motivated to be nonprejudicial and egalitarian, can develop chronic inhibition/reduction of the effect over time. That is, repeatedly pursuing egalitarian goals and thoughts when interacting with members of stereotyped groups may automate that goal (i.e., may make it into an "auto-motive"; Bargh, 1990, 1997), so that the goal operates autonomously when one interacts with such individuals, without needing to be consciously chosen and guided each time. There is already evidence that information-processing as well as behavioral goals can be activated nonconsciously and then operate to produce the same outcomes as when those goals are selected consciously and intentionally (see Bargh, 1997; Bargh & Gollwitzer, 1994; Chartrand & Bargh, 1996).

This reasoning leads to the interesting possibility that both the group stereotype *and* the motivation to be egalitarian can be automatically activated at the same time and in parallel, by the mere features of the stereotyped-group member. How will these two forces interact? Will the automatically activated goal to be fair stifle and defeat the automatic stereotype activation, all in a matter of a few hundred milliseconds or so, before conscious processing has a chance even to notice (if ever) what is going on?

Such a possibility was first raised by Gollwitzer and Moskowitz (1996), and evidence is starting to accrue in its support (Moskowitz, Wasel, Gollwitzer, & Schaal, 1998). This idea of an immediate automatic *inhibition* of stereotypes is also in harmony with the recent theoretical model of Bodenhausen and Macrae (1998), which calls for both automatic facilitative and inhibitional responses in the processing of information about a stereotyped-group member. Indeed, Stangor, Thompson, and Ford (1998) have explicitly linked the idea of an egalitarian "auto-motive" to the Bodenhausen–Macrae model.

If all this sounds too good to be true, well, it may be. How, for example, does the egalitarian motive or goal become automated if not by the individual's chronically pursuing it over time, consciously and intentionally? But doing so, as has been argued above, requires the awareness of possible (nonconscious) bias; knowledge of the effect of the bias on judgments (many of them quite subtle, implicit, and tacit); ability to engage in the effortful processing at the time; and the good intention to be egalitarian—all of which are problematic conditions in real life. Nevertheless, the good news is that if one can get one's egalitarian motivation to the automatic state,

it may then routinely win out over the automatic stereotype—a case, if you will, of fighting automatic fire with automatic fire.

CONCLUSIONS

One danger in any critical analysis of a research area is that it can be mistaken for criticism of the goals of the research itself. There is no question that the research described in this chapter has been driven by sincere and laudable goals to discover ways to reduce and ameliorate the social scourge of stereotyping and prejudice, and by optimism that such ways can be found. Such attempts should of course continue. But although the goals and purpose of the research are admirable and optimistic, we must nonetheless accept the findings of that research at face value, and not allow our optimism and hopefulness to color our interpretation of the evidence. In my opinion, the evidence to date concerning people's realistic chances of controlling the influence of their automatically activated stereotypes weighs in heavily on the negative side. The lesson to be learned from the tales of Frankenstein and King Kong is that monsters, once on the loose, cannot be controlled by chains.

Once a stereotype is so entrenched that it becomes activated automatically, there is really little that can be done to control its influence. Even under the scrutiny of a psychology experiment, in which most of the time not only the presented prime–target pairs but the participants' conscious expectations for the targets were counterstereotypic, participants produced the same pattern of means as when no such conscious expectancy existed. Even when a person has egalitarian values and motives, his or her facial expression and tone of voice and reaction to the behavior of a stereotyped-group member can often be a first strike that produces stereotype-confirming behavior and so perpetuates the stereotype. Realistically, there is little that will be done about such nonconscious effects in the real world—mainly because, in the words of Hall of Fame baseball pitcher Bob Feller, "You can't hit what you can't see," and because on those occasions when corrections are attempted, they are often (if not usually) guided by faulty lay theories as to the nature of the bias (Wilson & Brekke, 1994).

Hoping to stop the cognitive monster by trying to control already activated stereotypes is like mowing dandelions; they just sprout back up again. As with dandelions, the only way to kill stereotype effects is to pull them up by their roots—by removing their capability for automatic activation, or (better still) by preventing the seeds from taking root in the first place, through eradication of the cultural stereotype itself. Steele (1997) makes a similar argument in the context of African Americans' rocky road to holding an academic identity. In theory, he points out, one could perform intrapsychic interventions on each affected student in an attempt to counter the panoply of forces that can push them off the academic track, but this is not a realistic strategy. The real solution, Steele concludes, is to eliminate in the first place the culturally shared and transmitted assumption that blacks "can't cut it" academically. Until we find a way to kill the dandelions, reports of the death of the cognitive monster will be greatly exaggerated.

ACKNOWLEDGMENTS

Preparation of this chapter was supported in part by Grant No. SBR-9409448 from the National Science Foundation. Portions were previously presented as the keynote address to the annual meeting of the Netherlands Social Psychology Association, Leiden, December 1996. I am indebted to Mahzarin Banaji, Irene Blair, Galen Bodenhausen, Marilynn Brewer, Patricia Devine, Ap Dijksterhuis, Kimberly Duckworth, Ziva Kunda, Lorella Lepore, Neil Macrae, Bernadette Park, Denise Sekaquaptewa, Charles Stangor, Erik Thompson, and the editors for extensive feedback on an earlier version of the chapter; their generous help should not be taken as implying endorsement of any of the opinions expressed herein.

NOTES

1. Many (see below) have taken this conclusion to express the authors' belief that expectancies can prevent the initial automatic activation of the stereotype, but I. Blair (personal communication, November 11, 1997) considers that the successful elimination of stereotype effects on reaction times in "the 2000 ms SOA condition showed that participants could eliminate stereotype activation with

an expectancy strategy." In other words, the conclusion at the end of the Blair and Banaji (1996) article quoted here apparently referred to the 2,000-ms and not the 350-ms SOA condition results. Perhaps it is just a matter of personal taste, but I consider Devine's'(1989) distinction between the activation of a stereotype and its subsequent use in judgment to be a useful one. That is, 2 seconds is enough time to strategically control the effect of an automatically activated stereotype (Neely, 1977); therefore, a lack of stereotype effects in that condition does not demonstrate the elimination of the stereotype's automatic activation.

2. The experiment × prime × target interaction was in fact not reliable ($F < 1$; Blair, personal communication, November 11, 1997). Although it is my belief that this within-subjects test of the effect of expectancy (none vs. counterstereotypic) on automatic stereotype activation (i.e., the prime x target interaction) is more appropriate than comparing the significance levels of the two separate prime x target interactions (see Keppel, 1973), reasonable people can disagree on this point.

3. Importantly, this is not the position of Blair and Banaji concerning their results: "The position we took in our article and continue to support is that the priming effect was not eliminated under high cognitive constraints" (Blair, personal communication, November 8, 1997).

4. In the great majority of experiments, overt and explicit measures of stereotypic beliefs such as the Modern Racism Scale or the Attitudes Towards Women scale are uncorrelated with these implicit tendencies (for an exception, see Wittenbrink, Judd, & Park, 1997). In an insightful analysis of the role of encoding processes in the persistence of stereotypes, von Hippel et al. (1995) draw a distinction between the content of stereotypes and the way those stereotypes are used to process information about a person, akin to Jacoby's (e.g., Jacoby & Kelley, 1987) distinction between memory as an object and memory as a tool. It is possible, in other words, for the tendency to *process* person information in a way that confirms stereotypes to be somewhat dissociated or independent from the tendency to hold stereotypic *beliefs* (see also Dunning & Sherman, 1997; Trope & Alfieri, 1997).

REFERENCES

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68*, 181–198.

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology, 65*, 272–281.

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: Foundations of social behavior* (Vol. 2, pp. 93–130). New York: Guilford Press.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.

Bargh, J. A., & Chartrand, T. L. (in press). Studying the mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), *Research methods in the social sciences*. New York: Cambridge University Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230–244.

Bargh, J. A., & Gollwitzer, P. M. (1994). Environmental control over goal-directed action. In W. D. Spaulding (Ed.), *Nebraska Symposium on Motivation: Vol. 41. Integrative views of motivation, cognition, and emotion* (pp. 71–124). Lincoln: University of Nebraska Press.

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall–judgment link: The case of information overload. *Journal of Personality and Social Psychology, 49*, 1129–1146.

Bargh, J. A., & Tota, M. E. (1988). Context-dependent automatic processing in depression: Accessibility of negative constructs with regard to self but not others. *Journal of Personality and Social Psychology, 54*, 925–939.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Berkowitz, L. (1984). Some effects of thought on anti- and prosocial influences of media events: A cognitive-neoassociation analysis. *Psychological Bulletin, 95*, 410–427.

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

Blair, I., & Banaji, M. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142–1163.

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 11, pp. 1–52). Mahwah, NJ: Erlbaum.

Brewer, M. B. (1988). 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.

Brunstein, J. C., & Gollwitzer, P. M. (1996). Effects of failure on subsequent performance: The importance of self-defining goals. *Journal of Personality and Social Psychology, 70*, 395–407.

Carver, C. S., Ganellen, R. J., Froming, W. J., & Chambers, W. (1983). Modeling: An analysis in terms of category accessibility. *Journal of Experimental Social Psychology, 19*, 403–421.

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

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460–473.

Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of social information processing goals: Nonconscious priming reproduces effects of explicit conscious instructions. *Journal of Personality and Social Psychology, 71*, 464–478.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*, 541–560.

Darley, J. M., & Fazio, R. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35*, 867–881.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*, 20–33.

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, NJ: Erlbaum.

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelations among components and gender label. *Journal of Personality and Social Psychology, 46*, 991–1004.

DePaolo, B. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*, 203–243.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 680–690.

Dijksterhuis, A., & van Knippenberg, A. (1998). Automatic social behavior or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology, 74*, 865–877.

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In N. Macrae, M. Hewstone, & C. Stangor (Eds.), *Foundations of stereotypes and stereotyping* (pp. 276–319). New York: Guilford Press.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22–37.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.

Dunning, D., & Sherman, D. A. (1997). Stereotypes and tacit inference. *Journal of Personality and Social Psychology, 73*, 459–471.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88–106.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33*, 451–470.

Fazio, R. H., Jackson, J., Dunton, B., & Williams, C. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*, 889–906.

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 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.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

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.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54*, 733–739.

Gollwitzer, P. M., & Moskowitz, G. (1996). Goal effects on thought and behavior. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 361–399). New York: Guilford Press.

Hamilton, D. L. (1979). A cognitive-attributional analysis of stereotyping. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 53–85). New York: Academic Press.

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues, 46*(2), 35–60.

Hilton, J. L., & Darley, J. M. (1991). The effects of interaction goals on person perception. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 235–267). 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.

James, W. (1890). *Principles of psychology* (Vol. 2). New York: Holt.

Jones, E. E. (1990). *Interpersonal perception.* New York: Freeman.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220–266). New York: Academic Press.

Jones, E. E., & Nisbett, R. E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution:*

Perceiving the causes of behavior (pp. 79–94). Morristown, NJ: General Learning Press.

Jones, E. E., & Thibaut, J. W. (1958). Interaction goals as bases of inference in interpersonal perception. In R. Taguiri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 151–178). Stanford, CA: Stanford University Press.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review, 93*, 429–445.

Jussim, L. (1990). Social reality and social problems: The role of expectancies. *Journal of Social Issues, 46*(2), 9–34.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Kawakami, K. (1997). *Kicking the habit: The effects of suppression training on stereotype activation.* Unpublished manuscript, University of Nijmegen, The Netherlands.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192–241). Lincoln: University of Nebraska Press.

Keppel, G. (1973). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Koestler, A. (1967). *The ghost in the machine.* London: Hutchinson.

Langer, E. J. (1978). Rethinking the role of thought in social interaction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2, pp. 36–58). Hillsdale, NJ: Erlbaum.

Langer, E. J., & Abelson, R. P. (1972). The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology, 24*, 26–32.

Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology, 36*, 635–642.

Lee, Y.-T., Jussim, L. J., & McCauley, C. R. (Eds.). (1995). *Stereotype accuracy: Toward appreciating group differences.* Washington, DC: American Psychological Association.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72*, 275–287.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*, 981–993.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67*, 808–817.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology, 33*, 471–489.

Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology, 72*, 791–809.

Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review, 8*, 193–210.

Monteith, M. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65*, 469–485.

Monteith, M. J., Sherman, J. W., & Devine, P. G. (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review, 2*, 63–82.

Moskowitz, G. B., Wasel, W., Gollwitzer, P. M., & Schaal, B. (1998). *Volitional control of stereotype activation through chronic egalitarian goals.* Unpublished manuscript, Princeton University.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

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.

Neuberg, S. L. (1994). Expectancy-confirmation processes in stereotype-tinged social encounters: The moderating role of social goals. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario Symposium* (Vol. 7, pp. 103–130). Hillsdale, NJ: Erlbaum.

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology, 53*, 431–444.

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–238). New York: Guilford Press.

Paulhus, D. L., Graf, P., & Van Selst, M. (1989). Attentional load increases the positivity of self-presentation. *Social Cognition, 7*, 389–400.

Piaget, J. (1948). *La formation du symbole chez l'enfant.* Paris: Delachaux & Niestlé.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Pratto, F., & Bargh, J. A. (1991). Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology, 27*, 26–47.

Pratto, F., & John, O. P. (1993). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380–391.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action* (pp. 167–201). Berlin: Springer-Verlag.

Prinz, W. (1997). Explaining voluntary action: The role of mental content. In M. Carrier & P. Machamer (Eds.), *Mindscapes: Philosophy, science, and the mind* (pp. 153–175). Konstanz, Germany: Universitaetsverlag.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*, 557–568.

Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 248–305). New York: Academic Press.

Snyder, M., & Swann, W. B., Jr. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology, 14*, 148–162.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*, 656–666.

Sorrentino, R. M., & Higgins, E. T. (Eds.) (1986). *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 1). New York: Guilford Press.

Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Dunn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin, 24*, 1139–1152.

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.

Stangor, C., Thompson, E. P., & Ford, T. E. (1998). An inhibited model of stereotype inhibition. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 11, pp. 193–210). Mahwah, NJ: Erlbaum.

Steele, C. M. (1997, October). The rocky road to an academic identity for an African-American. In B. Major (Chair), *The psychological consequences of stigma*. Symposium presented at the annual meeting of the Society for Experimental Social Psychology, Toronto.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 249–288). New York: Academic Press.

Trope, Y., & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology, 73*, 662–674.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology, 66*, 490–501.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 27, pp. 177–254). San Diego, CA: Academic Press.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology, 33*, 490–509.

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

Wegner, D. M., Erber, R., Bowman, R., & Shelton, J. N. (1997). *On trying not to be sexist*. Unpublished manuscript, University of Virginia.

Weiner, B. (1974). *Achievement motivation and attribution theory*. Morristown, NJ: General Learning Press.

Wicklund, R. A., & Gollwitzer, P. M. (1982). *Symbolic self-completion*. Hillsdale, NJ: Erlbaum.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117–142.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47*, 237–252.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262–274.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10*, 109–120.