When Fair Is Foul and Foul Is Fair: Reverse Priming in Automatic Evaluation

Jack Glaser
University of California, Berkeley

Mahzarin R. Banaji
Yale University

Responses to information were facilitated by the rapid prior presentation of evaluatively congruent material. This fundamental discovery (R. H. Fazio, D. M. Sanbonmatsu, M. C. Powell, & F. R. Kardes, 1986) marked a breakthrough in research on automatic information processing by demonstrating that evaluative meaning is grasped without conscious control. Experiments employing a word naming task provided stringent tests of the automaticity of evaluation and found support for it. More strikingly, a previously unobserved reversal of these effects (i.e., slower responses to evaluatively matched rather than mismatched items) was found when primes were evaluatively extreme. Procedural variances across 6 experiments revealed that the reverse priming effect was highly robust. This discovery is analogous to demonstrations of contrast effects in controlled judgments. It is theorized that the reverse priming effect reflects an automatic correction for the biasing influence of the prime.

In the course of the daily business of science, unexpected findings in the hands of careful experimentalists can yield discoveries of new phenomena. As an example, we cite Henri Tajfel's expectation that a group constructed in a transparently arbitrary manner would be free of the intergroup biases prevalent in a "real group" (Tajfel, 1978, pp. 10–11). The unexpected result he obtained, that mere categorization of individuals into arbitrary groups provoked intergroup bias, yielded the important concept of "minimal groups." When an experimental finding not only contradicts expectation but also previous research, and when this result is so robust that virtually every participant's performance refutes expectation, it warrants closer attention. We report such a discovery in this paper.

This research was supported in part by a National Research Service Award postdoctoral fellowship from the National Institute of Mental Health (1F32MH12195) and grants from the National Science Foundation (DBC-9120987, SBR-9422241, and SBR-9709924), as well as the John Simon Guggenheim Fellowship and the Cattell Fund Fellowship.

We thank Hillary Haley and John Muse for their assistance with data collection and coding; John Bargh, Wil Cunningham, Russ Fazio, Tony Greenwald, Richard Hackman, Curtis Hardin, John Kihlstrom, Kristi Lemm, Jason Mitchell, Jim Neely, Norbert Schwarz, and Diederik Stapel for their helpful comments on earlier versions of this article; and Peter Frensch, John Jost, and John Turner for pointing us to important, relevant sources.

Correspondence concerning this article should be addressed to Jack Glaser, Institute of Personality and Social Research, University of California, 4143 Tolman, MC# 5050, Berkeley, California 94720-5050, or to Mahzarin R. Banaji, Department of Psychology, Yale University, P.O. Box 208205, New Haven, Connecticut 06520-8205. Electronic mail may be sent to glaserj@socrates.berkeley.edu or mahzarin.banaji@yale.edu.

The Ubiquity of Automatic Evaluation

Decades ago, psychologists established that semantic meaning could be activated rapidly and spontaneously by the mere presentation of a word, and that this meaning would, in turn, activate associated concepts in the mind. If proximally presented words shared meaning (i.e., were semantically associated), then responses to them would be facilitated (Meyer & Schvaneveldt, 1971; Posner & Snyder, 1975). Neely (1977) demonstrated that such concept activation occurs "automatically," that is, without conscious control. Given sufficient time between the presentation of two words in a given pair, people could counteract (i.e., control) the effects of the priming stimulus (i.e., the first word or concept presented) on the judgment of the target stimulus. With limited time (e.g., 250 ms), however, such control was not possible and the automatic effects of the prime were evident.

More recently, social psychologists have provided compelling evidence that people also evaluate objects automatically (Bargh, Chaiken, Govender, & Pratto, 1992; Bargh, Chaiken, Raymond, & Hymes, 1996; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald, Draine, & Abrams, 1996). Building on demonstrations of automatic semantic priming, Fazio et al. (1986) presented a series of word-pair sequences and had participants judge whether target adjectives were negative or positive. The first word (the prime) appeared for 200 ms, and the second (the target) appeared 100 ms after the prime had disappeared, resulting in a 300-ms stimulus onset asynchrony (SOA). Participants made faster responses when the prime and target were evaluatively congruent (i.e., both were negative or both were positive) than when they were incongruent. These results, given the conditions under which the stimuli were presented, necessitated that the participants were

Journal of Personality and Social Psychology, 1999, Vol. 77, No. 4, 669–687

evaluating the primes without conscious control. Just as the semantic priming work had shown that meaning is extracted automatically (Neely, 1977), Fazio et al. (1986) showed that evaluative associations, or attitudes, were activated automatically.

Bargh and colleagues (1992) replicated and extended this finding, arguing that automatic evaluation occurs regardless of the strength (i.e., extremity or accessibility) of the attitude toward the prime. Bargh et al. (1996; see also Hermans, de Houwer, & Eelen, 1994) made a case for the unmoderated generality of automatic evaluation by demonstrating the effect (a) in the absence of environmental evaluative cues (e.g., an instruction to specifically evaluate the target words) and (b) with subtly valenced primes and targets, as well as with those that are more clearly positive and negative. They achieved this by first employing a pronunciation task wherein participants simply read the target words aloud rather than evaluate them per se. Additionally, they selected consensually rated, subtly valenced words (e.g., dormitory, dentist) in addition to more clearly positive and negative words (e.g., flowers, disease). Thus, even when the judgment task did not require an explicit evaluative judgment, evaluatively matched items yielded faster pronunciations. The only way to account for faster pronunciations of target words preceded by evaluatively matched primes is to allow that both primes and targets are being evaluated even without explicit instructions or environmental cues, thus further demonstrating the ubiquity of automatic evaluation. Greenwald and colleagues (Greenwald, Klinger, & Liu, 1989; Greenwald et al., 1996) have also provided evidence for the spontaneous, unconscious evaluation of words by demonstrating that primes presented subliminally (i.e., not consciously recognizable) facilitated evaluations of evaluatively congruent target words. Together, these studies form the core of the evidence that people extract automatically the evaluative content of information.

Researchers interested in assessing intergroup stereotypes (i.e., trait ascriptions) and prejudices (i.e., evaluative biases) at the implicit level have employed similar procedures. Gaertner and McLaughlin (1983) and Dovidio, Evans, and Tyler (1986) provided the earliest demonstrations that Black and White race primes differentially affected latencies to evaluate trait adjective targets that were negative and positive, respectively, but these studies involved relatively long SOAs (e.g., 2,500 ms), thus precluding conclusions about automaticity. Fazio, Jackson, Dunton, and Williams (1995) subsequently demonstrated automatic race prejudice by pairing Black and White face primes with target adjectives, using a 300-ms SOA. They found that White participants were faster to respond to Black-negative and White-positive pairs. Perdue and Gurtman (1990), presenting age-related primes subliminally, showed college students to have negative attitudes toward old age. Similarly, Wittenbrink, Judd, and Park (1997) presented the words *Black* and *White* as subliminal primes and found them to facilitate responses to negative Black stereotypic trait words and positive White stereotypic trait words, respectively. These automatic prejudice effects have been bolstered by findings of automatic stereotyping as well (Banaji & Hardin, 1996; Blair & Banaji, 1996; Kawakami, Dion, & Dovidio, 1998). With such strong support from the literature, our expectation, as we conducted the first experiment on automatic evaluation and automatic prejudice, was that we would obtain results consistent with this ubiquitous phenomenon.

Experiment 1: First Sighting of a Reverse Priming Effect in Automatic Evaluation

Bem (1987) recommends, and we agree, that the empirical research paper ought to tell a story about a phenomenon and not detail every turn the research enterprise took in its journey to interim completion. It would be disingenuous, however, to suggest that we predicted or expected the reverse priming effect obtained in this first study. We simply did not. Experiment 1 produced evidence for automatic evaluation, but also the opposite effect. We report Experiment 1 in lesser detail because it served largely as the first sighting of the unpredicted discovery of the reverse priming effect in automatic evaluation that inspired the subsequent research aimed at testing the parameters of the effect.[1]

Experiment 1 was conducted to test for race prejudice under conditions of stricter automaticity than those in previous experiments. To do this, we adopted the procedure employed by Bargh and colleagues (1996), in which participants simply read aloud target words that followed primes. The importance of this task, as noted by Bargh et al. (1996), is that pronunciation of the target (as opposed to classification of the target as good versus bad) is, from the participant's perspective, independent of evaluation and hence affords a purer test of automatic evaluation (see also Hermans et al., 1994; Balota & Lorch, 1986; and Neely, 1991, for a review of experiments using pronunciation tasks). That is, if evaluatively congruent and incongruent prime–target pairs yield differential response times in the absence of an explicit goal to evaluate the target (as in a good–bad judgment task), evaluative priming can be viewed more confidently as an unconscious, automatic process. To further ensure that response latency differences reflected automatic processes, the SOA was reduced to 150 ms from the typical 300 ms used in supraliminal evaluative priming studies.

Four distinct categories of word stimuli were used (see the Appendix for the specific words): words with an African American association, words with a European American association, race-neutral words (heretofore referred to as "generic" words), and race-neutral food words. All the words (except the food words *bitter* and *sweet*) were nouns. The generic words were selected, based on the normative ratings obtained by Bellezza, Greenwald, and Banaji (1986), for their evaluative clarity (i.e., they were readily distinguishable as positive or negative). The race words were generated by a sample of Yale undergraduates, and all words were rated by another sample of Yale undergraduates on race and evaluative dimensions. Within each category of words, half were negative and half were positive. The use of the pronunciation task permitted a measure of response latencies to words across multiple dimensions (specifically evaluation and race) with a single type of response.

Word types in this study were paired in every possible combination for each participant, and all words served as both primes and targets. This created a large within-subjects design of 2 (prime valence: negative, positive) × 4 (prime category: food, generic, Black-associated, White-associated) × 2 (target valence: negative, positive) × 4 (target category: food, generic, Black-associated,

[1] The design and procedure for Experiment 2 is almost identical to that of Experiment 1. Much of the detail can therefore be obtained from the description of Experiment 2.

White-associated), allowing for simultaneous tests of multiple hypotheses about automatic evaluation, race categorization, and prejudice. A primary prediction was that responses would be fastest when primes and targets were evaluatively congruent (i.e., when both prime and target were positive or when both were negative). We refer to this as the *automatic evaluation effect.* Similarly, we predicted that the evaluative valence of the primes would interact with the race category of the target such that at least for non-Black participants, responses would be faster to negative-Black and positive-White pairings than to negative-White and positive-Black pairings. This would indicate relatively negative and positive evaluations of Black and White race categories, respectively.[2] We refer to this as the *automatic prejudice effect.* The combination of these measures within one procedure promised to reciprocally bolster construct validity for the measures. For example, the demonstration of automatic evaluation (i.e., the facilitation of responses to targets preceded by similarly valenced primes) with race-neutral primes would support the claim that differential response latencies to Black and White target words preceded by negative and positive primes reflect an evaluative bias.

Our hypotheses were supported, with one glaring exception. As Figure 1 depicts, when the food words served as primes, we obtained the predicted pattern of results: Evaluatively congruent prime–target pairings yielded faster responses than did evaluatively incongruent pairings.[3,4] A repeated measures analysis of variance (ANOVA) revealed this to be statistically significant, $F(1, 42) = 8.58$, $p < .01$, $r = .41$.[5] However, when the generic (i.e., race-neutral, nonfood) words served as primes, we obtained results that were the opposite of those predicted. Specifically, evaluatively congruent prime–target pairs yielded slower responses than did incongruent pairs, $F(1, 42) = 47.68$, $p < .0001$, $r = .73$. We dub this the *reverse priming effect.* Because the two types of primes (food versus generic) yielded such dramatically different results, the interaction of prime type by evaluative congruence was also highly significant, $F(1, 42) = 44.75$, $p < .0001$, $r = .72$.

Similar and even more dramatic results were obtained when the target words were race-related. As originally predicted, with the negative and positive food words serving as primes, participants were faster to respond to negative-Black and positive-White pairings. With the generic primes, on the other hand, the opposite result was again obtained such that negative-Black and positive-White pairs were responded to more slowly than were positive-Black and negative-White pairs. This striking interaction is depicted in Figure 2, where the predicted effect was clearly obtained with food primes, but the opposite effect was obtained with generic primes. These results are unlikely to be a consequence of Type I error. The standard priming effect with food primes was highly reliable, $F(1, 42) = 60.18$, $p < .0001$, $r = .77$, and so was the opposite effect with generic primes, $F(1, 42) = 123.19$, $p < .0001$, $r = .86$. The interaction effect was also convincing, $F(1, 42) = 197.13$, $p < .0001$, $r = .91$, making it clear that a feature of the primes dramatically moderated the manner in which participants responded to the stimuli. Both types of primes (food and generic) were clearly being evaluated; otherwise, their valence would not have affected responses to the targets at all. However, it is not clear why the generic primes appeared to be triggering an evaluation opposite to that of their own, yielding the reverse priming effect. The effect sizes in all cases were large enough to command our attention and warrant further exploration, both theoretical and experimental.

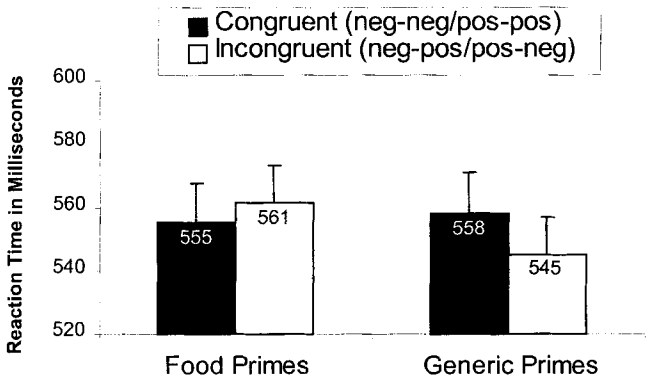

Figure 1. Automatic evaluation by prime type in Experiment 1. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

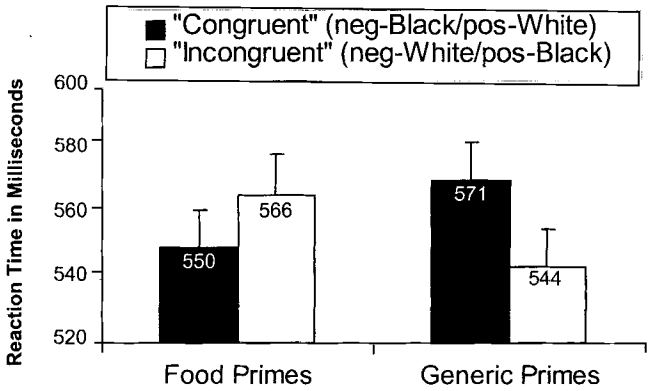

Figure 2. Automatic prejudice by prime type in Experiment 1. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

[2] Because race and race-neutral words served as both primes and targets and were combined in every possible way, we were also able to test for automatic prejudice in another way (i.e., the prime race by target valence interaction) and automatic race categorization (i.e., the prime race by target race interaction). Results for both of these analyses confirmed our hypotheses, but because they are not directly relevant to the primary questions at hand with regard to reverse priming as a function of prime extremity, we will not discuss them here in detail.

[3] Although statistical analyses were conducted on reciprocally transformed reaction times (see the Experiment 2 results section for a full description of the data analytic procedures), the data presented in the figures have been converted back into milliseconds to facilitate interpretation and comparison with other studies.

[4] Following the lead of Klauer, Rossnagel, and Musch (1997), we present the data in terms of evaluative congruence versus incongruence rather than the full design of prime valence by target valence.

[5] Pearson's r was calculated as a measure of effect size. For r, 0.1 is considered a small effect, 0.3 a medium effect, and 0.5 a large effect.

Assimilation and Contrast

Although the reverse priming effect obtained in Experiment 1 was not predicted, it is also the case that social psychologists are not strangers to such effects. Analogous findings have been reported in research on construct accessibility wherein priming stimuli influence judgments of targets in an assimilative manner, but under some conditions lead to judgments that are contrasted from the prime. Although the reaction time measures used in the present priming research do not lend themselves easily to an assimilation and contrast explanation of the results, there are important similarities that point to potential mediating mechanisms of reverse priming. To that end, we have drawn on theory and research on assimilation and contrast to begin formulating an explanation for why reverse priming effects might occur in automatic cognition.

Sherif and Hovland (1961) pioneered research on assimilation and contrast in social judgments. They provided early documentation that a comparison standard can either pull judgments toward it (assimilation) or push them away from it (contrast). Specifically, Sherif and Hovland argued that if the comparison standard is reasonably similar to the target, assimilation effects are likely; but if they are quite disparate, contrast may occur. For example, when estimating the weight of an object that is lifted following a slightly heavier object, assimilation occurs in the form of overestimating the weight of the second object. On the other hand, if a weight is lifted after one that is exceedingly heavy, its relative lightness by comparison produces a contrastive judgment in the form of underestimation of the second object.

Since the late 1970s, studies designed to measure the influence of accessible mental constructs (Higgins, 1996; Higgins, Rholes, & Jones, 1977) have also demonstrated assimilation and contrast effects. Lombardi, Higgins, and Bargh (1987), for example, reported that participants 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. On the other hand, if 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, with automatic processes accounting for assimilation and controlled processes engendering contrast.

Strack (1992; Strack & Hannover, 1996) argues that awareness of the influence of primes leads to corrective measures that can engender contrast effects. According to Strack, those making a judgment will engage in a "representativeness check" to determine if information is relevant to the judgment being made, but only when they are aware that such information may influence their judgment. If the information is determined to be nonrepresentative, the judgment will be corrected accordingly.

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. (1990) hypothesized that this would most likely occur when one has the cognitive resources to make such an adjustment, but not when such resources are depleted. Accordingly, they found that distracted participants showed assimilation toward primed concepts, whereas those who were not distracted showed contrast. Martin et al. corroborated this finding by reporting assimilation and contrast effects for participants 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 individuals who were reminded of a priming procedure showed contrast effects, whereas those who were not reminded exhibited assimilation. 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, contrast effects in judgments are more likely to result.

One determinant of information salience is its extremity. Accordingly, Herr, Sherman, and Fazio (1983; see also Herr, 1986) demonstrated that extreme primes yielded contrasted judgments, whereas moderate primes led to assimilative judgments of an ambiguous target. Specifically, participants were asked to judge the size (or, in another experiment, the ferocity) of ambiguous (i.e., fictitious) animals. When their judgments were preceded by presentations of animals that were moderately large (e.g., cow, lion) or moderately small (e.g., gopher, cat), judgments were influenced in an assimilative manner (e.g., primes of small animals led to judgments that target animals were relatively small). However, the presentation of extremely large (e.g., whale, elephant) or extremely small (e.g., flea, minnow) animals led to contrasted responses (e.g., primes of extremely small animals led to judgments that target animals were relatively large). The role of extremity as a determinant of prime salience is especially relevant here because the food and generic words employed in our Experiment 1 differed significantly in evaluative extremity. The food words (e.g., turnips, pears) were only mildly valenced, rated on average as -1.0 (for negative words) and $+1.03$ (for positive words) on an 11-point scale from -5 to $+5$, whereas the generic words (e.g., agony, paradise) were extreme, with mean ratings of -3.7 (negative) and $+3.85$ (positive) on the same scale.[6] This difference in our stimulus sets, and its parallel in the findings reported by Herr and colleagues, promised to explain when the automatic contrast effect observed in Experiment 1 was obtained. We discuss this possibility in greater depth later.

The effect of prime salience on assimilation and contrast effects may be moderated by the motivation to be accurate, which is known to influence how judgments are made (Kruglanski, 1990; Neuberg & Fiske, 1987). In one of their experiments, Martin et al. (1990) found that participants 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. This result suggests that anticipated accountability motivated participants to be vigilant and to adjust for the biasing influence of the primes. Other studies have more directly manipulated accuracy motivation,

[6] The difference in evaluative extremity between food and generic words was the inevitable result of the nature of food; foods (and the words that represent them) that are extremely unpleasant fall out of use (and language) quickly, or at least cease to be considered foods. As a consequence, it was difficult to find very negative food words, and so moderately positive food words were selected to avoid confounding valence with extremity among the food words.

finding that it attenuates assimilation effects (Ford & Kruglanski, 1995; Thompson, Roman, Moskowitz, Chaiken, & Bargh, 1994). Stapel, Koomen, and Zeelenberg (1998) have integrated these findings with new results to make the case that accuracy motivation leads to more careful processing of the target, thereby attenuating assimilation effects; but a correction strategy is required to bring about contrast effects. Of greatest relevance to the present research, 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. This, too, has the potential to explain why extreme primes, which are more blatently valenced, could elicit corrections in automatic evaluation and consequently yield reverse priming effects.

Correction has also been posited as a determinant of both contrast and assimilation effects by Wegener and Petty (1995; Petty & Wegener, 1993). These researchers 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, participants who expect assimilation effects correct away from the direction of the contextual information (i.e., the priming stimulus), whereas those who expect contrast effects correct toward the contextual information. These findings convey the complexity of perceivers' strategies as well as the excesses of their efforts when they attempt, at times unconsciously, to mitigate the effects of judgmental biases.

In sum, the research reviewed indicates that although contextual information can bias a response to supposedly unrelated stimuli, at times the result shows a contrastive pattern. Contrast occurs especially when the perceiver is aware of the potential biasing influence of the prime, perhaps as a result of its salience, and when the perceiver has the cognitive resources and motivation to recognize or remember the prime. It also appears to be the case that although assimilation effects occur spontaneously, contrast effects are more likely the result of an active correction. To date, studies of such corrections have been restricted to conditions under which judgments are relatively controlled and deliberate. Perhaps this is the case because of the assumption that corrections result from relatively deliberate processes. Given that assumption, one would expect to see only assimilative effects in automatic processing. However, the results of Experiment 1 suggest that corrections can occur under conditions where controlled processing is precluded, thus implicating an automatic correction process. One of the striking implications of automatic correction is that it would not likely be hampered by competing demands on cognitive resources (e.g., cognitive load) because automatic processes are understood to require virtually no effort and will occur inevitably, regardless of cognitive capacity (Bargh, 1994, 1996; Hasher & Zacks, 1979; Shiffrin & Schneider, 1977). Consequently, whereas distractions can preclude correction effects in more deliberate judgments (e.g., Martin et al., 1990), they would probably not interfere with automatic correction.

Although much of the literature attributes contrast effects to a judgmental distortion resulting from a salient comparison standard (i.e., "comparison contrast"), there is now evidence that contrast effects can result from overcorrection (Petty & Wegener, 1993; Stapel, Koomen, & Zeelenberg, 1998; Stapel, Martin, & Schwarz, 1998; Wegener & Petty, 1995). We theorize that the reverse priming effect in automatic evaluation represents such an overcorrection and may be, in this respect, akin to the contrast effects reported in the research described above. Consequently, predictors of contrast effects, such as prime salience, may apply to reverse priming as well.

If an inclination to correct for bias introduced by contextual information were to operate at the unconscious level, then we would expect to see contrast effects in automatic judgments varying as a function of the salience of the prime. In fact, as noted, perhaps the most clearly distinguishing feature of the food versus generic primes in Experiment 1 is evaluative extremity, and hence the salience of the evaluative valence. The food words were evaluatively subtle, whereas the generic words were extreme. Recall that the food primes yielded the expected "assimilative" effect (responses were faster to targets preceded by evaluatively congruent primes), whereas there were reversed effects with the generic primes (responses were slower to targets preceded by evaluatively congruent primes).

The effect obtained with the generic primes appears to represent an unconscious, automatic correction in the presence of contextual stimuli (primes) that by virtue of their extremity, present the potential to bias responses to the intended target of judgment. The purpose of the next experiment, therefore, is to determine whether prime extremity is indeed the moderating variable responsible for a reversal of the priming effect, as it is with more controlled judgments (Herr et al., 1983). Such a finding would support the proposition that it is the salience of the primes, as a consequence of their extremity, that triggers a correction, which is in this case an automatic correction.

Experiment 2: Extremity Matters

To test the hypothesis that extremity of the prime determines whether normal or reverse automatic evaluation will occur, Experiment 2 was designed to measure the differential priming effects of evaluatively moderate versus extreme primes. Although the food and generic words used as primes in Experiment 1 clearly differed in extremity, they also differed in other ways, such as semantically, with one set containing words referring to food (e.g., *broccoli, pancakes, cabbage, Spam*) and the other containing generic words (e.g., *pleasure, puppy, stress, leprosy*). Accordingly, we varied systematically the extremity of the primes, including one set of moderately valenced and one set of extremely valenced words. Because reverse priming effects have not been reported in other automatic evaluation studies, we retained all other features of Experiment 1, even though some variables (e.g., prime race and target extremity) are not essential to testing the current hypotheses.

Method

Overview. Participants were exposed to 640 experimental trials, each involving the rapid presentation of a prime word followed by a target. The task was to pronounce the target word quickly and accurately. Primes and targets were either positive or negative and were either stereotypically associated with Black American culture or White American culture, or were race neutral. Race-neutral words were either moderately or extremely valenced. Race words varied some in extremity, but generally fell soundly between the moderate and extreme sets.

Participants. Twenty-two Yale undergraduates (12 women, 10 men) were recruited and paid $6 each for their participation. Data from three African American participants were excluded from the analyses reported

below because, theoretically, they might exhibit a different pattern of results with regard to automatic race prejudice.[7]

Design. A 2 (prime valence: negative, positive) × 4 (prime type: moderate-race neutral, extreme-race neutral, Black-associated, White-associated) × 2 (target valence: negative, positive) × 4 (target type: moderate-race neutral, extreme-race neutral, Black-associated, White-associated) within-subjects design was employed.

Stimuli. The stimulus words used (see Appendix) were selected from two sources. The race-associated words were those used in Experiment 1. They were selected from among the race-related words generated most frequently by a sample of Yale undergraduates and were subsequently rated on valence and race-relatedness by separate samples from the same population. The positive and negative Black and White words were rated, on average, as equivalent in valence.

A new set of moderate and extreme race-neutral words was drawn from a dataset of words with normative ratings obtained by Bellezza et al. (1986). On the basis of the Bellezza et al. norms, words that were either near the extremes of the 7-point scale (from 1 = *extremely bad* to 7 = *extremely good*) or relatively close to the midpoint of the scale (4 = *neutral*) were selected. Extreme negative words were rated, on average, 1.7 (SD = 0.28); extreme positive, 6.2 (SD = 0.25); moderate negative, 3.36 (SD = 0.45); and moderate positive, 4.66 (SD = 0.33). All of the moderate and extreme words were nouns (as were the race words, though many of these were proper nouns) and were selected solely on the basis of valence and extremity. Effort was made to avoid confounding the valence and extremity variables with other features.

Procedure. Participants were presented with 640 successive word pairs. Each pair consisted of a prime word that was presented on a computer monitor for 100 ms. After a 50-ms interstimulus interval (blank screen), a target word was presented and remained on screen until it was pronounced. Prior to the appearance of each prime, a fixation stimulus ("+") appeared for 500 ms to orient the participant's gaze in the center of the screen, where the prime and target would subsequently appear. Between each trial there was a 1-s interval during which the screen remained blank. The 640 trials represent 10 repetitions of 64 conditions created by a full crossing of the four independent variables. The 40 different words in each category (20 negative and 20 positive) each appeared exactly eight times for each participant; four times as a prime and four times as a target. Primes were never paired more than once with the same target word, and although the same words appeared as primes and targets, a word was never paired with itself.

The 640 experimental trials were broken up into five sets of 128 trials. These sets were preceded by one set of 10 practice trials and an opportunity to ask questions before the experimental trials began. A rest period of 1 min was inserted between each set. Prior to the experimental trials in each set, six buffer trials, which serve to absorb error variance resulting from readjusting to the task, were added with additional positive and negative prime and target words. Each set contained two 64-trial blocks, and within each of these blocks, each condition (i.e., prime–target type and valence pairing) appeared exactly once. The pairings of the primes and targets as well as the order of the pairs was randomized by the computer program to create a unique ordering for each participant. The words appeared in white on a black background in a Helvetica font approximately one inch in height. MEL Pro, run on an IBM Aptiva personal computer, was the computer application used to execute the experiment. A microphone was attached through a Lafayette Instruments voice-activated relay box (Model 18010). Latencies to respond were recorded within 1-ms accuracy.

Participants were instructed to pronounce only the second word in each pair presented, quickly, clearly, and correctly. With regard to the prime, they were told only that the experiment was designed to measure how well people can make responses in the presence of distracting stimuli. Although the computer and software were not capable of recognizing the content of verbalizations offered by participants, a tape recording was made of each participant's vocal responses, and they were informed that the recording would be checked for accuracy. Participants completed the procedure alone in a quiet room. The procedure took approximately 30 minutes. All participants were thoroughly debriefed regarding the purpose of the experiment after being queried for their own hypotheses about it. No participants guessed the hypothesis being tested, and they typically reported that they had not attended to the primes. Despite the large number of trials, participants did not report much fatigue, which is corroborated by the fact that response times tended to be shorter over time.

Results

Data preparation. Because reaction time data are typically skewed in the positive direction, they must be transformed in order to conform to the assumptions of parametric techniques (Fazio, 1990; Ratcliff, 1993). A reciprocal transformation (i.e., dividing reaction times into one) yielded the closest approximation to a normal distribution. Furthermore, responses that were more than three of a given participant's standard deviations above or below her or his mean transformed reaction time were excluded from the analyses. These outlier response times typically reflect occasions when the participant is either already making noise (e.g., coughing) when the target appears, or when the response is not loud enough to trigger the mechanical sensor and must be repeated. In either case, the response does not generate meaningful data. Using these procedures, the percent of trials excluded from analyses ranged from 0.84% to 1.62% across the experiments, which is comparable to other similar studies. In order to test our hypotheses, we calculated mean latencies for each condition (i.e., type of prime–target pairing) for each participant and submitted them to repeated measures ANOVAs.

Automatic evaluation. We tested the hypothesis that the extremity of the primes would moderate the automatic evaluation effect, such that moderately valenced primes would yield the assimilative effects that are typical of evaluative priming studies, whereas extremely valenced primes would yield reverse priming effects. Specifically, we tested the interaction of prime type (moderate versus extreme) by prime–target congruence (congruent versus incongruent). We predicted that for moderate primes, the evaluatively congruent (i.e., positive–positive and negative–negative) prime–target pairs would yield faster responses than would incongruent (i.e., positive–negative and negative–positive) prime–target pairs. If the results of Experiment 1 were not an artifact of the particular stimuli used, evaluatively extreme primes should yield faster responses with incongruent prime–target pairs (i.e., a reverse priming effect).

The results confirmed the hypothesis that prime extremity determines the direction of priming.[8] With moderate primes the congruent pairs yielded the faster responses, $F(1, 18) = 42.31, p < .0001, r = .84$, reflecting the expected priming effect. With extreme primes, however, the incongruent pairs yielded the faster responses, $F(1, 18) = 19.2, p < .0001, r = .72$, replicating the reverse priming effect obtained in Experiment 1. The two-way interaction of prime type by prime–target congruence was also

[7] In fact, the pattern of data and significance levels was virtually identical with and without these participants.

[8] For Figure 3 (and in all subsequent figures depicting automatic evaluation) the targets are both moderate and extreme words collapsed together.

robust, $F(1, 18) = 35.13$, $p < .0001$, $r = .81$, confirming that the effects for moderate and extreme primes were reliably different. Figure 3, plotting the reaction times transformed back into milliseconds, illustrates this interaction, replicating the pattern obtained in Experiment 1.

It is worth noting that the effect is isolated to the extremity of the primes only. An analysis of the effect of target extremity (which was varied as well) reveals that target extremity does not moderate the automatic evaluation effect, $F < 1$. Furthermore, an analysis of the effect of prime extremity across the blocks of trials revealed that although there was some variability in the magnitude of the effect, there was no systematic pattern, $F < 1$. This indicates that the reverse priming effect does not result from a strategy that is developed (or attenuates) over the course of the procedure. It occurs immediately and persists throughout the procedure.

Automatic prejudice. We also predicted that consistent with past findings (e.g., Fazio et al., 1995), participants would exhibit automatic prejudice by responding faster to target words with Black associations when they were preceded by race-neutral negative primes than when preceded by race-neutral positive primes, and vice versa for target words with White associations. Based on the results from Experiment 1, we expected this pattern to hold only for trials with moderately valenced primes. When the primes were evaluatively extreme, on the other hand, responses should be faster to Black-associated words preceded by positive primes and White-associated words preceded by negative primes than to those preceded by negative and positive primes, respectively. Because the valence of the racial target words did not moderate the automatic prejudice effect (i.e., virtually identical results were obtained whether the Black and White targets were normatively negative or positive) in Experiment 1, and because it is the effect of prime valence that is of central importance here, we have collapsed across target valence for the race words. Consequently, in the following analyses, the Black and White target word categories are composed of both negative and positive words and, for the purposes of these analyses, vary only in terms of race.

As Figure 4 depicts, the automatic prejudice results obtained for food and generic primes in Experiment 1 are replicated with moderate and extreme primes.[9] With evaluatively moderate primes, responses were faster to negative-Black and positive-White prime-target pairs, $F(1, 18) = 6.73$, $p < .02$, $r = .52$. With the evaluatively extreme primes, the opposite result was obtained, $F(1, 18) = 72.57$, $p < .0001$, $r = .9$. Again, the two-way interaction, $F(1, 18) = 51.9$, $p < .0001$, $r = .86$, indicated that the effects of the moderate and extreme primes differed from each other reliably.

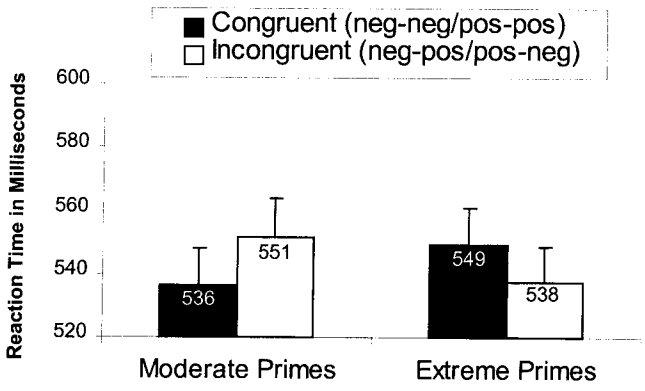

Figure 3. Automatic evaluation by prime extremity in Experiment 2. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

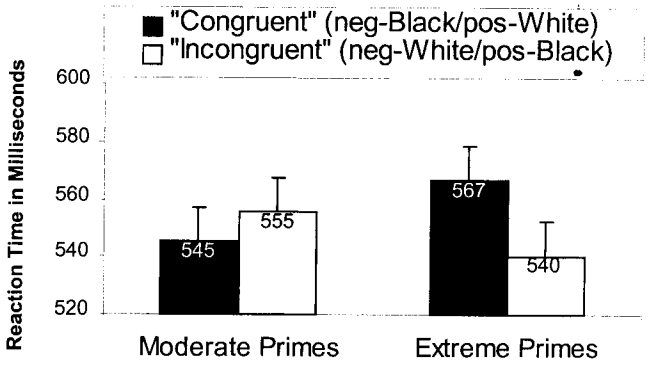

Figure 4. Automatic prejudice by prime extremity in Experiment 2. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

Discussion

The results of Experiment 2 provide verification that the extremity of primes dramatically influences automatic evaluation. This moderating effect is qualitative; the extremity of the primes does not simply determine the magnitude of the effect (i.e., larger effects with extreme primes), as one might intuit, but rather its direction. This result is informative in several ways. First, it is very robust, as indicated by the large effect sizes of both the interaction and the reverse priming effect itself. Second, the counterintuitive effect obtained in Experiment 1 is replicated, with stimuli deliberately selected to vary only on evaluative extremity, thus isolating extremity as a critical factor. Third, because the various conditions of the experiment (as with Experiment 1) are presented in random order (as opposed to separate blocks of moderate and extreme primes), the differential results for moderate and extreme primes is not likely due to participants' deliberate adoption of different response strategies. Finally, the effect appears to have some generality, occurring with race-associated as well as race-neutral targets. It is also noteworthy that we have obtained robust priming effects with words that are extremely subtle in their valence (the moderate words), thus supporting Bargh's (Bargh et al., 1992, 1996) position that automatic evaluation is a relatively universal phenomenon.

Experiment 3: Does Timing Matter?

The results of Experiments 1 and 2 demonstrate both normal and reverse priming in automatic evaluation and prejudice, thereby dramatically contradicting the findings of past studies of automatic

[9] In Figure 4 (and all subsequent figures depicting automatic prejudice effects) the target race words are collapsed across valence (e.g., Black words are both positive and negative).

evaluation wherein only assimilative effects have been reported. The most similar study is one conducted by Bargh and colleagues (1996) in which, as in the present studies, (a) a pronunciation task was employed and (b) both moderate and extreme primes were used. However, Bargh et al. did not obtain the reverse priming effect that we observed. In fact, it was the similarity of the effects across "weak" (i.e., moderate) and "strong" (i.e., extreme) primes that they cited in support of the argument that weak attitudes are activated as automatically as are strong attitudes.

There is, however, some precedent for reversed effects in automatic priming. Eimer and Schlaghecken (1998) presented symbols (e.g., arrows) for periods too brief to allow conscious perception and followed them with similar or different symbols, with an SOA of 118 ms. Participants indicated which type of symbol was presented by pressing the designated buttons. Eimer and Schlaghecken obtained reverse priming effects. That is, participants were faster to respond when symbols were incongruent than when they were congruent. Eimer and Schlaghecken's use of EEG readings to determine participants' lateralized readiness potential (LRP), which reflects the tendency to make one motor response or another, provided physiological evidence suggesting a corrective spike. The LRP charts revealed that participants had an initial response tendency to the subliminal primes that was consistent with the primes, but a subsequent tendency to respond in the opposite manner. This latter tendency typically coincided with the response to the target stimulus, thus yielding slower responses to congruent prime–target pairs.

An examination of the Eimer and Schlaghecken (1998) procedure may provide some insight regarding the discrepancy between our findings and those of Bargh et al. (1996). One potentially nontrivial difference in the procedures from Bargh et al. (1996) and our Experiments 1 and 2 is the SOA, which was 300 ms for Bargh et al. (1996; as well as for other automatic evaluation studies, e.g., Bargh et al., 1992; Fazio et al., 1986; and Hermans et al., 1994) and 150 ms in Experiments 1 and 2. It is possible that the correction that creates the reverse priming effect with the extreme primes occurs only briefly, as in Eimer and Schlaghecken's experiments, appearing as an aberrant spike in an otherwise assimilative response tendency. If the target were presented early (i.e., with a relatively short SOA such as 150 ms), around the time of the correction spike, the response to the target would be affected by the temporary disposition corresponding to the correction. A longer SOA (e.g., 300 ms) may capture the response tendency engendered by the prime after the correction spike has reverted. Recall that Eimer and Schlaghecken, who obtained contrast effects, employed an SOA of 118 ms, which is much closer to 150 than to 300. It is not clear at this stage why a correction would be so short-lived, unless a countercorrection ensues, like that of a driver making a series of diminishing corrections to regain his or her trajectory after swerving to avoid an object in the road. Nevertheless, Eimer and Schlaghecken's findings are compelling and seem to be relevant, and so an investigation of the role of SOA (at least insofar as it differs from that used in important, comparable studies) is warranted.

Further evidence exists for the importance of SOA in demonstrations of automatic evaluation from the work of Klauer and colleagues (1997) who, using an evaluative decision (key press) response, found evidence for evaluative priming at some SOAs (0 and 100 ms[10]), but not at others (−100, 200, 600, and 1,200 ms[11]).

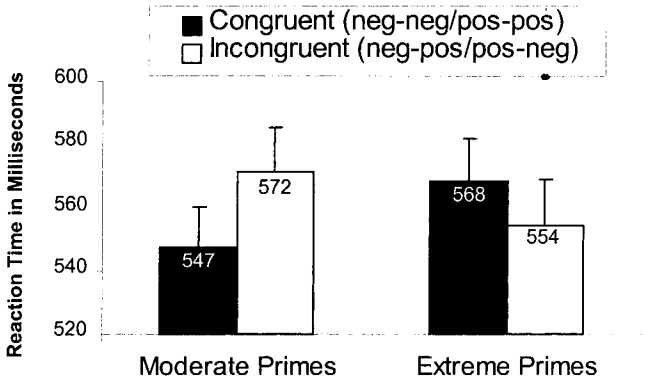

Figure 5. Automatic evaluation by prime extremity in Experiment 3. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

In fact, at some SOAs (−100 and 600 ms) they reported small (5 ms), nonsignificant reverse priming effects but, reasonably enough, did not consider them meaningful. SOA may then play an important role in determining whether priming effects and reverse priming effects (and the correction processes that yield them) are detected. On the other hand, if the disposition resulting from a correction is relatively stable, we should expect to obtain reverse priming effects regardless of the time course of the stimulus presentation. Accordingly, Experiment 3 was designed as an exact replication of Experiment 2, with the exception that the SOA was revised to 300 ms, exactly that used in the most comparable evaluative priming studies where only assimilative effects were obtained (e.g., Bargh et al., 1992, 1996; Fazio et al., 1986; Hermans et al., 1994).

Method

The design, apparatus, stimuli, and procedures for Experiment 3 were identical to those employed in Experiment 2, with the critical exception that the SOA was extended from 150 to 300 ms. In each trial, the prime was presented for 200 ms, followed by a 100-ms interstimulus interval (ISI), after which the target appeared. Seventeen Yale undergraduates (10 women, 7 men) were paid $6 each to participate.

Results

The design of Experiment 3 did not differ from that of Experiment 2, and so the same analyses were conducted, including adjustment for the positively skewed distribution of reaction time data.

Automatic evaluation. As Figure 5 depicts, moderate and extreme primes again produced dramatically discrepant patterns of automatic evaluation. With moderate primes, participants were faster to respond to targets that were preceded by evaluatively

[10] With a 0-ms SOA, the prime and target are presented simultaneously, one positioned above the other, and participants respond to the one appearing in a particular position or color.

[11] With negative SOAs, the response is made to the first stimulus that appears. In the case of a short SOA (e.g., −100 ms), the "prime" still appears before the response is made.

congruent primes, $F(1, 16) = 41.64$, $p < .0001$, $r = .85$. Conversely, participants were slower to respond to congruent prime–target pairs on trials with evaluatively extreme primes, $F(1, 16) = 18.39$, $p < .001$, $r = .73$. The test of the higher order interaction revealed that prime extremity dramatically reversed the automatic evaluation effect, $F(1, 16) = 83.83$, $p < .0001$, $r = .92$. These findings fully replicate those of Experiments 1 and 2.

Automatic prejudice. As shown in Figure 6, prime extremity also clearly moderated the relation between prime valence and target race. Consistent with the hypothesis of automatic prejudice, participants were faster to respond to negative-Black and positive-White prime target pairs than negative-White and positive-Black pairs when the primes were moderately valenced, $F(1, 16) = 19.52$, $p < .0005$, $r = .74$. However, once again the extreme primes led to the opposite result, $F(1, 16) = 46.37$, $p < .0001$, $r = .86$. Finally, the two-way interaction test confirmed that prime extremity qualitatively moderated evaluative priming, $F(1, 16) = 60.38$, $p < .0001$, $r = .89$. Again, these findings replicate those obtained in Experiments 1 and 2, as well as the general automatic evaluation effect (i.e., the interaction of prime and target valence).

Discussion

From Experiments 2 and 3, it is now clear that the contrast effect observed in Experiment 1 is not a function of the idiosyncratic stimuli or SOA employed. In fact, the pattern of results across Experiments 1, 2, and 3 is remarkably consistent despite the changes in stimuli and presentation timing. In all three experiments the automatic evaluation effect obtained in previous research was replicated with moderate primes, and reverse priming effects resulted when the primes were extreme. Experiment 2 demonstrated that with entirely different sets of evaluatively moderate and extreme primes (words selected, a priori, to vary on that dimension), normal and reverse effects, respectively, are nevertheless obtained. This provided strong evidence that prime extremity is a critical factor. Experiment 3 showed that when an SOA of 300 ms used in previous experiments is introduced, a reverse priming effect is still obtained with extreme primes. This result suggests that the corrections elicited by extreme primes are not likely to be short-lived spikes that revert in less than 300 ms. As the Eimer and Schlaghecken (1998) LRP readings indicate, there may first be an assimilative response tendency that is then reversed, but this reversal probably does not reflect a brief, indecisive spike. To the contrary, it appears that the correction is relatively stable and that the resulting reverse priming effect is robust and replicable across different conditions. Furthermore, virtually every participant shows the effect, hence the highly statistically significant results.

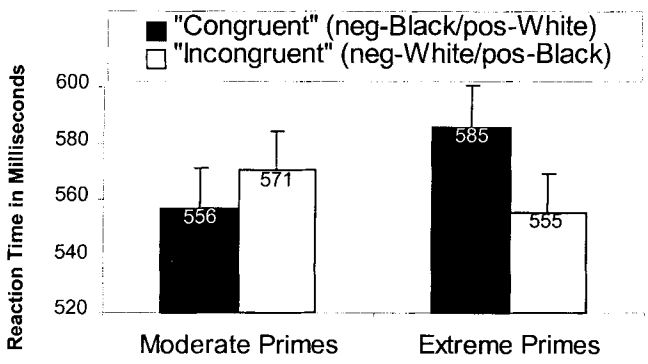

Figure 6. Automatic prejudice by prime extremity in Experiment 3. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

Experiment 4: Does Race Matter?

Despite the clarity and consistency of our findings, it is not evident why they contradict the results of past studies of automatic evaluation, particularly those involving similar procedures (e.g., Bargh et al., 1996; Hermans et al., 1994). There are several possible reasons why the procedures employed in the present studies may have allowed for the detection of the reverse priming effect when others have not. Differential extremity, we believe, is not a plausible explanation for the difference. The extreme primes employed by Bargh et al. (1996; e.g., *friend, holiday, cancer, funeral*) were comparable to ours (e.g., *blossom, peace, abuse, poison*). Like Bargh et al. (1996), we had participants rate the generic words used in Experiment 1 on an 11-point scale from −5 to +5. The mean extremity of our negative and positive generic primes (−3.85 and +3.7) is virtually identical to theirs (−3.75 and +3.65).[12]

Alternatively, we might compare the size of our stimulus sets with those of Bargh et al.'s (1996). In fact, there is a substantial difference here. We had 20 primes of each type, whereas Bargh et al. had only four of each type. It is also noteworthy that the primes and targets used by Bargh et al. were drawn from different sets. Is it possible that these differences account somehow for the striking discrepancy in results? For example, perhaps the knowledge that primes may appear as targets and vice versa causes participants to take greater pains to inhibit the primes, as in the negative priming effect (May, Kane, & Hasher, 1995). We will not feast here on such conjecture because we know from the existing data that these differences are not important. As reported in Experiment 2, the reverse priming effect emerges in the very first block of trials (and this is consistently the case across experiments) before words have appeared as both primes and targets, and when a smaller subset of the stimulus set has been presented. The first block more closely resembles those of the Bargh et al. experiments with respect to the overlap and repetition of primes and targets, and so these factors are unlikely to account for the reverse priming effects.

Racial Stimuli

One major difference between the present experiments and those in the most comparable study (Bargh et al., 1996) is that we included racial stimuli whereas they did not. The presence of words like *Harlem, homeboy, Nazi,* and *skinhead* in our procedure may have heightened participants' sensitivity to evaluating the stimulus words, instigating attempts to neutralize evaluations of the words, which resulted in overcorrections. As reported in Fig-

[12] Extreme words used in subsequent experiments were selected based on 7-point scale ratings (Bellezza, Greenwald, & Banaji, 1986), but mathematical conversion reveals the same compatibility with Bargh, Chaiken, Raymond, and Hymes (1996) stimuli.

ures 1, 3, and 5, we obtained reverse priming effects with extreme primes (and assimilative effects with moderate primes) even on trials where there were no race words. Thus, it is not necessarily the presence of a race word in a given trial, so much as the mere notion that race words had appeared and were likely to appear again that may have affected the manner in which participants reacted to the race-neutral primes.

On the other hand, it is possible that reverse priming effects in automatic evaluation are more commonplace, at least when a pronunciation task is employed. In Herr et al.'s (1983) demonstrations of contrast effects as a function of prime extremity, judgments were made of animal size and ferocity (in separate experiments). There were no racial stimuli, or any other known features that might have heightened participants' sensitivities. Consequently, it is within reason to expect that reverse priming effects will be obtained in automatic evaluation in the absence of racial stimuli. If the reverse priming effect is not obtained (with extreme primes) in the absence of racial stimuli, it will implicate the presence of racial stimuli as a critical antecedent of the observed reverse priming effect. On the other hand, if the reverse priming effect is obtained without racial stimuli present, we can have greater confidence that automatic correction is a relatively fundamental feature of unconscious evaluation, and we will have to look elsewhere to explain the discrepancy with Bargh et al.'s (1996) findings.

There is also another possible outcome of the removal of racial stimuli to consider. It is plausible that the racial stimuli do not serve to inspire correction, but do nevertheless serve to increase the salience of the evaluative nature of the stimuli. This could explain the unusual magnitude of the effects for both moderate and extreme primes. Such increased salience could enhance the ability of even the modestly valenced moderate primes (e.g., *chair, comparison, moment, patent*) to activate automatically an evaluative response. If this is the case, the moderate words will not yield automatic evaluation effects in the absence of racial stimuli.

Method

Participants. Thirty-three Yale undergraduates (17 women, 16 men) were paid $6 each for their participation. One participant indicated that her vision was not normal or corrected to normal and was, accordingly, dropped from the data analyses, as was another who indicated that English was not her primary language (this question was asked to ensure that participants would be able to process English words automatically). In this experiment, African American participants (there were three) were included in the analyses because without racial stimuli there was no reason to expect that they would respond differently.

Design. Because race primes and targets were not used, the design of the study was considerably simpler than that of the previous experiments. With the exclusion of the racial stimuli, a 2 (prime valence: negative, positive) × 2 (prime extremity: moderate, extreme) × 2 (target valence) × 2 (target extremity) within-subjects design resulted.

Stimuli. The primes and targets were the same moderate and extreme race-neutral words employed in Experiments 2 and 3 (see Appendix).

Procedure. The procedure was identical to that of the previous experiments except that with the exclusion of prime and target race as variables, only 160 experimental trials (10 per condition) were necessary. These were broken into two blocks of 80 trials, each with practice and buffer trials included as in Experiments 1, 2, and 3. Within each block, each condition (prime–target type pairing) was presented five times. Participants had a 1-min break between blocks.

Results

The abnormal reaction time data distribution was treated in the same manner as previously. Because there were no racial stimuli, only the automatic evaluation hypotheses (i.e., that responses will be faster to targets preceded by evaluatively congruent than incongruent primes) can be tested. As Figure 7 illustrates, evaluative priming did not occur at all with the moderately valenced primes ($F < 1$). However, with the evaluatively extreme primes, the reverse priming effect, wherein participants are faster to pronounce targets preceded by incongruent primes than those preceded by congruent primes, was replicated, $F(1, 30) = 11.44$, $p < .005$, $r = .52$. The two-way interaction testing the moderating role of prime extremity was significant, $F(1, 30) = 6.99$, $p < .05$, $r = .43$, but in this case this interaction reveals only that one effect is larger than the other—specifically, that one effect is present while the other is entirely absent. There is no reversal of patterns as in the previous experiments.

Discussion

Clearly, we can conclude from this experiment that reverse priming effects with extreme primes are not dependent on the presence of the racial stimuli, although their magnitude does appear to be affected by it. This result is most important given the goals of this study to demonstrate automatic correction and determine its moderators. The question regarding the disparity between the present findings and those of others, however, remains unanswered.

However, it appears that the presence of the racial stimuli does have an impact of a sort on automatic evaluation. It is not the cause of the reverse priming effects, because they have been evidenced now regardless, but from the results of this experiment we might conclude that the presence of racial stimuli, rather than instigating a correction, serves to heighten the salience of the evaluative aspect of the primes. This would be likely to enhance both assimilation and contrast effects. In the absence of the racial stimuli, the moderate primes seem to have lost their power to elicit evaluative responses. This result is not entirely surprising, given that these words are indeed very moderate, deviating only subtly from the neutral point on the evaluative rating scale.

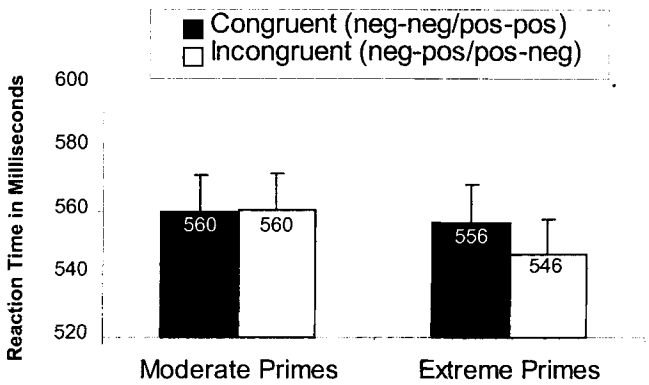

Figure 7. Automatic evaluation by prime extremity in Experiment 4. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

Experiment 5: Maybe Some Other Time

Experiments 1–4 have provided consistent evidence that evaluatively extreme primes yield reverse priming effects (i.e., faster responses to evaluatively incongruent targets than to congruent targets). And yet, similar studies (e.g., Bargh et al., 1996; Hermans et al., 1994) have reported only assimilative effects. As Experiment 4 revealed, the presence of racial stimuli does not account for the reverse priming effects. Experiment 3 showed that another procedural variant (stimulus onset asynchrony) is also not responsible for the difference in results from past studies. There is, however, another timing-related difference between the present experiments and those in which reverse priming effects have not been obtained, and this difference should also be addressed.

Because the design of Experiment 1 involved 64 conditions and consequently 640 experimental trials, we took steps to shorten the task so that it would take a reasonable amount of time. For example, we decreased the intertrial interval (ITI), the time between the participant's response to the target and the presentation of the next stimulus, from the four seconds used in previous studies to one second.[13] This seemingly trivial difference shortened the procedure by a full 32 minutes. Although four seconds, when considered out of context, may not seem like a very long time, when one is awaiting the appearance of a stimulus on a blank screen, it is a subjectively long period (hence the lengths the computer industry goes to in order to develop increasingly faster microprocessors). The relatively short ITI (1 s) we have employed may change the processing styles and strategies of the perceiver. This proposition is supported by results obtained by DeCarlo (1992), who found that longer ITIs decreased the influence of prior stimuli on the magnitude estimation of a target stimulus. A short ITI may make it considerably more difficult for the perceiver to distinguish between the primes and the targets; they may appear to flow more in a continuous stream than a series of pairs. With a longer ITI, the perceiver may have ample time to reset after his or her response to the target, preparing to ignore the prime. More desperate attempts, such as the correction that would yield a reverse priming effect, might be instigated by the short ITI. Or more passively, a longer ITI may make it more difficult for the participant to anticipate the onset of the prime. If ITI affects processing in any of these ways, the addition of time between trials could mitigate or even reverse the previously obtained reverse priming effects.

On the other hand, if the reverse priming effect reflects an attempt to correct for the influence of the prime on the response to the target, a short ITI could function to increase cognitive load, thereby making such a correction more difficult. A longer ITI could simplify the task, giving the participant more time to prepare for the next trial and counteract the effects of the prime. If this is the case, we would expect to see reverse priming effects persist with a longer ITI, perhaps with even greater magnitude.

In order to determine if our short ITI (1 s) was necessary to obtain reverse priming effects in automatic evaluation, we conducted an experiment employing a 4-s ITI, as used in previous demonstrations of automatic evaluation wherein reverse priming effects were not obtained (e.g., Bargh et al., 1996). This experiment also allows us to assess whether the failure to obtain priming effects with moderate primes in Experiment 4 was a Type II error or replicable evidence for the conditionality of automatic evaluation (Fazio et al., 1986). If priming effects are found with moderate primes, it would suggest that their absence in Experiment 4 was either a fluke or an idiosyncratic consequence of the 4-s ITI. If on the other hand, they are not obtained in Experiment 5, it would support Fazio's claim that attitude strength predicts activation.

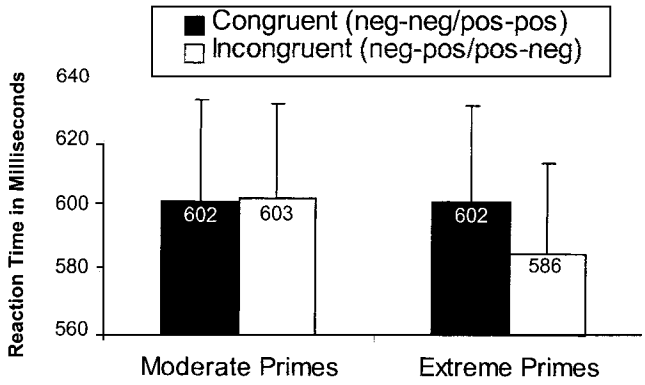

Figure 8. Automatic evaluation by prime extremity in Experiment 5. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

Method

Participants. Fifteen Yale undergraduates (12 women, 3 men) were paid $6 each for their participation.

Design. As in Experiment 4, a 2 (prime valence) × 2 (prime extremity) × 2 (target valence) × 2 (target extremity) factorial design was employed.

Stimuli. The same stimuli as those in Experiment 4 were used.

Procedure. The procedure was, in all respects, identical to that of Experiment 4, except that the ITI was increased from 1 to 4 s.

Results

The data were transformed and outliers removed following the procedures in Experiments 1–4. A repeated measures ANOVA was conducted to test the effects of prime extremity and prime–target congruence on latency to pronounce the target words. The mean reaction time latencies per condition (converted back into milliseconds) are presented in Figure 8. Note that as with Experiment 4, only the automatic evaluation-related hypotheses can be tested because no racial stimuli were included in the experiment. As in the previous experiments, the two-way interaction was significant, $F(1, 14) = 10.79$, $p < .01$, $r = .66$, indicating that prime extremity moderates the evaluative priming effect such that prime–target congruence does not matter with moderate primes ($F < 1$) while a reverse priming effect (i.e., faster responses to evaluatively incongruent pairs) was obtained for extreme primes, $F(1, 14) = 15.74$, $p < .005$, $r = .73$.

Discussion

The results of Experiment 5 provide further evidence for reverse priming effects in automatic evaluation with extreme primes, even

[13] It is very important to avoid confusing ITI (intertrial interval) with ISI (interstimulus interval). The latter is the time between the offset of the prime and the onset of the target within a given trial.

though the intertrial interval had been lengthened considerably. This experiment offered a fairly direct replication of Bargh et al. (1996), and yet we have again obtained contrast effects with evaluatively extreme primes and, as in Experiment 4, no effects with evaluatively moderate primes. Despite the procedural similarities, our results differ dramatically from those of Bargh et al. (1996). The replication of Experiment 4's null result with moderate primes diminishes the likelihood of a Type II error and lends further support to the hypothesis that the presence of racial stimuli in Experiments 1, 2, and 3 heightened the evaluative salience of the primes.

We postulated two possible effects of ITI on priming. The first was that the extension of the ITI from one to four seconds would make the task less confusing, enabling participants to better differentiate trials and thereby alleviating the necessity to correct excessively for the influence of the prime, which could now be more effectively ignored. The second hypothesis also held that the extension of the ITI would simplify the task, but that this would have the effect of reducing cognitive load and thereby enable participants to expend the resources required to monitor for potentially biasing primes and make the requisite corrections. The data appear to favor the second hypothesis. The reverse priming effect was obtained despite the increase in ITI. Furthermore, comparing the reverse priming effect size for Experiments 4 and 5 (r = .52 versus r = .73), it appears to be more robust with the 4-s ITI, although such comparisons across experiments can be made only speculatively. Issues of magnitude aside, Experiment 5 provides a clear replication of Experiment 4. An explanation for the discrepancies with Bargh et al. (1996) will have to be found elsewhere.

Experiment 6: Fixation

We have eliminated several procedural variables (SOA, ITI, and the presence of racial stimuli) that appeared to have the potential to account for the differences in results obtained herein versus those of Bargh et al. (1996). There remains one more potentially significant, identifiable difference that ought to be investigated. In our experiments, we presented a fixation point to help orient the participants' gazes where the primes and targets would appear. The fixation point was a plus sign ("+") that appeared in the same location just prior to each prime for 500 ms. Our intention in including this stimulus was to maximize the likelihood that participants would see the primes so that we could detect automatic evaluation and prejudice. Such fixation stimuli have not been used in other demonstrations of automatic evaluation (e.g., Bargh et al., 1992, 1996; Fazio et al., 1986).

This procedural discrepancy may seem trivial, but it stands to reason, and in fact is consistent with our original rationale for including the fixation point, that such a warning stimulus could serve to enhance the participants' ability to process and perhaps counteract the effect of the prime. In fact, part of the rationale for extending the ITI in Experiment 5 was that the greater length of time between trials may make the prime appear as more of a surprise, thereby undermining attempts to correct for it. If this were the case, the fixation point would negate such an effect, snapping the participant's attention into focus prior to the prime's appearance. In order to test whether the use of the fixation point is a necessary condition for reverse priming effects, we replicated Experiment 5 exactly, with the exception that we manipulated the presence of the fixation point.

Method

Participants. Thirty-eight University of California, Berkeley, undergraduates (31 women, 7 men) participated in exchange for partial course credit.[14] Two participants (both women) indicated that English was not their primary language and were consequently excluded from the data analyses.

Design. A 2 (fixation point: present, absent) × 2 (prime valence) × 2 (prime extremity) × 2 (target valence) × 2 (target extremity) mixed factorial design was employed with fixation point as a between-subjects variable.

Stimuli. The same stimuli as those in Experiments 4 and 5 were used.

Procedure. For half of the participants, the procedure was identical to that of Experiment 5. Of particular interest here is the presentation for 500 ms of a fixation point ("+") immediately prior to the presentation of each prime. For the other half of the participants, the procedure differed in that there was no fixation point presented. For both groups the SOA remained 300 ms and the ITI was 4 s. The fixation point does not count toward the ITI. It should also be noted that this experiment differed from previous experiments in that no tape recording was made of participants carrying out the word pronunciation procedure.

Results

The data were transformed and outliers removed following the procedures in Experiments 1 through 5. A repeated measures ANOVA was conducted to test the effects of fixation point, prime extremity, and prime–target congruence on latency to pronounce the target words. The mean reaction time latencies per condition (converted back into milliseconds) are presented in Figure 9. As in all the previous experiments, prime extremity moderated the automatic evaluation effect (the difference in response latency between evaluatively congruent and incongruent prime–target pairs), $F(1, 35) = 4.25$, $p < .05$, $r = .33$. As in Experiments 4 and 5, wherein there were no racial stimuli, there was no automatic evaluation effect for moderate primes ($F < 1$), but consistent with all the preceding experiments, a reverse priming effect was obtained with extreme primes, $F(1, 35) = 18.37$, $p < .0001$, $r = .59$. With regard to the between-subjects variable, fixation point, there was a nonsignificant main effect of fixation point, $F(1, 34) = 2.27$, $p < .15$, that although relatively large in milliseconds (34 ms), is unstable. The presence or absence of a fixation point did not moderate the effect of extremity on automatic evaluation ($F < 1$), nor did it moderate the separate priming effects for moderate or extreme primes (Fs < 1); there were no reliable effects with moderate primes (Fs < 1) and the reverse priming effect with extreme primes held for participants with ($F(1, 17) = 11.29$, $p < .005$, $r = .63$) and without ($F(1, 17) = 7.57$, $p < .05$, $r = .56$) fixation points.

[14] Although there is no compelling reason to expect the Yale and Berkeley undergraduate populations to differ with regard to automatic evaluation or reverse priming, it should allay any such concerns to note that we did replicate Experiment 3 with a Berkeley sample and obtained identical results.

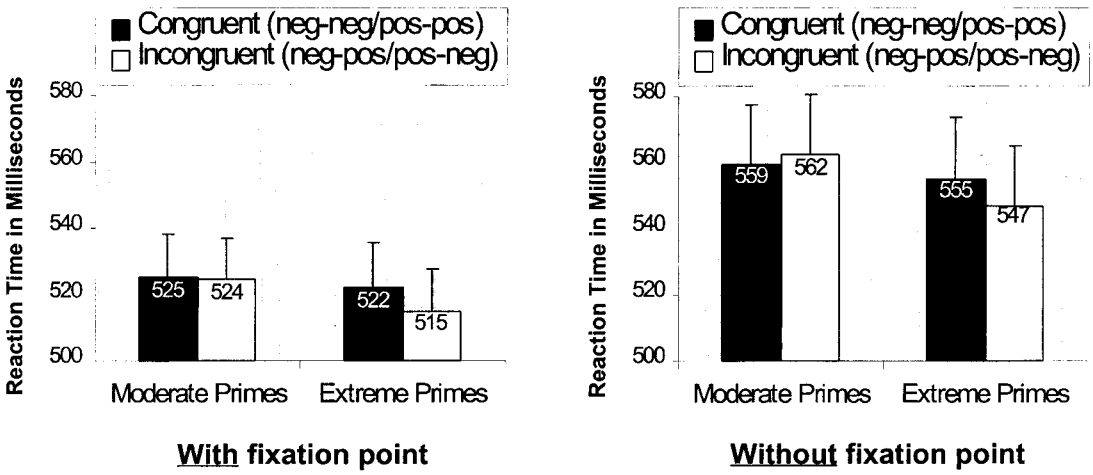

Figure 9. The effect of fixation point and prime extremity on automatic evaluation in Experiment 6. Neg = negative; pos = positive. Error bars represent one between-subjects standard error.

Discussion

As with the other procedural variables (SOA and ITI), the use of a fixation point does not account for the reverse priming effect. Although the main impetus for testing for an effect of fixation point was the fact that it had not been used in previous studies wherein reverse priming had not been observed, we did propose several mechanisms by which the fixation point would engender reverse priming. However, it appears that the focusing effect of the fixation point is not necessary to enable the kind of process that yields reverse priming.

As in the two preceding experiments, no priming effect was observed with moderate primes, again contradicting the position (Bargh et al., 1992, 1996) that automatic evaluation is a universal process that will happen under minimal circumstances, regardless of strength of attitude. Most important, this experiment has confirmed that the reverse priming effect appears to be more than merely a procedural artifact, but rather a remarkably persistent and stable phenomenon.

General Discussion

The results of these experiments provide strong corroborating evidence for the automaticity of evaluation. Specifically, the speed with which people pronounced target words was clearly affected by these words' evaluative congruence with preceding primes. Similarly, these experiments demonstrate the automaticity of prejudice. Participants were faster to read Black- and White-associated target words that were preceded by negative and positive race-neutral primes, respectively. This occurred under strict conditions of automaticity (i.e., precluding conscious, deliberate control of responses): The time between onset of prime and target was well within the established criterion for automaticity (Neely, 1977), and the dependent variable (latency to pronounce words) was unobtrusive.

Another important result of the present study is that the presence of racial stimuli appears to enhance the salience of the evaluative nature of the task. In the first three experiments, where half of the primes and targets had fairly obvious semantic associations with African American or European American subcultures, large automatic evaluation effects were observed with remarkably moderately valenced primes. However, in the latter three studies, when such racial stimuli were absent, so were the automatic evaluation effects with moderate primes. This is interesting in and of itself because it suggests that a motivational state (e.g., an apprehension about evaluating racial stimuli) will affect automatic processing. This result also has bearing on the debate over the conditionality of automatic evaluation. Fazio and colleagues (Fazio, 1993; Fazio et al., 1986) have argued that the strength of the attitude toward an object (e.g., a prime word) will moderate the speed and likelihood of activation of that attitude and that weak attitudes need not be automatically activated at all. In contrast, Bargh and colleagues (Bargh et al., 1992, 1996; Chaiken & Bargh, 1993) contend that automatic evaluation happens unconditionally and equally, for attitudes of all strengths. The present findings suggest a middle ground wherein all attitudes are not necessarily and unconditionally activated with the same speed and force, but that under the right conditions (e.g., when the salience of evaluation is increased) even the weakest of attitudes can be activated automatically.

Most importantly, in one particularly striking way, the results of the present experiments did not conform to those typically obtained in studies of automatic evaluation, or really semantic priming studies in general. Whereas previous studies (e.g., Fazio et al., 1986; Bargh et al., 1992, 1996) have reported only assimilative effects (i.e., responses are fastest to evaluatively congruent prime–target pairs), we found that when the priming stimuli were evaluatively extreme, responses were fastest for targets preceded by evaluatively incongruent primes. The present results are analogous to contrast effects obtained in studies of controlled judgments, and most notably those studies wherein extreme primes elicit contrast effects (Herr, 1986; Herr et al., 1983). Consistent with past studies of controlled judg-

ments, when primes were evaluatively moderate, we obtained assimilative effects or none at all.

Comparing the present findings to past research on assimilation and contrast should be done with caution. Studies of assimilation and contrast in priming, even when the priming has been implicit and unobtrusive or both, have typically employed quantitative judgments (e.g., ratings) as their dependent variables. These judgments are affected either in the direction of the prime (assimilation) or away from it (contrast). In the present experiments, however, the dependent variable was reaction time. One cannot conclude that slower responses to congruent prime–target pairs reflect a "contrast" effect in the same sense that judgments of magnitude would. In the present case, it is unlikely that the target is being "contrasted" against the prime, and the reaction time measure would not speak to that even if it were the case. Nevertheless, the similarities to past priming studies, especially in terms of the role of prime salience in reversing the direction of priming effects, warrant a consideration of the theories applied to such effects.

There are two commonly posited mechanisms by which prime salience can lead to contrast effects: comparison–contrast and correction (Stapel et al., 1998). As noted above, comparison–contrast cannot likely explain the present results. In order for that to be the case, the valence of the target would first have to be compared to the prime, and accordingly adjusted, and then the automatic evaluation effect (i.e., the facilitative or inhibitive effects of a congruent or incongruent prime on the evaluation of the target) would occur on the re-evaluated target. Such a complex, multistage, serial process would likely require more time overall, but the overall reaction time for trials with moderate versus extreme primes do not differ. Furthermore, not only is this explanation awkward and unparsimonious but it also predicts that all targets would be contrasted by extreme primes (those that are similar and different in valence alike). This state of affairs would yield "slow" responses to all targets following extreme primes of either valence, not just those of similar (initial) valence to the prime.

A simpler test of the comparison–contrast explanation for reverse priming is possible by examining the effect of target extremity. Recall that target words in these experiments were drawn from the same categories as were the primes, so there were moderate and extreme targets. Post hoc analyses revealed that target extremity did not qualitatively moderate the reverse priming effects with extreme primes (or the normal priming effects with moderate primes). The reverse priming effects were larger with moderate targets than with extreme, but the effect was still obtained with extreme targets. And yet research on assimilation and contrast has shown that comparison–contrast effects resulting from prime extremity will occur only when the targets of judgment are moderate in magnitude relative to the primes (Herr et al., 1983). If the prime and the target are of similar magnitude, the target cannot be "contrasted" against the prime (e.g., a large target will not seem smaller in the presence of an equally large prime). In the present case the primes and targets were drawn from the same pool of words and so, necessarily, the extreme primes and targets were equivalent. Therefore, comparison–contrast is unlikely to explain these results.

Automatic Correction

We propose that the reverse priming effects observed reflect not comparison–contrast, but the alternative mechanism for contrast effects: correction. This correction is instigated by the perceived potential of the peripheral prime to bias the response to the intended target. When a prime appears for which evaluation is particularly salient (vis à vis its extremity), the perceiver is all the more likely to recognize, quite unconsciously and automatically, its potential to bias the intended judgment of the target (Stapel et al., 1998; Strack et al., 1993). This, in turn, instigates a corrective or compensatory process (Stapel et al., 1998; Strack, 1992; Strack & Hannover, 1996).

If the strategy employed were simply to disregard or actively suppress the prime, we would see no differential effect for congruent versus incongruent prime–target pairs. However, the consistently obtained reverse priming effects indicate that an unconscious but active correction is taking place. As is the case with more deliberate judgments (e.g., Stapel et al., 1998), the correction is excessive (an overcompensation) and therefore results in a reversed effect. Specifically, extreme negative primes have the effect of positive primes (i.e., slower responses to negative than to positive targets) and vice versa for extreme positive primes. Previous findings implicate accuracy motivation as a mediator of contrast effects (e.g., Ford & Kruglanski, 1995; Martin et al., 1990; Stapel et al., 1998). A contribution of the present results is the suggestion that accuracy motivation, manifested in an attempt to avoid bias, can operate automatically and affect other automatic processes.

Typically, theories of knowledge accessibility and priming assume that assimilation effects reflect automatic processes, occurring without awareness and beyond control (Wilson & Brekke, 1994). It is the more deliberate, conscious attempts to attenuate priming that are theorized to lead to contrast effects (Lombardi, Higgins, & Bargh, 1987). The reverse priming effects reported here, however, cannot have resulted from deliberate, conscious processes. Participants could not have intentionally slowed their response times by an average of 10–20 ms when extreme primes were followed by similarly valenced targets. Furthermore, a conscious mediation explanation presupposes that the participants were aware that their latency to pronounce the target words was affected by the evaluative congruence of the prime and target—also highly untenable and not borne out in lengthy discussions with participants during debriefing.

Because the activation of the primed concepts under the present conditions is automatic (viz. Bargh et al., 1992, 1996; Fazio et al., 1986; Neely, 1977), it lies outside participants' control. It is perhaps time to entertain a new idea: Attempts to correct for priming can occur automatically. This does not preclude the possibility that conscious goals to respond accurately influence, and perhaps instigate, unconscious attempts to correct, but it does suggest that the motivation to avoid biased responding (an accuracy motivation) is capable of operating at the unconscious level as well. The idea of unconscious or automatic motivation is relatively new to psychological science (Bargh, 1990, 1997; Bargh & Barndollar, 1996; Kihlstrom, 1987), and empirical evidence for it is only now emerging. Recent studies have provided groundbreaking evidence for motivational effects on unconscious processes (Bargh

& Gollwitzer, 1994; Chartrand & Bargh, 1996).[15] Such studies have demonstrated that goals can be activated automatically (e.g., through subliminal presentation) and subsequently affect the manner and result of judgments and behaviors. These experiments have employed relatively deliberative judgment tasks and overt behaviors as dependent variables. The use of implicit reaction time measures in the present experiments (not to mention the unobtrusive nature of the pronunciation task) extends these findings, allowing for the conclusion that unconscious motives can influence purely automatic responses as well. In other words, as Bargh (1990, 1997) has theorized, a complete and complex mental response to a stimulus—from the goals activated with regard to that stimulus, to the cognitive or affective response, to the resulting behavior—can transpire automatically without conscious mediation or even awareness.

Alternative Explanations

Because the reverse priming effect was not initially predicted, and therefore explanations for it are necessarily post hoc, it is important to be particularly vigilant in further demonstrating the effect under varying conditions, and ruling out alternative explanations. Our a priori replications and procedural manipulations, as well as the robustness and consistency of the effect, go a long way toward ruling out Type I error and narrowing the pool of potential explanations. We have provided reasons above for dismissing comparison–contrast, as opposed to correction, as an explanatory mechanism. We now consider a few other explanatory candidates.

Phonology. It is conceivable that the pattern of results obtained may be an artifact of the similarities in sounds between primes and targets.[16] After all, it is rather remarkable, and for some incredible, that the automatic evaluation effect can be detected with as distal a dependent variable as pronunciation latency. In fact, at least one laboratory has demonstrated repeated failures to replicate it (Klauer & Musch, 1998). There is also reason to believe that phonemic similarities between primes and targets could affect latency to pronounce. The idea has intuitive appeal but is also supported by research on auditory perception (Shoaf & Pitt, 1998). Specifically, Shoaf and Pitt have shown that latencies to pronounce words are faster when they are preceded by word primes that have similar sounds in the first or second syllables, but longer when the third syllables are similar. Although this research was carried out with auditory stimuli, it suggests that word sound similarity can influence pronunciation latency, and not always in a facilitative manner.

For several reasons, we are certain that our results are not attributable to similarities in sounds between primes and targets. First, the probability that this unfortunate artifact infected two independent samples of race-neutral words (the food and generic, and then the moderate and extreme word sets) is low. Second, we examined the lists of moderate and extreme words (see Appendix) to identify the number of pairs sharing first sounds (e.g., *gloom–glory, revolt–respect*) and found the numbers for each potential list pairing (e.g., moderate–negative with extreme–positive) to be remarkably low, ranging from 1 to 5 (M = 2.7) out of the 400 possible combinations resulting for each pairing of 20-word lists.[17] As a consequence, the probability that any given participant would have a single trial (out of 10 per condition) wherein the prime and target shared first sounds was less than 1 in 10. Such a low probability event would not likely account for these robust and consistent results. Nevertheless, we plotted the results that one might predict using these slim probabilities and found that the predicted outcomes did not resemble the obtained results or their inverse (to allow for the alternative predictions that sound similarity either facilitates or inhibits responses).

Finally, the most conclusive test of the sound-similarity explanation does not rely on probability estimates but derives directly from the data itself. If evaluative congruence and extremity are just proxies for sound similarity, and the primes and targets are drawn from the same pools of words, then we would predict that target extremity would moderate the automatic evaluation effect in the same manner that prime extremity does. This is not the case (see the Results section for Experiment 2). For all these reasons, we can confidently reject the sound-similarity explanation for this effect.

Surprise! Another possible explanation for the reverse priming effects is that the inconsistency between primes and targets in incongruent pairs is surprising to perceivers and thus engenders a different, perhaps faster, mode of processing.[18] Superior memory for schema-incongruent information (e.g., Hastie & Kumar, 1979) has been attributed to enhanced attention to such information. If the inconsistency of primes and targets (which is most salient when stimuli are extreme) enhances attentional focus, then processing of the target may be more efficient, and responses consequently faster. Research on the effects of surprising stimuli on speed of judgments (e.g., Meyer, Niepel, Rudolph, & Schuetzwohl, 1991; Niepel, Rudolph, Schuetzwohl, & Meyer, 1994), however, indicates that although memory for schema-discrepant information may be better, its presentation tends to slow responses. This would militate against a "surprise" explanation for the present contrast effects. Furthermore, if surprise resulting from prime–target incongruence were to engender faster processing, then once again we would expect target extremity, enhancing the salience of the incongruence, to moderate the reverse-priming effect as well. However, as noted above, this is not the case.

Because, as noted, target extremity does not moderate the automatic evaluation effect in the manner that prime extremity does, it is apparent that the action (in terms of mechanisms underlying the reverse priming effect) is in the response to the prime alone, not the combination of the prime and target. The task is to correctly

---

[15] A recent study by Dijksterhuis et al. (1998) reporting assimilation and contrast effects in "automatic behavior" begs comparison with the present research. They demonstrate effectively that their contrast effects in measures of participants' performance are due to contrasted comparisons between the self and primed exemplars, while assimilation effects occur with stereotypic (categorical) primes that are not compared with the self. Thus, although the behavioral dependent variables (e.g., walking speed, intellectual performance) are deemed "automatic," the judgmental processes are very different from those occurring in the reverse priming reported herein and do not have direct implications for automatic motivation.

[16] We thank Anthony Greenwald for persistently suggesting this as the artifactual basis of automatic reverse priming.

[17] In cases where the same list of words was compared with itself, there were only 380 possible combinations because words were never paired with themselves in the experiments.

[18] Our thanks go to Diederik Stapel for his suggestion of surprise as a possible explanation.

pronounce the target word. This necessitates processing its meaning, which includes evaluation. The presentation of an extreme prime (an extraneous stimulus) poses a threat to the goal of objectively responding to the intended target. A correction is immediately instigated in order to neutralize that threat before the target appears. This correction is excessive. As a result, the extreme primes end up activating the evaluative associations that are opposite to that of their intrinsic meaning. Further experimentation is required to test directly if this is an accurate explanation of the present results, but for now, given the evidence and the inferences we can draw from our knowledge of assimilation and contrast as well as automaticity, it is a reasonable theoretical point of departure.

The Role of Pronunciation

It is possible that reverse priming effects have not been previously observed because they are unique to the pronunciation task, which is relatively rarely employed in semantic priming studies. The task itself may afford a necessary condition for reverse priming effects to occur. Balota and Lorch (1986), testing the theory of spreading activation, found that the pronunciation task can tap very different processes from other response modes. Using a pronunciation task, they found evidence for "indirect priming." Specifically, words that were not directly related, but did have an indirect semantic connection (e.g., *stripes* and *lion,* via *tiger*) facilitated each other in the pronunciation task. Such facilitation was not evidenced in lexical decision (word or nonword) judgments. It appears that the pronunciation task allows for more flexible, multistage priming, the type that may be required for the outcome of a correction process to be detected. This may explain why reverse priming effects have not been observed in prior priming studies, but this limitation does not make them any less significant than Balota and Lorch's (1986) demonstration of indirect priming.

Lurking Reverse Priming Effects

A more paradoxical explanation for the apparent novelty of the automatic reverse priming effect is that it is not, in fact, so novel. For example, we know that Eimer and Schlaghecken (1998) have demonstrated something akin to reverse priming (albeit not on the evaluative dimension), and that they have, in fact, replicated their finding seven times to ensure its validity. It is possible that other such findings, perhaps less robust and unambiguous than ours, are lying fallow in the file drawers of psychological science. With some effort we have detected more than a few previously understated contrast effects in studies of automatic and unconscious processes. As noted above, Klauer, Rossnagel, and Musch (1997) report assimilative effects in automatic evaluation with a good versus bad judgment task at certain SOAs, but also report small, nonsignificant reverse priming effects at other SOAs. Murphy and Zajonc (1993), in their study of unconscious affective priming, report and dismiss a marginally significant, unpredicted reverse effect. Additionally, De Houwer, Hendrickx, and Baeyens (1997) report a meta-analysis of their own evaluative conditioning studies,[19] finding that those participants who were aware of at least some subliminal primes exhibited contrast effects. Others may have interpreted past reverse priming effects as meaningless null results, reflecting a deficiency in methods or stimuli, and moved on to greener research pastures.

Fortunately, the effects we observed in Experiment 1 were striking enough to spark curiosity and a series of experiments that will extend beyond those presented here. In order to understand better the dynamics of reverse priming, and if it indeed reflects automatic correction, considerably more experimentation is required. Specifically, the role of accuracy motivation, and the specific mechanisms underlying the correction process, must be investigated. At this stage in the research, however, what is more evident than the exact delimiting conditions of the effect are the implications of the phenomenon; people can be unconsciously vigilant for the biasing influence of unintended objects of evaluation and, without their conscious awareness, can engage in corrective strategies that in turn lead to contrarily biased responses. Whether or not this generalizes to many types of judgments, it is apparent from the present research that the unconscious is more complex and sophisticated than previously conceived.

[19] In these evaluative conditioning experiments, participants learn, via repeated pairings of negative and positive subliminal stimuli with previously neutral target stimuli, to hold relatively negative or positive evaluations of the target stimuli.

References

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12,* 336–345.

Banaji, M. R., & Hardin, C. (1996). Automatic gender stereotyping. *Psychological Science, 7,* 136–141.

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

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (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 Press.

Bargh, J. A. (1997). The automaticity of everyday life. In R. Wyer (Ed.), *Advances in social cognition* (Vol. X, pp. 1–62). 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 Press.

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., & Gollwitzer, P. M. (1994). Environmental control of goal-directed action: Automatic and strategic contingencies between situations and behavior. 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.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, & Computers, 18,* 299–303.

Bem, D. (1987). Writing the empirical journal article. In M. P. Zanna & J. M. Darley (Eds.), *The complete academic.* Hillsdale, NJ: Erlbaum.

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

Chaiken, S., & Bargh, J. A. (1993). Occurrence versus moderation of the automatic attitude activation effect: Reply to Fazio. *Journal of Personality and Social Psychology, 64,* 759–765.

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.

DeCarlo, L. T. (1992). Intertrial interval and sequential effects in magnitude scaling. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1080–1088.

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

Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D. A., Koomen, W., van Knippenberg, A., & Scheepers, D. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology, 75,* 862–871.

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. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (Vol. 11, pp. 74–97). London: Sage.

Fazio, R. H. (1993). Variability in the likelihood of automatic attitude activation: Data reanalysis and commentary on Bargh, Chaiken, Govender, and Pratto (1992). *Journal of Personality and Social Psychology, 64,* 753–758.

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.

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23–30.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996, September 20). 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.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37,* 25–38.

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 Press.

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

Kawakami, K., Dion, K. L., & Dovidio, J. F. (1998). Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin, 24,* 407–416.

Kihlstrom, J. F. (1987, September 18). The cognitive unconscious. *Science, 237,* 1445–1452.

Klauer, K. C., & Musch, J. (1998). *Affective priming: The puzzle of the naming task.* Unpublished manuscript, Rheinische Friedrich-Wilhelms-Universität Bonn.

Klauer, K. C., Rossnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 246–255.

Kruglanski, A. W. (1990). Motivations for judging and knowing: Implications for causal attribution. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 333–368). New York: Guilford Press.

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

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.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Meyer, W., Niepel, M., Rudolph, U., & Schuetzwohl, A. (1991). An experimental analysis of surprise. *Cognition and Emotion, 5,* 295–311.

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.

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., & 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.

Niepel, M., Rudolph, U., Schuetzwohl, A., & Meyer, W. (1994). Temporal characteristics of the surprise reaction induced by schema-discrepant visual and auditory events. *Cognition and Emotion, 8,* 433–452.

Perdue, C. W., & Gurtman, M. B. (1990). Evidence for automaticity in ageism. *Journal of Experimental Psychology, 26,* 199–216.

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

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and performance* (Vol. V, pp. 669–682). New York: Academic Press.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

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

Shoaf, L. C., & Pitt, M. A. (1998, November). *Inhibition in phonological priming: Lexical or strategic effect?* Poster session presented at the annual meeting of the Psychonomics Society, Dallas, Texas.

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. (1992). The different routes to social judgment: Experiential versus informational strategies. In L. L. Martin and A. Tesser (Eds.), *The construction of social judgments* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Strack, F., & Hannover, B. (1996). Awareness of influence as a precondition for implementing correctional goals. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 579–596). New York: Guilford Press.

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.

Tajfel, H. (1978). *Differentiation between social groups.* London: Academic Press.

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

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.

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

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.

Appendix

Stimulus Words Used in Experiments 1–6

Experiment 1		Experiments 1–3		Experiments 2–6	
Food Words	Generic Words	Black Words	White Words	Moderate Words	Extreme Words
		Negative			
BEETS	ACCIDENT	CRACK	ARISTOCRACY	AMBULANCE	ABUSE
BITTER	AGONY	GRAFFITI	ARYAN	ARMY	ANGER
CABBAGE	COFFIN	HAITI	DAHMER	CAPACITY	CANCER
HERRING	DEATH	HARLEM	EMBEZZLER	CHAIR	CRASH
KALE	DEVIL	HOMEBOY	HICK	COMPARISON	FAILURE
LENTILS	EXECUTION	HOOD	HILLBILLY	CONTEXT	GARBAGE
MAYONNAISE	FUNERAL	ISLAM	HIPPIES	FIRE	GLOOM
MEATLOAF	HELL	JACKSON	HITLER	HAMMER	GRIEF
OKRA	HORROR	JAIL	HUNTING	HIDE	MAGGOT
PIMENTO	JEALOUSY	JERRICURL	KLAN	INDUSTRY	MILDEW
PRUNES	LEPROSY	MINORITY	NAZI	INK	POISON
RADISH	LONELINESS	MULATTO	OPPRESSOR	KEROSENE	RAT
RHUBARB	MOSQUITO	NEGRO	PALE	MONTH	RIDICULE
SARDINES	PARALYSIS	PLANTATION	POLITICIAN	PUNISHMENT	SLAP
SPAM	STRESS	RAP	POLKA	RATTLE	TERMITE
SPROUTS	SUICIDE	RIOT	PURITAN	REVOLT	TRAGEDY
SQUID	TOOTHACHE	SEGREGATION	SKINHEAD	RUST	TRAITOR
TURNIPS	TORTURE	SLUM	SNOB	SHADOW	TROUBLE
VINEGAR	TUMOR	TYSON	SUNBURN	SQUARE	VENOM
WHISKEY	VIRUS	WELFARE	YUPPIE	VANITY	VOMIT
		Positive			
BRAN	BABY	ATHLETE	AEROBICS	BUTTERFLY	BLOSSOM
BROCCOLI	BIRTHDAY	BASKETBALL	ARGYLE	CLOTHING	BUNNY
CANTALOUPE	ENJOYMENT	BRAIDS	BALLET	CUSTOM	DELIGHT
CELERY	FRIDAY	BROTHER	BIRKENSTOK	GLACIER	FRIEND
CLAMS	HEAVEN	CALYPSO	CANADA	HISTORY	GLORY
COCONUT	JOY	CARIBBEAN	CARDIGAN	HORSE	HEALTH
COUSCOUS	KINDNESS	COSBY	COCKTAIL	MOMENT	HUG
CUSTARD	LAUGHTER	ELLINGTON	COTTAGE	OPINION	HUMOR
FUDGE	LIFE	EMANCIPATION	ENGLAND	PATENT	JUSTICE
JELLO	PARADISE	GILLESPIE	EUROPE	PILLOW	LOVE
MARMALADE	PASSION	GOSPEL	HOCKEY	PLANT	NATURE
MUSHROOM	PLEASURE	HENDRIX	LETTERMAN	PRAIRIE	PEACE
PANCAKES	PUPPY	JAMAICA	MAYFLOWER	RIVER	RESCUE
PEARS	RAINBOW	JAZZ	MONK	SALUTE	RESPECT
PICKLES	SUMMER	MUSIC	PRESIDENT	SAPPHIRE	SLEEP
RELISH	SUNRISE	REGGAE	SEINFELD	THEORY	SUCCESS
SOUP	SUNSET	RHYTHM	SKIING	VEHICLE	TRIUMPH
SWEET	TRUST	ROOTS	TENNIS	VILLAGE	VACATION
WALNUTS	TRUTH	SOUL	VERMONT	WINDOW	VICTORY
ZUCCHINI	WARMTH	WINFREY	WALTZ	WORLD	VIRTUE

Received October 19, 1998
Revision received April 9, 1999
Accepted April 22, 1999 ■