7

Compensatory Automaticity: Unconscious Volition Is Not an Oxymoron

Jack Glaser and John F. Kihlstrom

After nearly three decades of research on automaticity and construct activation, it is increasingly clear that much of human mental life operates without awareness or intent. In the literature of cognitive psychology, automatic processes are held to be inevitably evoked by the presence of a relevant environmental stimulus; once triggered, their execution proceeds rapidly, effortlessly, and incorrigibly to completion, leaving no traces accessible to conscious recollection. The concept of automaticity, long central in cognitive psychology (e.g., Schneider & Shiffrin, 1977), has come to occupy an important place in social psychology as well (e.g., Bargh, 1994, 1997; Kihlstrom, 1996, 1999; Wegner & Bargh, 1998). At the same time, some of the features canonically attributed to automaticity have been called into question (e.g., Logan, 1997; Shiffrin, 1997), and newer research reveals that automatic responses, while unintended, may not be inevitable (e.g., Glaser, 2003; Glaser & Banaji, 1999; Moskowitz, 2001; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). It appears that unconscious vigilance for bias can lead to corrective processes that also operate without conscious awareness or intent. Given these developments, we contend that the unconscious, in addition to being a passive categorizer, evaluator, and semantic processor, has processing goals (e.g., accuracy, egalitarianism) of its own, can be vigilant for threats to the attainment of these goals, and will proactively compensate for such threats. One might call this "compensatory automaticity"; strategic yet nonconscious compensations for unintended thoughts, feelings, or behaviors. For some, this will pose a paradox because automaticity has been equated with lack of control or intent. We believe, however, that it is important at this stage to move beyond that conflation and to entertain the possibility that intention operates at multiple levels of consciousness. There can be nonconscious intentions (e.g., goals) that, when the potential for their imminent frustration becomes evident,

automatic compensatory processes will promote and protect. All of this can operate outside of conscious awareness and control, thereby rendering the unconscious relatively "complete," although not, of course, the entirety of mental life.

The Unconscious and the Triarchic Mind

Arguments that the unconscious is complex and wide ranging have been made compellingly, and with increasing empirical support, for several decades. It is probably not a coincidence that the person who wrote definitively about the mental trilogy of affect, cognition, and conation (Hilgard, 1980) also conducted groundbreaking work on the complexity of the unconscious, even arguing that the unconscious is comprised of multiple levels of awareness (Hilgard, 1977). Hilgard's (1977) research revealed that percepts not available to conscious recollection could nevertheless guide a person in a nonconscious, hypnotic state. Most important to the present thesis, Hilgard provided evidence of "divided consciousness" and specifically the concept of the *hidden observer*, wherein a hypnotized subject was able to perceive and respond to auditory instruction even while otherwise "deaf" by hypnotic suggestion and accordingly nonresponsive to startling noises.

Partly inspired by Hilgard's (1977) insights, Kihlstrom (1987) made the case that the cognitive unconscious was expansive and sophisticated. At the time that Kihlstrom first argued that a modern, cognitive conception of the unconscious was likely to be wide ranging, evidence was concentrated on the cognitive component of the mind, with ample demonstrations of implicit memory and automatic cognition (e.g., Neely, 1977; Schacter, 1987; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Unconscious affect and motivation were, for the most part, theoretical constructs with smatterings of empirical support. But the early trickle of research on unconscious affect (e.g., Kunst-Wilson & Zajonc, 1980; Zajonc, 1980), gave way to a steady flow in the late 1980s with Fazio's (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) demonstration that evaluative responses can occur automatically, and Bargh's (Bargh, Chaiken, Govender, & Pratto, 1992; Bargh, Chaiken, Raymond, & Hymes, 1996) extension that this automatic evaluation was a highly general phenomenon. Greenwald (Greenwald, Draine, & Abrams, 1996; Greenwald, Klinger, & Liu, 1989) as well as Niedenthal (1990) and Murphy and Zajonc (1993) made a convincing case for unconscious affective responses, demonstrating that they occurred even for stimuli that were subliminal and therefore not consciously perceived (see Kihlstrom, Mulvaney, Tobias, & Tobis, 2000, for a review).

In recent years, unconscious goals and motives have also appeared on the radar screen (e.g., McClelland, Koestner, & Weinberger, 1989). Bargh and his

colleagues (e.g., Bargh, 1996, 1997; Bargh & Barndollar, 1996; Bargh & Chartrand, 1999; Bargh & Ferguson, 2000; Chartrand & Bargh, 1996), in particular, have brought their research to bear to make the compelling argument that all major mental processes, including motivation (see also Gollwitzer, 1999), can operate automatically. Most recently, Bargh, Gollwitzer, Lee-Chai, Barndollar, and Trötschel (2001) have demonstrated that nonconscious goal pursuits possess properties similar to those deemed fundamental to conscious motivation, specifically, vigorous action toward goal satisfaction, persistence, and resumption after disruption. Although to date less comprehensive than the research on unconscious cognition and affect, their work strongly indicates that goals and behaviors can also be activated automatically and will be pursued nonconsciously.

If, as theorists (e.g., Bargh, 1997; Hilgard, 1977; Kihlstrom, 1987, 1999) have suggested for decades, nonconscious mental life is sophisticated and comprehensive, we must entertain the possibility that in addition to the trilogy of affect, cognition, and conation being represented there, another critical aspect of human psychology, self-awareness and metacognition, may also reside outside of consciousness. The very terms *self-awareness* and *metacognition*, raised in the context of the unconscious, will likely raise consternation among those who, for good reason, equate awareness with consciousness. The very absence of a term like *awareness* or *introspection* in the lexicon of the study of the unconscious is a testament to the prevalence of the belief that nonconscious processes operate outside any monitoring capability. Hilgard (1977) questioned this rigidity with his concept of the hidden observer, but to date the idea is still counterintuitive, and we are left to ponder the oxymoronic nature of the proposition. Nevertheless, evidence for the apparent automatic control, or attempts at control, of automatic processes (e.g., Glaser, 2003; Glaser & Banaji, 1999; Moskowitz, 2001; Moskowitz et al., 1999), leads us to theorize that people are capable of nonconscious vigilance for nonconscious bias and further of automatic compensatory processes that are triggered outside of conscious awareness or control. Evidence for this comes from developments in the priming literature, specifically with regard to assimilation and contrast in construct activation and stereotype suppression effects. These findings are discussed as they relate to compensatory automaticity and unconscious volition.

The Role of Awareness in Construct Activation and Inhibition

Research on construct activation (Higgins, 1996; Higgins, Rholes, & Jones, 1977) has been guided to some extent by the assumption that unconscious/automatic processes lead to assimilation effects while more deliberate pro-

cesses, such as correction, lead to contrast. Lombardi, Higgins, and Bargh (1987), for example, reported that subjects were likely to judge an ambiguous target person in a manner consistent with a construct (e.g., "stubborn" or "persistent") that was made accessible (an assimilation effect), if the priming event presenting the construct was not explicitly remembered. If, on the other hand, the priming event was remembered at all, the target person was judged in a manner inconsistent with the construct (a contrast effect). Lombardi et al. attributed this difference to distinctions between automatic and controlled processing—automatic processes accounting for assimilation, and controlled processes engendering contrast.

Support for the role of awareness and deliberation in determining assimilation versus contrast also comes from a study by Martin, Seta, and Crelia (1990). Theorizing that contrast effects result from an overgeneralization in attempts to counteract the biasing influence of priming stimuli (Martin, 1986), Martin et al. hypothesized that this would be most likely to occur when one has the cognitive resources to make such an adjustment, but not when such resources are depleted (a limitation to which automatic processes are, by the way, immune). Accordingly, they found that distracted subjects showed assimilation toward primed concepts, while those who were not distracted showed contrast. Martin et al. corroborated this finding by reporting similar effects for subjects who were low and high in need for cognition, respectively. Similarly, Newman and Uleman (1990) found that contrast effects occurred when primes were blatant, and Strack, Schwarz, Bless, Kübler, and Wänke (1993) reported that subjects who were reminded of a priming procedure showed contrast effects, whereas those who were not reminded exhibited assimilation.

It appears from the research described that the salience of the prime, to the extent that salience is related to awareness, may determine whether priming is assimilative or contrastive. One determinant of salience is extremity. Indeed, Herr, Sherman, and Fazio (1983; see also Herr, 1986), studied the effects of extremely large or small (and, in another experiment, ferocious and meek) animal primes on judgments of the size (or ferocity) of target animals. They found that extreme primes yielded contrastive judgments, whereas moderate primes led to assimilative judgments of an ambiguous target. Taken together, the results of such experiments suggest that as the priming stimulus, or at least its potential to influence the judgment of the target, becomes more salient, and therefore more accessible to conscious awareness, contrast effects in judgments are more likely to result.

The effect of prime salience on assimilation and contrast effects may be moderated by the motivation to be accurate, which, not surprisingly, influences how judgments are made (Neuberg & Fiske, 1987). In fact, in one of their experiments, Martin et al. (1990) found that subjects who believed that

their judgments would be averaged with those of others made assimilative responses, whereas those who believed that their judgments would be evaluated individually exhibited contrast effects, most likely attempting to compensate, but regrettably overcompensating, for the biasing effect of the prime. This result suggests, consistent with other research on accountability and accuracy (e.g., Lerner & Tetlock, 1999), that anticipated accountability motivated subjects to be vigilant and adjust for the biasing influence of the primes. Other studies have more directly manipulated accuracy motivation, finding that it attenuates assimilation effects (Ford & Kruglanski, 1995; Thompson, Roman, Moskowitz, Chaiken, & Bargh, 1994). Stapel, Koomen, and Zeelenberg (1998) have drawn upon such findings to make the case that accuracy motivation leads to more careful processing of the target, thereby attenuating assimilation effects, but that a correction strategy is required to bring about contrast effects. Of most relevance to the present thesis, Stapel, Martin, and Schwarz (1998) have shown that the corrections that engender contrast effects are made spontaneously when biasing information is blatant, but not when it is subtle. Such corrections would almost certainly be dependent on an accuracy goal (in the absence of such a goal, why would one correct for biasing information?).

Correction has also been posited as a determinant of both contrast and assimilation effects by Petty and Wegener (1993; Wegener & Petty, 1995), who provide evidence that people's lay theories about assimilation and contrast predict the direction of their corrections for the potential biasing effects of contextual stimuli. Specifically, in Petty and Wegener's studies, subjects who expected assimilation effects corrected away from the direction of the contextual information (i.e., the priming stimulus) while those who expected contrast effects corrected toward the contextual information. These findings convey the complexity of perceivers' strategies when they attempt to mitigate the effects of judgmental biases.

In sum, a substantial body of research indicates that while contextual information can bias a response to supposedly unrelated stimuli, at times the result shows a contrastive pattern. It appears that contrast effects occur especially when the perceiver is aware of the potential biasing influence of the prime, perhaps as a result of its salience, and/or when the perceiver has the cognitive resources and motivation to recognize or remember the prime. It also appears to be the case that, while assimilation effects occur spontaneously, contrast effects are more likely the result of an active correction. To date, studies of contrast effects have been restricted to conditions under which judgments are relatively controlled and deliberate (e.g., rating a target stimulus on a scale). Perhaps this is the case because of the assumption that contrast effects result from deliberate processes (e.g., Lombardi et al., 1987; Wilson & Brekke, 1994). Given that assumption, one would expect to see only

assimilation effects in automatic processing. However, the results of recent research provide clear evidence, apparently the first, of what might be called contrast effects under conditions where controlled processing is precluded, thus suggesting an automatic correction process. The evidence for automatic correction calls into question prevailing conceptions of unconscious processes as passive and reactive, and therefore warrants further examination.

Reverse Priming: Automatic Correction for Automatic Evaluation

Fazio et al. (1986) first demonstrated automatic evaluation (aka automatic attitude activation, affective priming) using a semantic priming procedure in which research subjects categorized target adjectives as being positive or negative. The presentation of each target was preceded by the presentation of another word (a prime) that was either positive or negative (evaluatively neutral nonword letter strings were also used as primes). Responses were faster when the prime and target in a given pair were evaluatively congruent. This necessarily indicated that the primes had been evaluated too. Prior research on automaticity had shown that when the time from the onset of the prime to the onset of the target, otherwise known as the stimulus onset asynchrony (SOA), was brief (e.g., under 500 ms), the effects of controlled processes on priming effects could be precluded (Neely, 1977). Fazio et al. (1986), employing a 300 ms SOA, made a strong case that the evaluations of the primes were automatic and unintended. Furthermore, comparisons with response times when the prime was "neutral" revealed that subjects were evaluating both the positive and negative primes, which were having facilitative and inhibitory effects on responses to evaluatively congruent and incongruent targets, respectively. Bargh and colleagues (1992) expanded on this finding, demonstrating that automatic evaluation is a very general phenomenon that occurs even for objects toward which attitudes are weak. Further, Bargh et al. (1996) employed more subtle techniques, including the mere pronunciation, rather than evaluative categorization, of the targets, to make a compelling case that automatic evaluation occurs very spontaneously and relatively unconditionally (see also Hermans, de Houwer, & Eelens, 1994). In the absence of cues that the experiment was about evaluating the stimuli, and the consequent elimination of corresponding demand characteristics, Bargh et al. (1996) obtained robust automatic evaluation effects (i.e., faster pronunciation of targets preceded by evaluatively congruent primes) for both weak and strong attitude objects.

Inspired by the effectiveness of the Bargh et al. (1996) procedure, Glaser and Banaji (1999) set out to adapt the paradigm to measure implicit race

prejudice unobtrusively. Researchers had previously employed similar procedures to measure racial bias (e.g., Dovidio, Evans, & Tyler, 1986; Fazio, Jackson, Dunton, & Williams, 1995), but such procedures, involving judgments about the valence or category of the target, while effectively nonreactive, were relatively obtrusive. With the use of a pronunciation, rather than an evaluation or a categorization task, subjects would be unlikely to presume the procedure was designed to assess their racial attitudes. Glaser and Banaji (1999) expanded the stimulus categories beyond those of Bargh et al. (1996) to include words and names that are stereotypically associated with African American and European American culture (e.g., basketball, homeboy, Cosby; and golf, hippies, Letterman). In addition, they used a list of extremely positive and negative but race-neutral words (e.g., kindness, puppy, accident, tumor) termed "generic" stimuli, and another set of race-neutral positive and negative words that were all related to food (e.g., fudge, soup, beets, meatloaf).[1] Pairing these categories of words in all possible types of combinations allowed for a test for automatic evaluation (race-neutral primes with race-neutral targets), race categorization (race primes with race targets), and race prejudice (race-neutral primes with race targets, or vice versa), and further enhanced the unobtrusiveness of the measure by lending the appearance of true arbitrariness. These multiple possible outcomes within one paradigm also promised to bolster the internal validity of the findings. For example, if race-neutral (i.e., generic or food) primes and targets produced automatic evaluation effects (e.g., faster responding to positive-positive and negative-negative prime-target pairs), and race primes and targets produced race categorization effects (e.g., faster responding to Black-Black and White-White pairings) that would support the interpretation of an interaction of race-neutral primes with race targets (e.g., faster responding to negative-Black and positive-White than to negative-White and positive-Black prime-target pairings) as reflecting prejudice (i.e., an association between evaluations and groups).

Indeed, all three types of unconscious association were evident with both types of race-neutral stimuli (generic and food), and the effects tended to be large. However, when the generic words served as primes, the results were counterintuitive and perplexing. The effects were in the opposite direction of what had been predicted and of what had been obtained on trials where the food words served as primes. Specifically, on trials where the generic words served as primes, subjects were faster to respond to evaluatively incongruent targets than to the congruent targets. Similarly, regarding the test for automatic prejudice, with the food primes subjects were faster to pronounce Black-associated and White-associated targets when preceded by negative and positive primes, respectively, indicating the predicted pro-White or anti-Black bias. However, as with the automatic evaluation effect, the opposite

was true for generic primes, giving the impression of an anti-White/pro-Black bias that might have been interpreted as such had it not been for the similar pattern of results observed with the race-neutral stimuli in the automatic evaluation analysis.

Glaser and Banaji (1999) dubbed the effect "reverse priming" because the priming effect was the reverse of what one would expect. Importantly, these reverse priming effects were very large and highly statistically significant. Furthermore, careful inspection of the data revealed that virtually all subjects showed this pattern, and that the pattern of results was very consistent across trials, with subjects showing reverse priming with generic primes, and normal priming with food primes in the early, middle, and late phases of the task. The inferential statistical tests were, accordingly, very conclusive. Consequently, despite having predicted otherwise, Glaser and Banaji were not inclined to dismiss the finding as random.

Clearly, something interesting was happening in the automatic evaluation evidenced in this study, as a function of the type of prime. What was it about the food and generic words that led to such dramatically different patterns? One thing that was evident immediately was that the food and the generic words differed in evaluative extremity. Because food words are unlikely to be extremely negative (anything that is extremely aversive is probably not edible, or at least not considered a food), and because the negative and positive words were selected to be balanced on evaluative extremity, the food words tended to be only mildly valenced (averaging −1.0 for negative and +1.03 for positive food words on an 11-point evaluation scale from −5 to +5). In contrast, the generic words, selected to be unambiguously positive or negative, were fairly extreme in valence (−3.7 and +3.85). Evaluative extremity seemed a likely candidate for the critical difference between the food and generic words. Indeed, subsequent, post hoc analyses revealed that the most extremely valenced subset of the generic primes showed an even more exaggerated reverse priming effect, while the least extreme of them (still considerably more extreme than the food primes) showed a flat line—no priming effect at all.

The post hoc analyses were suggestive, but a priori replication was required to isolate prime extremity as a determinant of reverse priming. Accordingly, Glaser and Banaji replicated this experiment selecting new race-neutral stimuli that varied only in evaluative extremity, to replace the generic and food words. The results (see figure 7.1) matched those of the first experiment almost precisely, with mild (instead of food) primes leading to normal effects and extreme (instead of generic) primes yielding reversed effects (Glaser & Banaji, 1999, Experiments 2 and 3).

Nevertheless, the reverse priming effect was counterintuitive, and so Glaser and Banaji (1999) sought to identify other conditions that gave rise to it.

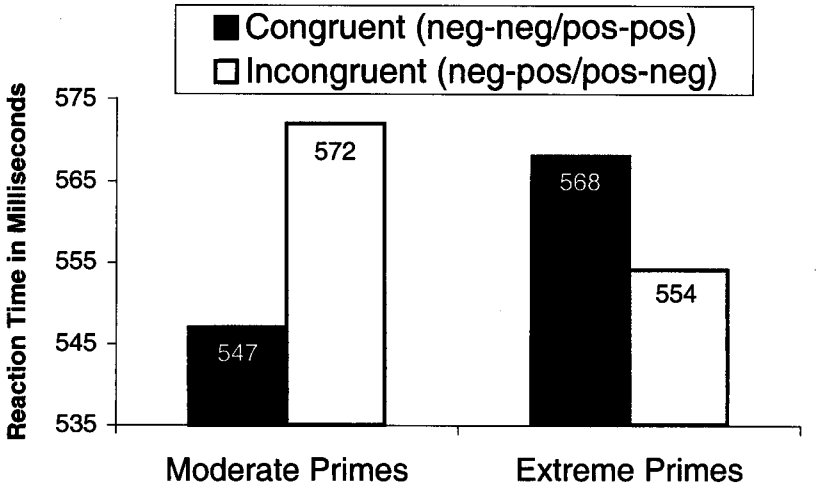

Figure 7.1 Automatic evaluation by prime extremity in Glaser and Banaji (1999), Experiment 3 (with racial stimuli present, but not in the trials represented by these data). Neg = negative; pos = positive.

Most prominently, it seemed possible that the presence of the racial stimuli in the procedure caused subjects to react in an unusual, perhaps acutely self-conscious manner, even on trials where there were no racial stimuli, when prime and target were race neutral, as with the data presented in figure 7.1. A series of three experiments (Glaser & Banaji, 1999, Experiments 4, 5, and 6) excluded racial stimuli from the procedure and also tested the effects of procedural variables, such as stimulus presentation durations and the presence of an orienting stimulus prior to each prime, in an effort to rule out obscure procedural variants as explanations for reverse priming. While replications of the experiment without any racial stimuli revealed that the racial stimuli did make a difference, it was not with regard to reverse priming. As figure 7.2 illustrates, whereas with racial stimuli present (see figure 7.1) normal priming was evident with mild primes and reverse priming with extreme primes, in the absence of racial stimuli, there was no longer any priming effect with the mild primes, but with the extreme primes the reverse priming effect was still evident. We can only speculate at this stage, but it seems plausible that the presence of the racial stimuli served to enhance the salience of the evaluative aspect of the stimuli, thereby enabling even mild primes to activate an associated attitude. Importantly, though, reverse priming (i.e., priming in the opposite direction of that predicted given the valence of the prime) persisted, as before, when the primes were evaluatively extreme, across various samples, stimulus sets, and procedures.

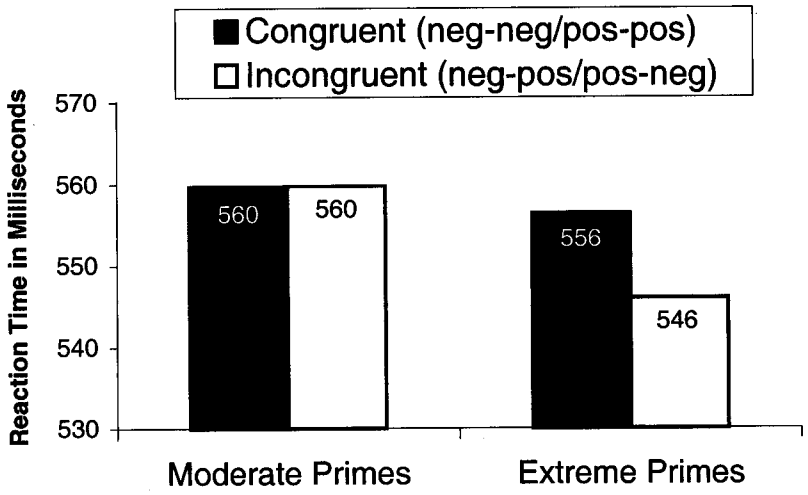

Figure 7.2 Automatic evaluation by prime extremity in Glaser and Banaji (1999), Experiment 4 (with no racial stimuli present in the experiment). Neg = negative; pos = positive.

Automatic Correction for Unintended Bias

We propose that the observed reverse priming effects reflect a correction that is instigated by the perceived potential of the peripheral prime to bias the response to the intended target. The task is to pronounce the target word correctly. In an orthographically deep (i.e., nonphonetic) language like English, this very likely involves processing the target's meaning, which includes, perhaps centrally, its evaluative valence (i.e., whether it is good or bad) (Osgood, Suci, & Tannenbaum, 1957; Tesser & Martin, 1996). When an extraneous stimulus (the prime) that is particularly obtrusive (by virtue of its evaluative extremity) appears, the perceiver is all the more likely to recognize its potential to bias the response (i.e., the identification of meaning and resultant correct pronunciation) to the intended target of judgment. If the strategy employed were simply to disregard or actively inhibit the prime, we would see no priming effect at all. However, the consistently obtained reverse priming effects indicate that an unconscious, but tactical correction is taking place in order to neutralize the threat to the accuracy of the intended response. As is the case with more deliberate judgments (e.g., Lombardi et al., 1987; Martin et al., 1990; Stapel et al., 1998), the compensatory process is excessive; an overcompensation. As a result, the extreme primes end up activating the evaluative associations that are opposite to that of their intrinsic meaning and consequently facilitate responses to words of opposite valence, and perhaps inhibit responses to words of the same valence.

Previous findings implicate accuracy motivation as a mediator of correction effects (e.g., Ford & Kruglanski, 1995; Martin et al., 1990; Stapel et al., 1998). The contribution of the reverse priming findings is the suggestion that accuracy motivation, manifested in an attempt to avoid bias, can operate automatically and outside of conscious awareness.

Correction or Comparison?

The effect of prime extremity, yielding reverse priming, is reminiscent of the findings of Herr et al. (1983), wherein extreme primes led to contrast effects. Such contrast effects are typically attributed to comparison-contrast, wherein the judgment of the target is contrasted away from a presumably irrelevant prime or comparison standard (Sherif & Hovland, 1961), and this could be the case with Herr et al. (1983) as well. Thoughts of extremely large (small) animals may have made target animals of an ambiguous size seem small (large) by comparison. Such an explanation is not, however, likely for Glaser and Banaji's (1999) reverse priming results, for several reasons. First, we must consider that the dependent variable was reaction time to pronounce the target word, not a qualitative judgment (e.g., evaluative rating, or size or weight estimation) of the target as employed in typical studies of comparison-contrast. In order for comparison-contrast to explain the reverse priming results in a reaction time paradigm, we would have to allow that, after seeing an extreme prime, the subject judges the target in the opposite direction of the prime (via comparison, e.g., "Compared to a tumor, war isn't such a bad thing"), and then the automatic activation process has to start all over (with the newly contrasted target), yielding differential activation. In other words, in a two-stage priming process, the prime and target would both be perceived, and then the judgment of the target would have to be radically adjusted for its comparative relation to the prime, such that it is now perceived to have the opposite valence of what it normatively has. Subsequent to that adjustment, the response to the newly adjusted (in terms of evaluation) target would be either facilitated or inhibited by the prime. Not only is this explanation awkward and unparsimonious, it also predicts that all targets, or at least all mildly valenced targets, would be contrasted by extreme primes (those that are similar and different in valence alike). This would yield relatively slow responses to all targets following extreme primes, not just those targets of similar valence to the prime. In all the Glaser and Banaji (1999) experiments, priming with extreme primes was characterized by slower responses to evaluatively congruent (e.g., positive-positive and negative-negative) prime-target pairs.

Perhaps more profoundly, the comparison-contrast explanation fits poorly in this context because the reverse priming effects, while limited to trials with

evaluatively extreme primes, were obtained for trials with evaluatively extreme as well as moderate targets. The primes and targets were drawn from the same pool of words (although the same word was never presented with itself as prime and target in the same or even a proximal trial). It is unlikely that the comparison of extreme targets to equally extreme primes would yield comparison contrast, since, by virtue of their equality, there is no basis for contrasting one with the other.

Finally, another comparison with Herr et al.'s (1983) findings is illuminating. They found assimilation effects only when the targets were ambiguous (i.e., fictitious animals). When targets were unambiguous (real animals), contrast effects occurred with moderate and extreme primes alike. If the comparison-contrast explanation typically applied to Herr et al.'s results applies to the reverse priming findings, we would expect to see contrast effects when the prime is moderate and the target is extreme (i.e., unambiguous). Again, this is not the case, and correction is therefore a more plausible explanation. In fact, it is difficult to envision how comparison-contrast could explain any priming effect with a reaction-time-dependent variable. It is for this reason that we have been careful to adhere to the term *reverse priming* and to endorse a correction explanation, so as to avoid confusion with comparison-contrast. While there are some striking parallels to research on assimilation and contrast, and specifically comparison-contrast (e.g., the Herr et al. findings), the relation appears to be more one of analogy than similarity. For the reasons stated above, the cause of the reversal of the priming effect appears to lie in the response to the prime alone, not the comparison of the prime and target. While reverse priming results in what may look like a contrast effect, because the hypothesized underlying mechanism, correction, is distinct from comparison, we are careful not to call this a contrast effect, lest it be confused with the more commonplace comparison-contrast effects.

Unconscious Volition?

If the reverse priming effects do represent an automatic correction, they provide evidence for a complex, sophisticated, and even volitional unconscious. Not only are subjects spontaneously correcting for an unwanted bias, they are doing so in the context of having no conscious awareness of the potential for bias. Because of the unobtrusiveness of the procedure (e.g., due to the use of the pronunciation task and the generic, nonsensitive—i.e., not race-related—nature of the stimuli), subjects do not even know that they are evaluating the targets, let alone the primes, or that they might respond faster when the prime and target are evaluatively congruent. Discussions with subjects during debriefing confirmed this lack of awareness.

The goal to pronounce the target accurately may well be a deliberate one,

given that the experiment instructions demand as much. The goal to evaluate the target, however, and any intention to correct for threats to such a goal, would almost certainly have to be unconscious, because participants are unaware and unlikely to infer that evaluative priming is the subject of the experiment. Furthermore, the use of short SOAs (150 and 300 ms), coupled with relatively short mean response times (approximately 550 ms), make it almost certain that latencies to respond to the targets as a function of their relation to the primes reflect automatic processes that are, therefore, not subject to conscious control (e.g., Neely, 1977). Given these conditions, the likelihood is very low that the correction derives from any deliberate process. Additionally, the accuracy motivation itself is not a sufficient cause of reverse priming; rather, a sense that accuracy might be threatened must also be present, and this may derive from a chronic unconscious vigilance for bias.[2]

There is further evidence in the Glaser and Banaji (1999) data that the vigilance necessary to trigger the accuracy-motivated correction is nonconscious. First of all, the pattern (normal priming for mild primes and reverse priming for extreme primes) obtains throughout the procedure, in the earliest, middle, and latest trials of the experiment. Therefore, it does not appear to result from a developing strategy arising from aroused suspicion on the part of subjects. More important, the effect obtained for a remarkably high number of subjects (virtually all) in the earliest experiments (those including racial stimuli) despite the fact that only a handful, during debriefing, guessed that prime and target congruence would affect response time (and no subjects surmised that evaluative congruence mattered). Past research has demonstrated convincingly that evaluation can occur spontaneously, unintentionally, and without awareness (e.g., Bargh et al., 1992, 1996; Fazio et al., 1986; Greenwald et al., 1989, 1996; see Fazio, 2001; Klauer, 1998, for reviews). The reverse priming effects appear to indicate that on some nonconscious level we know—we are "aware," if you will—that we evaluate extraneous, potentially distracting or biasing stimuli, and we will consequently attempt to correct accordingly.

The Prevalence of Reverse Priming

If reverse priming is restricted to the specific conditions of the Glaser and Banaji (1999) research, even with its many procedural variances and differing stimulus sets across experiments, then its generality and, consequently, importance is debatable. However, reverse priming effects reported by others suggest that the phenomenon is not altogether rare and that it is not limited to specific conditions such as pronunciation tasks.

Hermans (1996, as described in Banse, 2001) obtained reverse priming results in an automatic evaluation experiment using an evaluative judgment

task. More recently, Banse (2001) found reverse priming with an evaluation task with subliminal face and name primes, but not when the primes were supraliminal. Although at this stage the discrepancy in Banse's findings with regard to the perceptibility of the primes is inexplicable, the fact that reverse priming occurred with subliminal primes, which subjects could not consciously perceive, makes an even stronger case that the compensatory processes, if that is what they are, transpire outside of conscious awareness or control.

Another study may shed light on the issue of underlying mechanisms for reverse priming. Hypothesizing that people with high anxiety may be more likely to elaborate affective stimuli (Mogg & Marden, 1990), which, in turn, would lead to reverse priming among high- but not low-anxiety subjects, Maier, Berner, and Pekrun (2003) adopted the Glaser and Banaji (1999) procedure (as in the later experiments, excluding racial stimuli), and segmented their sample into low-, moderate-, and high-anxiety subjects. Consistent with Glaser and Banaji's (1999) findings, moderate primes yielded no priming effects. However, with extreme primes, subject anxiety moderated the priming effect qualitatively. Low- and moderate-anxiety subjects showed normal priming effects, while high-anxiety subjects showed reverse priming. In addition to providing a replication of reverse priming and offering a moderating variable, this finding could be interpreted as evidence that reverse priming reflects a vigilance for biasing information and an accuracy-motivated correction, which may be higher among highly anxious people.

Perhaps of greatest relevance to the issue of automatic correction, Wentura (2000) has found that, with subliminal priming and an evaluation task, when the speed of the response is emphasized in the instructions to subjects, normal priming effects are obtained. When accuracy is emphasized, however, reverse priming results. Wentura offers a different explanation than automatic correction. Drawing on Milliken, Joordens, Merikle, and Seiffert's (1998) application of selective attention to negative priming,[3] Wentura argues that accuracy motivation will lead subjects to try to discriminate between stimuli, thereby leading to slower responding when prime and target are congruent, presumably because discrimination is harder when they are similar. Such an explanation would not likely apply to the Glaser and Banaji (1999) findings because only prime, and not target, extremity determined reverse priming. If incongruence effects resulted merely from difficulty in discriminating between prime and target, and even if one argued that such difficulty would occur only when stimuli are clearly valenced (as with extreme words), Wentura's explanation would predict target extremity to play at least as large a role as prime extremity. That Wentura's discrimination explanation probably does not apply to the Glaser and Banaji findings is not as important as the fact that this explanation, like that of automatic correction, requires unconscious

vigilance for biasing information, considering that Wentura's primes were presented subliminally.

Like Wentura, Glaser (2003) has directly tested the role of accuracy motivation in reverse priming. In an effort to reconcile the findings of Glaser and Banaji (1999) with those of Bargh et al. (1996) wherein, despite very similar procedures, no reverse priming was evidenced, Glaser (2003) replicated one of the Bargh et al. (1996) experiments, with identical word stimuli and highly similar procedures. Hypothesizing that accuracy motivation was a necessary condition for the corrective processes that yield reverse priming, Glaser (2003) manipulated between subjects the presence of an accuracy-enhancing instruction. Specifically, in earlier research in which reverse priming effects were consistently obtained with extreme primes (e.g., Glaser & Banaji, 1999), experiments had included in the instructions to subjects a directive which stated that trials on which they made errors would be repeated later, and indicated explicitly that it was therefore in the subjects' interests to respond accurately. Such an instruction could serve to boost the drive to be accurate, lest subjects have to endure a longer procedure. In the more recent experiment (Glaser, 2003), this instruction was given to only one group of subjects, while the others were given no such warning. The results indicate that the instruction was influential. Specifically, those who did not receive the instruction showed a pattern of results more consistent with that of Bargh et al. (1996), with normal priming for extreme primes.[4] More important, for the subjects who were given the accuracy-enhancing instructions, no priming effect was obtained with extreme primes. Interestingly, in this instruction condition (with the accuracy imperative), a reliable priming effect was obtained with moderate primes, suggesting that the accuracy instructions were sufficient to bolster attention to the stimuli enough to strengthen priming with even weakly valenced primes, but perhaps also sufficient to instigate some degree of correction with the extreme primes, but not as much as in previous experiments where overcorrection (and hence reverse priming) was evident. The presence of the priming effect with moderate primes in this condition is important in ruling out the possibility that the accuracy instructions simply prevented priming of any sort, a possibility that was nevertheless unlikely given that such instructions have been used in previous experiments (Glaser & Banaji, 1999) that repeatedly obtained both normal and reverse priming with similar procedures.

The absence of a reverse priming effect with strong accuracy instructions is not entirely compelling evidence for an accuracy-mediated correction account. However, in the presence of a normal priming effect with moderate primes, the absence of any effect with extreme primes, with which reverse priming effects have been repeatedly obtained in the past, suggests that some degree of reverse priming is occurring when accurate responding is encour-

aged. Consistent with this, Wentura (2000) entertained the possibility, and we agree, that a null result in an automatic evaluation experiment, "might reflect a heterogeneous distribution of some subjects showing a congruence effect and some showing an incongruence effect" (p. 4). In other words, reverse priming, and the unconscious vigilance for biasing information that gives rise to it, may be relatively common but masked by individual differences within an experiment wherein normal and reverse priming effects may cancel each other out. Similarly, null results in automatic evaluation studies may result from the use of priming stimuli that straddle the evaluative extremity continuum enough to engender competing, and self-canceling, normal and reverse priming responses, as may be the case in Glaser and Banaji's (1999) post hoc analyses of the least extreme of their extreme primes, which showed no priming effect at all. This is not to say that all or even most null results in affective priming experiments are confounded by reverse priming; this would be an unfalsifiable and reckless assertion. Rather, it simply seems likely that, because until recently reverse priming was undiscovered, it may be more prevalent than one would think.

Glaser and Banaji (1999) suggested that, because of the counterintuitive nature of reverse priming effects and the likelihood that they will confound predictions, many such findings may never have been published and allowed to see the proverbial light of day. In fact, as Glaser and Banaji noted, a quick survey of the published research on nonconscious priming reveals more than a few unexpected, and often unexplained, contrast effects (e.g., Banaji, Hardin, & Rothman, 1993; De Houwer, Hendrickx, & Baeyens, 1997; Eimer & Schlaghecken, 1999; Murphy & Zajonc, 1993) that may reflect unconscious compensatory mechanisms and at least raise the question of the prevalence of such findings, published or not.

Controlling Automatic Stereotypes

Another literature promises to shed further light on the issue of unconscious correction for unintended thoughts and biases. In recent years, acknowledging that exposure to cultural stereotypes is virtually inevitable (Devine, 1989; but see Lepore & Brown, 1997, 1999), social psychologists have sought to identify conditions under which the automatic activation and application of such stereotypes can be derailed. Although most conceptions of automatic processes hold that they are beyond deliberate control (e.g., Bargh, 1994; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), there have been compelling arguments allowing for some measure of control (e.g., Logan, 1989). With regard to controlling automatic stereotypes, initial findings indicate that, while stereotype suppression is difficult and perhaps even counterpro-

ductive (e.g., Macrae, Bodenhausen, Milne, & Jetten, 1994; Macrae, Bodenhausen, Milne, & Wheeler, 1996; Wegner, 1994), under some circumstances people appear capable of moderating the effects of automatic stereotype activation (see Monteith, Sherman, & Devine, 1998; and Blair, 2002, for reviews). Research has now documented that people who are highly motivated to control prejudice can dampen the explicit expression of automatically activated biases (e.g., Fazio et al., 1995), perhaps with the subsequent activation of egalitarian "replacement thoughts" (Monteith, 1993), and this has been demonstrated even when measuring the automatic activation of the stereotypes, as in a semantic priming paradigm (Blair & Banaji, 1996). Similarly, Kawakami, Dovidio, Moll, Hermsen, and Russin (2000) have shown that following considerable counterstereotype training, subjects will show less automatic stereotype activation. This concept of competing impulses is also expressed in Wilson, Lindsey, and Schooler's (2000) model of dual attitudes, where the concept of automatic override (p. 106) holds that explicit attitudes may trump implicit ones. However, promising as these stereotype elimination effects are, they require a somewhat deliberate strategy and/or the complete relearning of automatic associations, which cannot be readily equated with unconscious volition and control.

The question of whether the automatic activation of stereotypes can be prevented or intercepted due to nonconscious motivations has only recently been addressed. Moskowitz et al. (1999; see also Moskowitz, 2001) demonstrated that people with high egalitarian (e.g., antiprejudice) goals exhibit less automatic activation of stereotypes even though they have the same knowledge of these stereotypes as do those with lower egalitarian motives, who showed greater automatic activation of stereotypes. Having identified an indirect measure of chronic egalitarian goals, Moskowitz et al. (1999) submitted subjects who scored high and low in egalitarianism to a procedure with striking similarities to that employed by Glaser and Banaji (1999) and Glaser (2003); an automatic gender stereotyping test employing a semantic priming paradigm with a 200 ms SOA (the time from onset of the prime to the onset of the target stimulus) and latency to pronounce the target word as the dependent variable. Photographs of men and women were used as primes, while attributes stereotypical of men and women, as well as gender-irrelevant words, were presented as targets. The use of the short SOA as well as the relatively ambiguous pronunciation task served a function similar to that intended by Glaser and Banaji (1999); to ensure that differential response times as a function of prime-target gender stereotype match reflected automatic activation of gender stereotypes by the mere perception of the primes, and not an intentional response. Predicting that chronic egalitarian goals would serve to obstruct even the automatic activation of gender stereotypes, Moskowitz et al. (1999) indeed found that subjects who scored high

on their indirect measure of chronic egalitarianism exhibited no automatic gender stereotyping. Importantly, those who had scored low in chronic egalitarianism did exhibit automatic gender stereotype activation in this experiment.

In order to demonstrate more directly that those high in chronic egalitarianism actively, albeit nonconsciously, inhibited the automatic activation of stereotypes, Moskowitz et al. (1999) followed up this experiment with a negative priming experiment wherein participants were presented with two primes simultaneously; one that was supposed to be ignored but that varied in its female stereotype relevance. Target words, which were again to be pronounced, were also either stereotype relevant or not, and they were either the same word as the prime that was to be ignored or not. Moskowitz et al. found again that only those subjects who were low in chronic egalitarianism exhibited automatic stereotype activation. In contrast, those who had scored high in chronic egalitarianism were actually slower to respond to feminine-stereotypical target words following female distractor primes than following gender-irrelevant primes, indicating that they were effectively inhibiting, or perhaps correcting for, the stereotypical content that might have been activated by these primes. Moskowitz et al. concluded that egalitarians, while sharing knowledge of cultural stereotypes, are able to counteract the automatic activation of those stereotypes without conscious intent.

We are very sympathetic to Moskowitz et al.'s view, similar to that expressed by Glaser and Banaji (1999), that volition, in the form of vigilance for unintended bias, can occur outside of consciousness: "Despite the fact that the English language vernacular equates intent with conscious and effortful forms of pursuing a desired end state, volition can be exerted preconsciously. A passive process like stereotype activation could be controlled by goal pursuit, which could be activated as passively as stereotype activation" (Moskowitz et al., 1999, p. 169).

Although the Moskowitz et al. findings make a strong case for unconscious volition and compensation for unwanted thoughts, the reverse priming in automatic evaluation effects (Glaser & Banaji, 1999) add some value to the argument because, in the absence of group-relevant information (e.g., faces, names, stereotypes), motivations to avoid bias are unlikely to originate consciously. In other words, the Moskowitz et al. findings illustrate compellingly that chronic goals will motivate the suppression of an automatic response, but it is not yet entirely clear that such goals would be activated without the conscious recognition that something about bias was being measured. In fact, Wasel and Gollwitzer (1997) found, using a similar paradigm, that high egalitarians inhibited automatic stereotype activation only when primes were supraliminal (i.e., consciously perceptible), but when primes were presented subliminally, even high egalitarians failed to inhibit automatic stereotyping, suggesting that at least the potential for bias may need to reach conscious-

ness for inhibition or correction to occur. On the other hand, reverse priming in automatic evaluation with a highly nonreactive pronunciation task, especially in experiments excluding racial stimuli (Glaser & Banaji, 1999, Experiments 4–6), holds considerably less chance that the correction is related to a consciously activated goal. Furthermore, Banse's (2001) finding of reverse priming with subliminal primes indicates that conscious awareness of the potential for bias is not necessary. Finally, the demonstrated moderating role of accuracy instructions (Glaser, 2003; Wentura, 2000) and trait anxiety (Maier et al., 2003) strongly suggest that reverse priming reflects a motivated, specifically accuracy-motivated, corrective response. Even though in the Glaser and Banaji (1999) studies and for some subjects in Glaser (2003), accuracy is explicitly encouraged in the instructions, such exhortations are likely to only indirectly influence the nonconscious accuracy motivation that we posit gives rise to reverse priming. Subjects are typically unaware that evaluation is being measured, and consequently any motivation to correct for bias in a judgment that one does not even know is being made, let alone that the potential for bias exists, must itself have nonconscious bases. With reverse priming, therefore, we have a full representation of the evaluative process, including the evaluative response itself, the goal to evaluate accurately, and the corrective measures to ensure accuracy, all operating under the proverbial hood—nonconsciously.

Conclusion

Evidence from studies of automatic affect (e.g., Glaser & Banaji, 1999) and cognition (e.g., Moskowitz et al., 1999) suggests that, in addition to the ability to process the meaning of, categorize, and evaluate perceived stimuli automatically, the human mind is capable of maintaining unconscious vigilance over its own automatic processes. This suggests a volitional nature of the unconscious, an idea that to many may seem self-contradictory. In the construct activation literature so central to social cognitive theory and research, the unconscious has been credited with (or blamed for, as the case may be) simplistic, assimilative influences on judgments, while the compensatory efforts necessary to obtain contrast effects have been ascribed strictly to deliberate, conscious processes (e.g., Wilson & Brekke, 1994). However, when observed in responses that most likely reflect automatic processes, "contrast" effects stemming from correction (e.g., Glaser & Banaji, 1999) indicate that people can unconsciously monitor and correct for bias in judgments, just as they might consciously.

That goals can operate at the unconscious level, and subsequently influence explicit judgments and behaviors, is now well demonstrated (e.g.,

Bargh & Barndollar, 1996; Bargh et al., 2001; Chartrand & Bargh, 1996). Evidence that corrective processes can occur even when a person is unaware that she is making a judgment, as in the case of a priming procedure with a nonreactive response task (i.e., the pronunciation task), and especially considering that accuracy motivation moderates such automatic compensatory effects (Glaser, 2003; Wentura, 2000), however, implicates unconscious, metacognitive processing goals. Similarly, demonstrations of the effects of chronic ideological goals (e.g., egalitarianism) on suppressing automatic, and therefore previously presumed uncontrollable, responses (Moskowitz et al., 1999) bolster the thesis of unconscious volition. Building on the trend to allow for a full spectrum of mental life (i.e., affect, cognition, and motivation) in the unconscious (Bargh, 1996, 1997; Kihlstrom, 1987, 1999), we go a step further to suggest that the unconscious is indeed capable of holding such metacognitive processing goals (e.g., accuracy) which it will pursue through self-monitoring, and that it will, under some conditions, compensate for anticipated threats to the attainment of those goals. This thesis, and the findings supporting it, represents a departure from traditional conceptions of the unconscious as passive and reactive, suggesting an unconscious that is, paradoxically, "aware."

Acknowledgments Production of this chapter was supported by a National Research Service Award Postdoctoral Fellowship from the National Institute of Mental Health, Grant MH12195 to the first author and Grant MH35856 to the second author, as well as institutional support from the Institute of Personality and Social Research at the University of California, Berkeley. We are grateful to Molly Parker Tapias for incisive comments and Miranda Chiu and Brian Penrod for assistance with the research.

Notes

1. The food words were included because of an interest in a possible "disgust" component of racial bias.

2. This concept of unconscious vigilance should be distinguished from Pratto and John's (1991) theory of "automatic vigilance," which holds that people are chronically vigilant for negative information.

3. "Negative priming" is an inhibitory effect wherein responses are slower to targets that served as primes in preceding trials (May, Kane, & Hasher, 1995). It is distinct from what we are calling reverse priming, which appears to be a within-trial phenomenon.

4. Inconsistent with Bargh et al. (1996), no priming effect was obtained with moderately valenced primes in the low-accuracy instruction condition, suggesting that attitude strength may under some conditions moderate attitude activation (e.g., Fazio et al., 1986).

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

Banse, R. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition and Emotion, 15,* 501–520.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1. Basic processes* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A. (1996). Principles of automaticity. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169–183). New York: Guilford.

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

Bargh, J. A., & Barndollar, K. (1996). Automaticity in action: The unconscious as repository of chronic goals and motives. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 457–481). New York: Guilford.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62,* 893–912.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32,* 104–128.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462–479.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126,* 925–945.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81,* 1014–1027.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242–261.

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

Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology, 71,* 464–478.

De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with "subliminally" presented stimuli. *Consciousness and Cognition, 6,* 87–107.

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

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.

Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1737–1747.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion, 15,* 115–141.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (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.

Ford, T. E., & Kruglanski, A. W. (1995). Effects of epistemic motivations on the use of accessible constructs in social judgment. *Personality and Social Psychology Bulletin, 21,* 950–962.

Glaser, J. (2003). Reverse priming: Implications for the (un)conditionality of automatic evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 87–108). Mahwah, NJ: Erlbaum.

Glaser, J., & Banaji, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology, 77,* 669–687.

Gollwitzer, P. M. (1999). Implementation intentions. *American Psychologist, 54,* 493–503.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science, 273,* 1699–1702.

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition, 17,* 35–47.

Hermans, D. (1996). *Automatic stimulus evaluation: An experimental analysis of the preconditions for evaluative stimulus discrimination using an affective priming paradigm.* Unpublished doctoral dissertation, University of Leuven, Belgium.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion, 8,* 515–533.

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology, 51,* 1106–1115.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology, 19,* 323–340.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.

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

Hilgard, E. R. (1977). *Divided consciousness.* New York: John Wiley.

Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences, 16,* 107–117.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78,* 871–888.

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

Kihlstrom, J. F. (1996). Unconscious processes in social interaction. In S. Hameroff, W. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness: The first Tucson discussions and debates* (pp. 93–104). Cambridge, MA: MIT Press.

Kihlstrom, J. F. (1999). The psychological unconscious. In L. A. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 424–442). New York: Guilford.

Kihlstrom, J. F., Mulvaney, S., Tobias, B. A., & Tobis, I. P. (2000). The emotional unconscious. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas, & P. M. Niedenthal (Eds.), *Cognition and emotion* (pp. 30–86). New York: Oxford University Press.

Klauer, K. C. (1998). Affective priming. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (pp. 67–103). New York: John Wiley and Sons.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*, 557–558.

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

Lepore, L., & Brown, R. (1999). Exploring automatic stereotype activation: A challenge to the inevitability of prejudice. In D. Abrams & M. A. Hogg (Eds.), *Social identity and social cognition* (pp. 141–163). Malden, MA: Blackwell.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*, 255–275.

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

Logan, G. D. (1997). The automaticity of academic life: Unconscious applications of an implicit theory. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 157–179). Mahwah, NJ: Erlbaum.

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

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., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition, 14*, 1–20.

Maier, M. A., Berner, M. P., & Pekrun, R. (2003). Directionality of affective priming: Effects of trait anxiety and activation level. *Experimental Psychology, 50*, 116–123.

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

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

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin, 118*, 35–54.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*, 690–702.

Milliken, B., Joordens, S., Merikle, P. M., & Seiffert, A. E. (1998). Selective attention: A reevaluation of the implications of negative priming. *Psychological Review, 105*, 203–229.

Mogg, K., & Marden, B. (1990). Processing of emotional information in anxious subjects. *British Journal of Clinical Psychology, 29*, 227–229.

Monteith, M. J. (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. (2001). Preconscious control and ompensatory cognition. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 334–358). Mahwah, NJ: Erlbaum.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167–184.

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*, 723–739.

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*, 225–254.

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.

Newman, L. S., & Uleman, J. S. (1990). Assimilation and contrast effects in spontaneous trait inference. *Personality and Social Psychology Bulletin, 16*, 224–240.

Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology, 26*, 505–527.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

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

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380–391.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84*, 1–66.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven, CT: Yale University Press.

Shiffrin, R. M. (1997). Attention, automatism, and consciousness. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 49–64). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127–190.

Stapel, D. A., Koomen, W., & Zeelenberg, M. (1998). The impact of accuracy motivation on interpretation, comparison, and correction processes: Accuracy × knowledge accessibility effects. *Journal of Personality and Social Psychology, 74*, 878–893.

Stapel, D. A., Martin, L. L., & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin, 24*, 797–806.

Strack, F., Schwarz, N., Bless, H., Kübler, A., & Wänke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology, 23*, 53–62.

Tesser, A., & Martin, L. (1996). The psychology of evaluation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 400–432). New York: Guilford.

Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology, 66*, 474–489.

Wasel, W., & Gollwitzer, P. M. (1997). Willful control of "automatic" stereotype activation: The role of subliminally vs. supraliminally presented stimuli. *Sprache und Kognition, 16*, 198–210.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology, 68*, 36–51.

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

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, 4th ed., pp. 446–496). Boston: McGraw-Hill.

Wentura, D. (2000). *Masked priming in the evaluation task: A switch from positive to negative priming due to speed-accuracy instructions.* Unpublished manuscript, University of Münster, Germany.

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

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*(2), 151–175.

THE NEW UNCONSCIOUS

Edited by
Ran R. Hassin
James S. Uleman
John A. Bargh

OXFORD
UNIVERSITY PRESS

2005

OXFORD
UNIVERSITY PRESS

Oxford New York
Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
The new unconscious / edited by Ran R. Hassin, James S. Uleman, and John A. Bargh.
p. cm — (Oxford series in social cognition and social neuroscience)
Includes bibliographical references and index.
ISBN 0-19-514995-5
1. Subconsciousness. I. Hassin, Ran R. II. Uleman, James S. III. Bargh, John A. IV. Series.
BF315 .N47 2004
154.2—dc22
2003021368

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper