

Attentional Influences on Affective Priming: Does Categorization Influence Spontaneous Evaluations of Multiply Categorizable Objects?

Bertram Gawronski
University of Western Ontario

William A. Cunningham
Ohio State University

Etienne P. LeBel
University of Western Ontario

Roland Deutsch
University of Würzburg

Previous research suggests that spontaneous evaluative responses to a stimulus depend on how that stimulus is categorized. The present research indicates that such categorization effects depend on task-specific aspects of the measure, thereby concealing or overriding effects of unattended category cues. Results showed that affective priming effects in a paradigm based on response interference depended on participants' attention to the category membership of the primes. These effects were reflected in (a) reduced effect sizes, (b) reduced internal consistencies, and (c) reduced correlations to corresponding self-reports when attention was directed toward alternative categories. Such attention-related decrements were not obtained for a priming paradigm based on affect misattribution, which showed reliable priming effects irrespective of participants' attention to the relevant categories. These results challenge the ubiquity of categorization effects on spontaneous evaluations, suggesting that the impact of unattended category cues depends on conditions inherent in specific tasks.

Keywords: Affect Misattribution; Affective Priming; Categorization; Implicit Measures; Reliability; Response Interference

Over the last decade, psychologists witnessed what could be called a measurement revolution. This revolution has been stimulated by the development of a new class of indirect measurement procedures to assess mental representations and spontaneous evaluative responses (for reviews, see Fazio & Olson, 2003; Petty, Fazio, & Briñol, 2009; Wittenbrink & Schwarz, 2007).¹ In contrast to standard self-report measures, these measures are typically based on experimental paradigms, such as sequential priming (Neely, 1977) or other types of compatibility tasks (Kornblum, Hasbroucq, & Osman, 1990). The most prominent examples include the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) and variants of affective priming (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Payne, Cheng, Govorun, & Stewart, 2005). Other examples include variants of semantic priming (e.g., Banaji & Hardin, 1996; Wittenbrink, Judd, & Park, 1997), the extrinsic affective Simon task (De Houwer, 2003a), and the go/no-go association task (Nosek & Banaji, 2001).

Even though there is still controversy about the nature of the constructs that are assessed by these measures (e.g., Arkes & Tetlock, 2004; De Houwer, 2006; Gawronski, LeBel, & Peters, 2007), their usefulness has been demonstrated in numerous studies showing that they predict judgments and behavior over and above standard self-report measures (for a review, see Frieze, Hofmann, & Schmitt, 2008). Based on this evidence, researchers became increasingly interested in factors that have the potential to change the psychological attributes assessed by indirect measures (Gawronski & Sritharan, in press). The available research on this question suggests that spontaneous evaluations assessed by indirect measures may not be

as rigid and inflexible as it has been assumed by earlier models (e.g., Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). Rather, there is accumulating evidence that spontaneous evaluations often vary as a function of the context, such that the same object can elicit different responses depending on the particular context in which it is encountered (e.g., Barden, Maddux, Petty, & Brewer, 2004; Rydell & Gawronski, in press; Van Bavel & Cunningham, 2009; Wittenbrink, Judd, & Park, 2001; for a review, see Gawronski & Sritharan, in press).

To account for these findings, several models argue that spontaneous evaluative responses to a given object depend on how the object is categorized (e.g., Cunningham, Zelazo, Packer, & Van Bavel, 2007; Fazio, 2007; Gawronski & Bodenhausen, 2006). As virtually any object can be categorized in multiple ways, spontaneous responses to the same object may vary if that object is categorized differently in different contexts and the employed categories are associated with different valence.² For instance, spontaneous evaluative responses to Michael Jordan may be more favorable when he is categorized as an athlete than when he is categorized as African American (e.g., Mitchell, Nosek, & Banaji, 2003; Wheeler & Fiske, 2005). Moreover, which of the multiple applicable categories is used to categorize a given object may be determined by the momentary salience of relevant category cues, which in turn influences the type of evaluative associations that become activated.

The main goal of the present research was to investigate the relative power of categorization processes on spontaneous evaluations assessed by two distinct indirect measures. Specifically, we were interested in how attention to categories may interact

with task-specific mechanisms underlying different indirect measures, such that categorization may influence priming effects on some measures, but not on others.³ Such evidence would indicate that task-specific mechanisms can influence the effects revealed by indirect measures, and that theorizing about spontaneous evaluations may be distorted if the crucial role of these mechanisms is not taken into account (see also Gawronski, Deutsch, LeBel, & Peters, 2008; Sherman, Gawronski, Gonsalkorale, Hugenberg, Allen, & Groom, 2008). In addition, evidence of this kind would qualify the universal power of categorization processes, such that unattended category cues may influence spontaneous evaluations under various conditions that are constrained by task-specific mechanisms. That is, category-related features (e.g., race-related cues) may sometimes influence spontaneous evaluative responses even when the object is categorized in terms of an alternative category (e.g., age); however, such effects may not be detected by the employed task if categorization processes interact with task-specific mechanisms in a manner such that (a) unattended category cues do not elicit any evaluative response in that particular task, or (b) unattended category cues do elicit an evaluative response, but the task-specific mechanism does not reliably translate this evaluation into an observable response to the target stimuli (Gawronski et al., 2008; Moors, Spruyt, & De Houwer, in press).

In the present studies, we tested these assumptions by comparing the impact of categorization processes on two variants of affective priming: Fazio et al.'s (1995) evaluative priming task and Payne et al.'s (2005) affect misattribution procedure. This research was inspired by earlier evidence showing that the two measures can produce opposite effects of the same experimental manipulation (e.g., Deutsch & Gawronski, 2009; Deutsch, Kordts-Freudinger, Gawronski, & Strack, in press), suggesting that task-specific mechanisms influence the effects revealed by these measures in a significant manner. Thus, before we discuss how categorization processes may influence affective priming effects in the two measures, it is important to specify their presumed underlying mechanisms, which are response interference for Fazio et al.'s (1995) paradigm and affect misattribution for Payne et al.'s (2005) measure.

Response Interference

A useful example to illustrate the notion of response interference is the Stroop color-naming task (Stroop, 1935). In this task, participants are asked to name the color of a word presented on a screen as quickly as possible. The critical items in this task are words that themselves represent a color label. On these items, people usually show better performance when the ink color of the word corresponds to the color label

depicted by the word (e.g., the word *RED* written in red ink) than when ink color and color label do not correspond to one another (e.g., the word *RED* written in blue ink). These differences in performance can be explained by the influence of two independent response tendencies elicited by the ink color and the semantic meaning of the stimulus. For instance, the word *RED* written in blue ink may elicit two response tendencies that interfere with a quick and accurate response to that stimulus, namely a response tendency to say "red" elicited by the word meaning and a response tendency to say "blue" elicited by the ink color. Conversely, the word *RED* written in red ink may elicit two response tendencies that facilitate quick and accurate responses, namely a response tendency to say "red" elicited by the word meaning and a response tendency to say "red" elicited by the ink color. Put differently, the first case results in two response tendencies that have antagonistic effects on participants' responses, whereas the latter case results in two response tendencies that have synergistic effects. Thus, performance in the Stroop task depends (among other factors) on the relative strength of two competing response tendencies that can be compatible or incompatible with each other (De Houwer, 2003b).

Accumulating evidence suggests that processes of response interference also play a significant role in Fazio et al.'s (1995) paradigm (e.g., De Houwer, Hermans, Rothermund, & Wentura, 2002; Gawronski, Deutsch, & Seidel, 2005; Klauer, Roßnagel, & Musch, 1997; Klinger, Burton, & Pitts, 2000; Spruyt, Hermans, De Houwer, Vandromme, & Eelen, 2007; Wentura, 1999). In this task, participants have to indicate the valence of positive and negative target words as quickly as possible. Shortly before the presentation of a target word, participants are briefly presented with either a positive or a negative prime stimulus. Affective priming effects are typically reflected in faster responses to positive words after priming with positive as compared to negative stimuli, and in faster responses to negative words after priming with negative as compared to positive stimuli (for a review, see Klauer & Musch, 2003).

Priming effects in Fazio et al.'s (1995) paradigm have originally been interpreted as being due to processes of spreading activation in associative memory (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Hermans, De Houwer, Eelen, 1994; see Collins & Loftus, 1975); however, recent evidence suggests that they are better understood as being driven by Stroop-like processes of response interference (e.g., De Houwer et al., 2002; Gawronski et al., 2005; Klauer et al., 1997; Klinger et al., 2000; Spruyt et al., 2007; Wentura, 1999). According to this account, the valence of the prime stimulus triggers a prepotent response tendency that can be compatible or incompatible with

the response tendency elicited by the target word. If the prime stimulus and the target word share the same valence, the two response tendencies have synergistic effects. If, however, the prime stimulus and the target word have a different valence, the two response tendencies have antagonistic effects. Thus, affective priming effects in Fazio et al.'s (1995) task depend (among other factors) on the relative strength of two competing response tendencies, thereby implying a response interference mechanism similar to the one operating in the Stroop task (De Houwer, 2003b; Klauer & Musch, 2003).⁴

Affect Misattribution

Payne et al. (2005) recently introduced an affective priming variant that resembles Fazio et al.'s (1995) paradigm on the surface, but substantially differs in methodological details and in the task-specific mechanism underlying the measure. In this paradigm, participants are briefly presented with a positive or a negative prime stimulus, which is followed by a neutral Chinese ideograph (see also Murphy & Zajonc, 1993). Participants' task is to indicate whether they consider the Chinese ideograph as more pleasant or less pleasant than the average Chinese ideograph. Affective priming effects in this paradigm are reflected in assimilation effects, such that the neutral Chinese ideographs are evaluated more positively when they are preceded by a positive prime stimulus than when they are preceded by a negative prime stimulus.

Even though Payne et al.'s (2005) task may appear similar to Fazio et al.'s (1995) paradigm, the mechanisms underlying the two measures are quite distinct. Different from the response interference mechanism underlying Fazio et al.'s (1995) task, Payne et al.'s (2005) paradigm is assumed to be driven by a misattribution mechanism, whereby the affect elicited by the prime is (mistakenly) used to evaluate the Chinese ideograph. Specifically, it is assumed that the affective state elicited by the prime persists during the presentation of the neutral Chinese ideograph, thereby biasing participants' evaluations of the target. Thus, as Payne et al. (2005) argued, participants seem to mistakenly assume that their affective reaction stems from the target ideograph, which may result from their inability to disentangle the relative contributions of prime-related versus target-related responses to their momentary affective state. Moreover, the target stimuli in Payne et al.'s (2005) paradigm are typically selected to be evaluatively neutral in order to maximize the likelihood that both positive and negative valence can be attributed to them. As such, they do not trigger the same kind of response tendencies as the target words in Fazio et al.'s (1995) task, which could be congruent or incongruent with the response tendencies elicited by the primes. Thus, the mechanisms underlying the two affective priming variants are quite different, such that

Fazio et al.'s (1995) task is based on the interference of two independent response tendencies resulting from the prime and target stimuli, whereas Payne et al.'s (2005) paradigm is presumably based on the misattribution of prime characteristics to the neutral target stimuli. Based on this difference, we will use the acronym *RIT* for Fazio et al.'s (1995) response-interference task, and the acronym *AMP* for Payne et al.'s (2005) affect misattribution procedure.

Categorization and Spontaneous Evaluation

Because affective priming tasks typically do not require a categorization of the stimuli used as primes, several researchers argued that affective priming tasks are more likely to reflect evaluative associations pertaining to particular features of the individual exemplars, rather than associations to the categories these exemplars belong to (e.g., Livingston & Brewer, 2002; Olson & Fazio, 2003). For instance, priming effects resulting from exposure to a black face prime may be more likely to reflect evaluative associations related to the individual exemplar rather than black people in general. This feature has also been used to explain the lack of correspondence that is sometimes observed for different kinds of indirect measures. For instance, investigating differences between the RIT and the IAT in the domain of racial attitudes, Olson and Fazio (2003) found a higher correlation between the two measures when participants were instructed to pay attention to the race of black and white face primes presented in the priming task. According to Olson and Fazio (2003) this increase in correlation is due to the fact that the IAT requires an explicit categorization of the presented stimuli, which is typically not the case in affective priming tasks. Thus, the IAT may differ from affective priming tasks, such that the IAT is more likely to capture category-related associations (see also De Houwer, 2001) whereas affective priming tasks are more strongly influenced by associations related to the particular exemplars used as prime stimuli (see also Livingston & Brewer, 2002).

Olson and Fazio's (2003) findings are consistent with the assumption that spontaneous evaluative responses to a given object depend on how that object is categorized. At the same time, task-specific mechanisms have been shown to shape affective priming effects in a significant manner, such that the same experimental manipulation can produce opposite effects on otherwise identical measures (e.g., Deutsch & Gawronski, 2009; Deutsch et al., in press). Importantly, task-specific mechanisms may also be relevant for the impact of categorization processes on spontaneous evaluations assessed by affective priming tasks. Priming effects in the AMP are presumably based on diffuse affective states that are (mistakenly) used to make an evaluative judgment of the target stimulus. The RIT, in contrast, seems to assess focused

evaluations that are more closely related to the primes rather than being based on diffuse affective states. As such, the AMP may be more likely to integrate evaluative information from multiple sources, whereas the RIT may capture only those pieces of information that are in the attentional focus during the task. In line with this assumption, Deutsch and Gawronski (2009) found that two sequentially presented prime stimuli produced evaluative contrast effects in the RIT, such that priming effects of the second prime were more pronounced when this prime was preceded by a first prime of the opposite valence than when it was preceded by a first prime of the same valence (see also Fockenberg, Koole, & Semin, 2008; Gawronski et al., 2005; Klauer, Teige-Mocigemba, & Spruyt, 2009); however, the same procedural setup produced additive effects in an otherwise identical variant of the AMP, such that priming effects were more pronounced when the two primes had the same valence than when they had the opposite valence. Even though the mechanisms that are responsible for evaluative contrast effects in the RIT are still under debate (see Gawronski et al., 2008; Klauer et al., 2009), these findings suggest that the task-specific mechanism underlying the AMP is more likely to integrate evaluative information from multiple sources, in this case the independent contributions of two sequentially presented prime stimuli (see also Murphy, Monahan, & Zajonc, 1995). As such, it is likely that previously obtained categorization effects on spontaneous evaluations may differ for the RIT and the AMP, such that the AMP is more likely to reveal previously unidentified effects of category cues that are ignored or unattended. That is, even though priming effects in the RIT may be modulated by participants' attention to a given category, priming effects in the AMP may be influenced by category cues regardless of whether participants do or do not pay attention to that category.

To test these assumptions, we compared priming effects of multiply categorizable objects in the RIT and the AMP. For this purpose, participants were primed with either black or white faces of either young or old age. Adopting Olson and Fazio's (2003) categorization manipulation, participants were asked to keep a mental tally of either (a) how many black versus white individuals were presented over the course of the task (race categorization) or (b) how many young versus old individuals were presented during the task (age categorization). Based on our theoretical considerations about task-specific influences of categorization, we expected reliable priming effects of a given category dimension (e.g., race) in the RIT only when participants paid attention to that category in the task, but not when their attention was directed toward an alternative category dimension (i.e., age). These effects were expected to be reflected in (a) reduced effect sizes

of priming effects, (b) reduced internal consistencies of corresponding preference scores, and (c) reduced correlations to criterion measures of self-reported category preferences. Such attention-related decrements should not occur for an otherwise identical variant of the AMP, which was expected to show reliable priming effects of a given category dimension irrespective of whether participants did or did not pay attention to that dimension.

Pilot Studies

Two pilot studies provided preliminary support for these assumptions. The first one ($N = 82$) investigated categorization effects on Fazio et al.'s (1995) RIT; the second one ($N = 81$) tested categorization effects on Payne et al.'s (2005) AMP. In both studies, the face primes were presented for 200 ms, after which they were replaced with the respective target stimuli (i.e., positive or negative word in the RIT; neutral Chinese ideographs in the AMP). In line with the standard procedures of the RIT and the AMP, the target stimuli in the RIT remained on the screen until participants pressed one of the two response keys (see Fazio et al., 1995); the target stimuli in the AMP were replaced with a black-and-white pattern mask after 100 ms, which remained on the screen until participants gave their response (see Payne et al., 2005). Participants in the RIT study were asked to categorize the target words as positive or negative; participants in the AMP study were asked to judge the Chinese ideographs as visually more pleasant or less pleasant than average (AMP). Analyses were conducted on priming scores reflecting preferences for whites over blacks (*implicit racism*) and preferences for young over old people (*implicit ageism*).⁵

For the RIT, a 2 (preference score: implicit racism vs. implicit ageism) \times 2 (attention to prime category: race vs. age) mixed-model ANOVA revealed a significant two-way interaction, $F(1, 80) = 6.91$, $p = .01$, $\eta_p^2 = .079$, showing that implicit racism scores were significantly higher when participants paid attention to race ($M = 7.78$, $SD = 35.71$) than when they paid attention to age ($M = -6.38$, $SD = 20.03$), $F(1, 80) = 4.97$, $p = .03$, $\eta_p^2 = .059$. Conversely, implicit ageism scores showed a tendency to be higher when participants paid attention to age ($M = 11.00$, $SD = 33.56$) than when they paid attention to race ($M = 0.07$, $SD = 22.13$), $F(1, 80) = 3.00$, $p = .09$, $\eta_p^2 = .036$. Moreover, reliability analyses revealed that implicit racism scores showed higher split-half correlations when participants paid attention to race ($r = .26$, $p = .11$) than when they paid attention to age ($r = -.38$, $p = .01$). Conversely, implicit ageism scores showed higher split-half correlations when participants paid attention to age ($r = .24$, $p = .12$) than when they paid attention to race ($r = -.15$, $p = .36$). The difference between split-half correlations as a function of attention was

significant for implicit racism scores, $z = 2.90$, $p = .004$, and marginally significant for implicit ageism scores, $z = 1.73$, $p = .08$.

Results for the AMP were remarkably different, such that implicit racism scores did not differ as a function of whether participants paid attention to race ($M = 0.047$, $SD = 0.21$) or age ($M = 0.047$, $SD = 0.14$). Interestingly, implicit ageism scores tended to be higher when participants paid attention to race ($M = 0.056$, $SD = 0.24$) than when they paid attention to age ($M = 0.039$, $SD = 0.22$); however, the two-way interaction of attention and preference score was far from statistical significance, $F(1, 74) = 0.06$, $p = .81$, $\eta_p^2 = .001$. Both implicit racism and implicit ageism scores were higher than zero, $t(75) = 2.36$, $p = .02$ for implicit racism, and $t(75) = 1.83$, $p = .07$ for implicit ageism. Reliability analyses further revealed that implicit racism scores showed high split-half correlations irrespective of whether participants paid attention to race ($r = .76$, $p < .001$) or age ($r = .62$, $p < .001$). The same was true for implicit ageism scores, which showed high split-half correlations irrespective of whether participants paid attention to age ($r = .81$, $p < .001$) or race ($r = .82$, $p < .001$). Split-half correlations did not significantly differ as a function of attention for both implicit racism scores, $z = 1.13$, $p = .26$, and implicit ageism scores, $z = 0.12$, $p = .90$.⁶

Main Study

The primary objective of the main study was to directly compare categorization effects on the RIT and the AMP in a between-subjects design. In addition, we aimed at providing further evidence for differences in the criterion validity of the respective preference scores by including a measure of self-reported category evaluations (see Gawronski, LeBel, Peters, & Banse, 2009). Expanding on the findings of our pilot studies, we expected reliable priming effects in the RIT only for the category dimension that participants pay attention to, but not for the category dimension that is unattended. In contrast, priming effects in the AMP were expected to be unaffected by our attention manipulation. These effects were expected for both the overall size of priming effects as well as internal consistencies of the corresponding preference scores. Moreover, RIT scores were predicted to show significant correlations to corresponding self-reported category-evaluations only when participants pay attention to these categories in the priming task; however, such attention-related modulations should not occur for the AMP, which was expected to show significant correlations to corresponding self-reported category-evaluations regardless of whether participants do or do not pay attention to these categories in the priming task.

Method

Participants and design. A total of 194 undergraduates at the University of Western Ontario (138 female, 56 male) participated in the Main Study. Half of the participants completed the RIT (Fazio et al., 1995); the remaining half completed an otherwise identical variant of the AMP (Payne et al., 2005). All subjects received course credit for their participation.

Materials. As prime stimuli, we used a total of 40 head-and-shoulder color photographs depicting young white, old white, young black, or old black males, with 10 pictures for each of the four prime categories. As target words for the RIT, we used 10 positive nouns (*paradise, summer, harmony, freedom, honesty, pleasure, sunrise, love, peace, vacation*) and 10 negative nouns (*cockroach, poison, vomit, bomb, virus, disaster, terror, rotten, accident, pollution*). In choosing the target words, we made every effort to avoid words that are stereotypically related to the social stereotypes regarding race and age. As target stimuli for the AMP, we used a set of 160 Chinese ideographs from Payne et al. (2005).

Procedure. Each of the two priming tasks consisted of 160 trials, including 40 trials for each of the 4 possible prime categories implied by our manipulation of race and age. Order of trials was randomized for each participant. Each trial began with the presentation of a fixation cross for 500 ms, which was replaced by a face prime for 200 ms. The face prime was immediately followed by either a positive or negative target word (RIT) or by a Chinese ideograph (AMP). To maximize the comparability between the two tasks, the target stimulus was replaced by a black-and-white pattern mask after 100 ms in both the RIT and the AMP. Participants were asked to indicate as quickly as possible whether they consider the target stimulus as positive or negative (RIT) or as more pleasant or less pleasant than the average Chinese ideograph (AMP). Participants were asked to press a right-hand key (5 of the number pad) for positive / more pleasant responses, and a left-hand key (A) for negative / less pleasant responses. The employed software (DirectRT) recorded response latencies from the onset of the last stimulus of a given trial (i.e., the black-and-white pattern mask).⁷ The inter-trial interval was 1000 ms. If participants did not respond within 1500 ms after the onset of the target stimulus, the target word was replaced with the message *Please try to respond faster!* for 2000 ms before the next trial started. Incorrect responses in the RIT were followed by the message *Error!* for 1500 ms. The instructions in the RIT and AMP conditions were kept as similar as possible. Specifically, participants in the RIT condition were told:

The next component of this study is concerned with how well people can do two things at the same time (e.g., driving a car and having a conversation with a

passenger). For this purpose, you will be presented with positive and negative words. In addition, you will be presented with faces that briefly appear before the words are presented. Your task is to indicate as quickly as possible whether the word on the screen is a positive or a negative word. Please press the "A" key on the left side of the keyboard when you see a negative word, and please press the "5" key on the right side of the keyboard when you see a positive word. In order to facilitate faster responses to the words, please keep your main left-hand finger on the "A" key and your main right-hand finger on the "5" key. In addition to this task, we want you to keep a mental tally of how many of the presented pictures show an old (black) person and how many show a young (white) person. After you have finished the task, you will be asked to estimate the number of pictures that have shown an old (black) person and the number of pictures that have shown a young (white) person. Please note that the words and the faces will be presented only for a very brief time. So, please pay close attention to the words and try to keep a mental tally of how many of the presented pictures show an old (black) person and how many show a young (white) person. **IMPORTANT!** Note that the faces can sometimes bias people's judgments of the words. Because we are interested in how people can avoid being biased, please try your absolute best not to let the faces bias your judgments of the words. Again, please press the "A" key on the left side of the keyboard when you see a negative word, and please press the "5" key on the right side of the keyboard when you see a positive word. In addition, please try to keep a mental tally of how many young (white) and old (black) faces are presented during the task.

Participants in the AMP condition were given identical instructions, the only difference being that (a) instructions referred to Chinese ideographs rather than positive and negative words as target stimuli, and (b) participants were asked to indicate whether they consider the Chinese ideographs as more pleasant or less pleasant than average. After the priming task, all participants were asked to rate the warmth or coldness of their personal feelings associated with (a) young people, (b) elderly people, (c) black people, and (d) white people, on four 7-point scales ranging from 1 (very cold) to 7 (very warm).

Results

Data aggregation. The data aggregation procedures were identical to the ones employed in the Pilot Studies. Specifically, we discarded all latencies from RIT trials on which participants did not respond within the response deadline of 1500 ms (3.0%) or showed an incorrect response (10.9%). Following recommendations by Ratcliff (1993), we also controlled for outliers by excluding all trials with

response latencies greater than 1250 ms (1.8%). Latencies were then averaged for the 8 prime-target combinations. The same procedures were applied to the AMP, such that we first discarded all responses from trials on which participants did not respond within the response deadline of 1500 ms (2.7%) or showed a response latency greater than 1250 ms (1.6%) and then calculated the mean proportion of pleasant responses for each of the four prime categories (i.e., young-White, old-White, young-Black, old-Black). Means and standard deviations for the RIT are printed in Table 1; means and standard deviations for the AMP are printed in Table 2.⁸

To test our main hypotheses, we calculated affective priming scores of implicit preference for whites over blacks and implicit preference for young over old, respectively. RIT scores of implicit preference for whites over blacks were calculated by subtracting the mean response latency to positive words preceded by a white face prime from the mean response latency to positive words preceded by a black face prime (i.e., higher positivity for white compared to black), and by subtracting the mean response latency to negative words preceded by a black prime from the mean response latency to negative words preceded by a white prime (i.e., higher negativity for black compared to white). The resulting scores were then averaged as an index of implicit preference for whites over blacks (for the sake of simplicity referred to as implicit racism). RIT scores of implicit preference for young over old were calculated accordingly by subtracting the mean response latency to positive words preceded by a young face prime from the mean response latency to positive words preceded by an old face prime (i.e., higher positivity for young compared to old), and by subtracting the mean response latency to negative words preceded by an old prime from the mean response latency to negative words preceded by a young prime (i.e., higher negativity for old compared to young). The two difference scores were again averaged as an index of implicit preference for young over old (for the sake of simplicity referred to as implicit ageism). AMP scores of implicit preference for whites over blacks (implicit racism) were calculated by subtracting the mean positivity scores for black face primes from the mean positivity scores for white face primes. Correspondingly, we calculated scores of implicit preference for young over old (implicit ageism) by subtracting the mean positivity scores for old face primes from the mean positivity scores for young face primes.

Size of priming effects. Submitted to a 2 (preference score: implicit racism vs. implicit ageism) \times 2 (attention to prime feature: race vs. age) mixed-model ANOVA with the first variable as a within-subjects factor, RIT preference scores revealed a significant

two-way interaction, $F(1, 95) = 4.13, p = .04, \eta_p^2 = .042$ (see Table 3). Replicating the pattern obtained in the Pilot Studies, implicit racism scores tended to be higher when participants paid attention to race ($M = 9.23, SD = 32.64$) than when they paid attention to age ($M = 1.25, SD = 31.67$), $F(1, 95) = 1.47, p = .23, \eta_p^2 = .015$, whereas implicit ageism scores showed a tendency to be higher when participants paid attention to age ($M = 9.24, SD = 38.06$) than when they paid attention to race ($M = -0.58, SD = 24.23$), $F(1, 95) = 2.37, p = .13, \eta_p^2 = .024$.

The same ANOVA on AMP preference scores did not reveal any significant effects of our attention manipulation (see Table 3). If anything, implicit racism scores tended to be higher when participants paid attention to age ($M = 0.039, SD = 0.14$) than when they paid attention to race ($M = 0.022, SD = 0.16$), whereas implicit ageism scores tended to be higher when participants paid attention to race ($M = 0.050, SD = 0.21$) than when they paid attention to age ($M = 0.035, SD = 0.21$); however, the two-way interaction of attention and preference score was far from statistical significance, $F(1, 95) = 0.36, p = .55, \eta_p^2 = .004$. Both implicit racism and implicit ageism were significantly higher than zero, $t(96) = 2.01, p = .05$ for implicit racism, and $t(96) = 2.01, p = .05$ for implicit ageism.

To test whether attention effects differ for the RIT and the AMP, scores of implicit racism and implicit ageism were standardized separately for the two measures. The resulting scores were then submitted to a 2 (preference score: implicit racism vs. implicit ageism) \times 2 (attention to prime feature: race vs. age) \times 2 (task: RIT vs. AMP) mixed-model ANOVA with the first variable as within-subjects factor and the latter two as between-subjects factors. Consistent with the assumption that attention effects differ for the two tasks, this ANOVA revealed a marginally significant three-way interaction of task, preference score, and attention, $F(1, 190) = 3.41, p = .07, \eta_p^2 = .018$.

Reliability of preference scores. To estimate the reliability of the two preference scores for the RIT and the AMP, we used an odd-even split in which we divided the trials of the two priming tasks in two blocks depending on whether these trials were associated with an odd or an even trial number. We then calculated two affective priming scores for each category dimension, one for odd and one for even trials, according to the procedures outlined above (see Table 3).

Reliability analyses for the RIT revealed that implicit racism scores showed higher split-half correlations when participants paid attention to race ($r = .25, p = .08$) than when they paid attention to age ($r = -.13, p = .40$). Conversely, implicit ageism scores showed higher split-half correlations when participants paid attention to age ($r = .27, p = .08$) than when they

paid attention to race ($r = -.29, p = .03$). The difference between split-half correlations as a function of attention was statistically significant for implicit ageism scores, $z = 2.73, p = .006$, and marginally significant for implicit racism scores, $z = 1.83, p = .07$.

Reliability analyses for the AMP revealed that implicit racism scores showed equally high split-half correlations irrespective of whether participants paid attention to race ($r = .58, p < .001$) or age ($r = .57, p < .001$). The same was true for implicit ageism scores, which showed high split-half correlations irrespective of whether participants paid attention to age ($r = .79, p < .001$) or race ($r = .82, p < .001$). Split-half correlations did not significantly differ as a function of attention for both implicit racism scores, $z = 0.07, p = .94$, and implicit ageism scores, $z = 0.41, p = .68$.⁹

Correlations to self-report measures. To provide evidence for corresponding differences in the criterion validity of the respective priming scores, we analyzed the correlations between implicit preference scores and corresponding self-reports for the two measures. For this purpose, we subtracted participants' ratings of black people from their ratings of white people (*explicit racism*) and ratings of elderly people from ratings of young people (*explicit ageism*). The results showed that correlations to self-reported preferences were generally lower for the RIT when participants did not pay attention to the relevant category dimension in the priming task (see Table 3). Specifically, explicit and implicit racism scores tended to show higher correlations when participants paid attention to race ($r = .26, p = .06$) than when they paid attention to age ($r = -.05, p = .72$), $z = 1.50, p = .13$, whereas explicit and implicit ageism scores showed higher correlations when participants paid attention to age ($r = .39, p = .01$) than when they paid attention to race ($r = -.10, p = .47$), $z = 2.43, p = .01$. In contrast, correlations between self-reported preferences and AMP scores were moderately high regardless of attention instructions (see Table 3). That is, explicit and implicit racism scores were significantly correlated regardless of whether participants paid attention to race ($r = .30, p = .03$) or age ($r = .29, p = .04$), $z = 0.05, p = .96$. The same was true for explicit and implicit ageism scores, which were significantly correlated regardless of whether participants paid attention to age ($r = .31, p = .03$) or race ($r = .29, p = .04$), $z = 0.10, p = .90$.

Discussion

Results from the Main Study further corroborate our assumption that the mechanisms underlying the RIT and the AMP are differentially sensitive to categorization processes, in that categorization processes influence affective priming effects in the RIT, but not the AMP. Replicating the results of the Pilot Studies, the RIT showed significant decrements in the overall size and reliability of priming effects when

participants did not pay attention to the relevant category dimension. Such attention-related decrements were not obtained for the AMP, which showed equally high effect sizes and reliability estimates irrespective of whether participants did or did not pay attention to the relevant category dimension. Moreover, RIT scores showed significant correlations to corresponding self-reported category-evaluations only when participants paid attention to these categories in the priming task. In contrast, AMP scores showed significant correlations to corresponding self-reports regardless of whether participants did or did not pay attention to the relevant categories in the priming task.

Even though the present findings are generally consistent with our claim that the mechanisms underlying the RIT and the AMP are differentially sensitive to categorization processes, one could object that the identification of a response to the target stimuli may be more difficult for the evaluatively ambiguous targets in the AMP as compared to the evaluatively unambiguous targets in the RIT. As such, participants' performance in categorizing the face primes may have been lower in the AMP than the RIT, which would imply a rather trivial explanation for the obtained differences. Even though the processing requirements for the target stimuli may indeed be partially responsible for the obtained differences (see below), we consider this explanation as implausible for two reasons. First, follow-up analyses of participants' estimates of the number of black/white or young/old faces revealed that the average deviation from the actual numbers of faces presented throughout the task (i.e., 80 faces for each category) did not significantly differ between the RIT ($M = 30.76$, $SD = 23.38$) and the AMP ($M = 32.16$, $SD = 22.16$), $F(1, 188) = 0.51$, $p = .61$, $\eta_p^2 = .001$.¹⁰ Second, if attention to categories was indeed impaired in the AMP, one may expect the same outcome that was obtained for the RIT when attention was directed away from a given category dimension, such that the size of priming effects, reliability estimates, and external correlations should be close to zero in all conditions. This conclusion is based on Olson and Fazio's (2003) findings, who obtained effect sizes and reliability estimates of zero when participants did not pay attention to the category membership of the primes in the RIT. This was clearly not the case for the AMP, which showed significant priming scores higher than zero, reasonably high reliability estimates, and significant correlations to corresponding self-report measures regardless of attention conditions. Thus, impaired performance in categorizing the face primes seems rather unlikely to account for the present findings.

General Discussion

The main goal of the present research was to investigate the relative impact of categorization

processes on spontaneous evaluations of multiply categorizable objects, as assessed by different kinds of affective priming tasks. Drawing on earlier evidence showing that task-specific mechanisms can influence the effects revealed by indirect measures (e.g., Deutsch & Gawronski, 2009; Deutsch et al., in press; Gawronski & Bodenhausen, 2005), we were interested in how attention to the category membership of prime stimuli may interact with task-specific mechanisms, such that categorization may modulate affective priming effects on some tasks, but not on others. In line with this contention, the present studies showed that categorization processes influenced spontaneous evaluations in Fazio et al.'s (1995) RIT, but not in Payne et al.'s (2005) AMP. Specifically, we found reliable priming effects of a given category dimension (e.g., race) in the RIT only when participants paid attention to that category, but not when their attention was directed toward an alternative category dimension (i.e., age). These effects were reflected in (a) reduced effect sizes, (b) reduced internal consistencies, and (c) reduced correlations to corresponding self-report measures when attention was directed toward alternative categories. Such attention-related decrements were not observed in otherwise identical variants of the AMP, which produced reliable priming effects of a given category dimension irrespective of whether participants did or did not pay attention to that dimension. These results not only corroborate earlier conclusions that task-specific mechanisms can shape priming effects in a significant manner (e.g., Deutsch & Gawronski, 2009; Deutsch et al., in press; Gawronski & Bodenhausen, 2005); they also qualify the universal power of categorization processes in modulating spontaneous evaluations. In fact, the present results indicate that category cues can influence spontaneous evaluations even when these cues are unattended or ignored (see also Blair, Judd, & Fallman, 2004; Ito & Urland, 2003); however, these effects may not be reflected in the relevant measurement scores if categorization processes interact with task-specific mechanisms in a manner such that the effect of unattended categories is concealed or overridden by the effects of salient categories. These results are important as task-specific effects may have the potential to distort theorizing about attitudes and evaluation, if the crucial role of task-specific mechanisms in shaping affective priming effects is not taken into account.

Differences in Evaluation or Translation?

In our introduction, we argued that the mechanism underlying the AMP is more likely to integrate evaluative information from multiple sources (e.g., Deutsch & Gawronski, 2009; Murphy et al., 1995), which in turn might make the task more sensitive to potential effects of unattended category cues; however, a major question is how the obtained differences

between the RIT and the AMP should be understood in a more specific sense. Moors et al. (in press) recently argued that affective priming tasks involve two sequential processes, both of which are essential for the emergence of affective priming effects (see also Gawronski et al., 2008). The first step involves the elicitation of an evaluative response to the prime stimulus (*evaluation*); the second step involves the influence of this evaluation on participants' responses to the target stimulus (*translation*). From this perspective, the absence of an affective priming effect can have at least two possible reasons. First, there should be no priming effect if the first process (evaluation) failed to occur, as there would be no input for the second process (translation). Second, there should be no priming effect if the first process (evaluation) occurred, but the second process (translation) failed to translate the output of the first process into an observable response to the target stimulus. These general considerations imply at least two possible explanations for why the RIT failed to show priming effects of unattended categories.

First, it seems possible that the failure to obtain priming effects of unattended category cues in the RIT reflects the complete absence of any evaluative response to these category cues (Step 1). According to this interpretation, task-specific characteristics may shape the type of evaluative responses that are activated in the first process, such that unattended category cues elicit evaluative responses in the AMP, but not the RIT. Second, it is possible that unattended category cues elicit evaluative responses in both tasks; however, these evaluations may be properly translated into observable responses to the targets only in the AMP, but not in the RIT (Step 2). According to this interpretation, task-specific characteristics of the two measures do not constrain the type of evaluative response that is activated in the first process, but the ability of the translation process in picking up these responses (Gawronski et al., 2008).

As outlined by Moors et al. (in press), it is notoriously difficult, if not impossible, to distinguish between these two possibilities (but see Ito, in press, for a discussion of neuropsychological evidence supporting an explanation in terms of failed translation). Nevertheless, the mere fact that the AMP shows reliable priming effects of unattended category cues suggests that categorization may not be as powerful in modulating spontaneous evaluations as it is often assumed (e.g., Mitchell et al., 2003; Wheeler & Fiske, 2005). After all, a priming effect can emerge only if (a) an evaluation is activated in the first step and (b) this evaluation is properly translated in the second step (Moors et al., in press). Moreover, systematic investigations may identify which of the procedural differences between the RIT and the AMP are

responsible for the obtained differences in priming effects. For instance, one could argue that the target stimuli in the RIT usually have a strong positive or negative connotation, whereas the target stimuli in the AMP are typically selected to be as neutral as possible. Thus, in combination with the frequently employed error feedback in the RIT, it is possible that participants tend to adopt an analytic processing goal in the RIT, but a holistic processing goal in the AMP. To the degree that analytic processing goals may constrain the type of evaluative responses that are elicited in response to a given object in the first process (Cunningham & Zelazo, 2007) or, alternatively, the translation of an evaluative response into an observable response to a target stimulus in the second process (Gawronski et al., 2008), analytic processing goals may play an important role for the obtained decrements in priming effects in the RIT; however, as the list of procedural differences between the RIT and the AMP is rather long (see Deutsch & Gawronski, 2009), any attributions to specific features remains speculative at this point. Hence, the most useful strategy to address this question might be a bottom-up approach that systematically varies particular features of the two measures and then compares their responsiveness to different kinds of manipulations that have been shown to produce different outcomes.

In this context, it seems important to address one additional interpretation, namely that evaluative responses assessed by the AMP might be more "explicit" compared to the RIT, and that the higher correlations to self-report measures may in fact indicate a stronger influence of controlled processes. From this perspective, the present results may be driven by a stronger influence of controlled processes on the AMP, rather than by a lacking influence of categorization processes. Even though this interpretation may seem plausible at first glance, it is important to note that the AMP produced reliable effects of a given category dimension even when participants did *not* pay attention to that category. As such, one could refute the above interpretation by making exactly the opposite claim, namely that the AMP is in fact more likely to reflect automatic processes, given that priming effects in the AMP—in contrast to the RIT—are unaffected by momentary processing goals (see Moors & De Houwer, 2006). Nevertheless, future research that systematically investigates other features of automaticity (see De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009) may help to further clarify whether controlled processes may contribute to the AMP's robustness against categorization processes.

Conclusions

In summary, the present results provide important insights into factors that influence spontaneous evaluations assessed by different kinds of affective

priming tasks, namely Fazio et al.'s (1995) RIT and Payne et al.'s (2005) AMP. Whereas priming effects in the RIT depended on participants' attention to the category membership of the primes, implicit preference scores derived from the AMP were generally unaffected by categorization processes. Given the widespread use of these tasks in various areas of psychology, we believe that such insights are important to improve our understanding of the employed measures and to avoid potential distortions in the interpretation of data obtained with these measures. In addition, the present findings challenge the ubiquity of categorization effects on spontaneous evaluations, suggesting that the impact of unattended category cues depends on conditions inherent in specific tasks. Based on these conclusions, researchers would be well-advised to replicate their findings with multiple measures that differ with regard to their task-specific mechanisms. For instance, if a given manipulation shows identical effects on the RIT and the AMP (e.g., Rydell & Gawronski, in press), the obtained correspondence would provide preliminary evidence for the generality of these effects. If, however, a given manipulation shows different effects on the RIT and the AMP (e.g., Deutsch & Gawronski, 2009), the results point to a significant role of task-specific mechanisms, which have to be taken into account in explaining the obtained effects.

References

- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?" *Psychological Inquiry*, 15, 257-278.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7, 136-141.
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, 87, 5-22.
- Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, 87, 763-778.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, 11, 97-104.
- Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The Iterative reprocessing model: A multi-level framework for attitudes and evaluation. *Social Cognition*, 25, 736-760.
- De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, 37, 443-451.
- De Houwer, J. (2003a). The extrinsic affective Simon task. *Experimental Psychology*, 50, 77-85.
- De Houwer, J. (2003b). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219-244). Mahwah, NJ: Erlbaum.
- De Houwer, J. (2003c). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory and Cognition*, 31, 360-368.
- De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA: Sage Publishers.
- De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *Cognition and Emotion*, 16, 643-666.
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135, 347-368.
- Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, 45, 101-114.
- Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (in press). Fast and fragile: A new look at the automaticity of negation processing. *Experimental Psychology*.
- Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25, 603-637.
- 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., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297-327.
- 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.
- Fockenberg, D. A., Koole, S. L., & Semin, G. R. (2008). Priming in concert: Assimilation and contrast with multiple affective and gender primes. *Social Cognition*, 26, 647-669.
- Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behaviour?

- Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, 19, 285-338.
- Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation versus retrieval experiences. *Journal of Personality and Social Psychology*, 89, 672-685.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692-731.
- Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, 24, 218-225.
- Gawronski, B., Deutsch, R., & Seidel, O. (2005). Contextual influences on implicit evaluation: Additive versus contrastive effects of evaluative context stimuli in affective priming. *Personality and Social Psychology Bulletin*, 31, 1226-1236.
- Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition*, 15, 485-499.
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193.
- Gawronski, B., LeBel, E. P., Peters, K. R., & Banse, R. (2009). Methodological issues in the validation of implicit measures. Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin*, 135, 369-372.
- Gawronski, B., & Sritharan, R. (in press). Formation, change, and contextualization: Determinants and principles of variations in implicit measures. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4-27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- 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.
- Ito, T. A. (in press). Implicit social cognition: Insights from social neuroscience. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85, 616-626.
- Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7-49). Mahwah, NJ: Erlbaum.
- Klauer, K. C., Musch, J., & Eder, A. (2005). Priming of semantic classifications: Late and response-related, or earlier and more central? *Psychonomic Bulletin and Review*, 12, 897-903.
- 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.
- Klauer, K. C., Teige-Mocigemba, S., & Spruyt, A. (2009). Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology*, 96, 265-287.
- Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 441-455.
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—a model and taxonomy. *Psychological Review*, 97, 253-270.
- Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, 82, 5-18.
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, 109, 376-400.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132, 455-469.
- Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin*, 132, 297-326.
- Moors, A., Spruyt, A., & De Houwer, J. (in press). In search of a measure that qualifies as implicit: Recommendations based on a decompositional view of automaticity. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition:*

- Measurement, theory, and applications*. New York, NY: Guilford Press.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposure. *Journal of Personality and Social Psychology*, 64, 723-739.
- Murphy, S. T., Monahan, J. L., & Zajonc, R. B. (1995). Additivity of nonconscious affect: Combined effects of priming and exposure. *Journal of Personality and Social Psychology*, 69, 589-602.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106, 226-254.
- Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, 19, 625-666.
- Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, 14, 636-639.
- Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277-293.
- Petty, R. E., Fazio, R. H., & Briñol, P. (Eds.) (2009). *Attitudes: Insights from the new implicit measures*. New York: Psychology Press.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.
- Rydell, R. J., & Gawronski, B. (in press). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion*.
- Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, 115, 314-335.
- Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, 71, 888-898.
- Spruyt, A., Hermans, D., De Houwer, J., Vandromme, H., & Eelen, P. (2007). On the nature of the affective priming effect: Effects of stimulus onset asynchrony and relatedness proportion in naming and evaluative categorization. *Memory & Cognition*, 35, 95-106.
- Stroop, J. R. (1935). Studies on the interference in serial verbal reactions. *Journal of Experimental Psychology*, 59, 239-245.
- Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, 35, 321-335.
- Wentura, D. (1999). Activation and inhibition of affective information: Evidence for negative priming in the evaluation task. *Cognition and Emotion*, 13, 65-91.
- Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, 16, 56-63.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.
- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262-274.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81, 815-827.
- Wittenbrink, B., & Schwarz, N. (Eds.) (2007). *Implicit measures of attitudes*. New York: Guilford Press.

Footnotes

¹ Following De Houwer (2006), we use the terms *direct* and *indirect* to refer to features of measurement procedures, and the terms *implicit* and *explicit* to refer to features of the constructs assessed by a given measurement procedure. Moreover, given ambiguous evidence regarding the nature of the constructs assessed by indirect measurement procedures (De Houwer, 2006; Gawronski, Hofmann, & Wilbur, 2006), we limit our interpretation of the term *implicit* to “unintentional” (i.e., evaluative responses to an object that do not require an intention to evaluate that object).

² In this context, it is important to distinguish between categorization effects on evaluation (e.g., Olson & Fazio, 2003) and evaluation effects on categorization (e.g., Smith, Fazio, & Cejka, 1996). The present study is primarily concerned with categorization effects on evaluation.

³ Even though we consider attention and categorization as *conceptually* distinct processes, we treat attention to categories and categorization as *functionally* equivalent in the present research, given that attention to a category has been argued to imply categorization and categorization implying attention to that category (e.g., Logan, 2002).

⁴ Note that both the Stroop task and Fazio et al.’s (1995) affective priming task involve variations in stimulus-response compatibility as well as stimulus-stimulus compatibility (De Houwer, 2003b); however, the actual contribution of stimulus-stimulus compatibility seems relatively minor compared to the contribution of stimulus-response compatibility in the Stroop task (De Houwer, 2003c) as well as Fazio et

al.'s (1995) affective priming task (Klauer, Musch, & Eder, 2005), which makes response interference the primary source of Stroop and priming effects in these tasks.

⁵ More information about procedural details, data aggregation, and the calculation of preference scores is provided in the Main Study.

⁶ To provide further evidence for the AMP's robustness against attentional influences, we ran an additional study that used a 75 ms prime presentation followed by blank screen for 125 ms before the target stimulus appeared, as recommended by Payne et al. (2005). The results of this study were identical, in that the AMP produced reliable priming effects of a given category dimension regardless of whether participants paid attention to race or age.

⁷ Note that the RIT Pilot Study did not include a masking stimulus. Therefore, latencies in the Pilot Study were recorded from the onset of the target stimulus. For the Main Study, corresponding latencies from the onset of the target stimulus can be calculated by adding a constant of 100 ms to all latencies (equivalent to the 100 ms target presentation). Results of the resulting priming scores are identical irrespective of whether latencies are scored from the onset of the target or the onset of the masking stimulus. In the following sections, we report the actual data recorded by the software.

⁸ Note that the standard procedure for analyzing AMP data does not involve an exclusion of trials based on RTs. In the present study, trials with latencies higher than 1250 ms were excluded to avoid a potential confounding resulting from different data treatments. To confirm that the obtained AMP data are

independent of the employed exclusion criterion, we also ran our analyses without the employed cutoff of 1250 ms. Results were identical to the ones reported below, with effect sizes and split-half correlations differing only at the level of the second decimal.

⁹ Reliability analyses produced the same pattern of results when the trials were divided on the basis of whether they occurred in the first versus second half of the task instead of odd versus even trial numbers.

¹⁰ Two participants did not provide an estimate of the number of faces.

Authors' Notes

Bertram Gawronski, Department of Psychology, University of Western Ontario, Canada; William A. Cunningham, Department of Psychology, Ohio State University, USA; Etienne P. LeBel, Department of Psychology, University of Western Ontario, Canada; Roland Deutsch, Department of Psychology, University of Würzburg, Germany.

This research has been supported by grants from the Canada Research Chairs program (CRC), the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Academic Development Fund of the University of Western Ontario (ADF) to the first author. We thank Keith Payne for providing the stimulus materials for the Affect Misattribution Procedure; and Xenia Avvakumova, Allison Boyd, William Dunlop, and Arjun Sharma for their help in collecting the data.

Correspondence concerning this article should be addressed to Bertram Gawronski, Department of Psychology, University of Western Ontario, Social Science Centre, London, Ontario, N6A 5C2, Canada, Email: bgawrons@uwo.ca.

Table 1

Means and standard deviations of latencies and error rates in responses to target words in an affective priming measure based on response interference (Fazio et al., 1995) as a function of target valence (positive vs. negative), race of face prime (white vs. black), age of face prime (young vs. old), and attention to prime feature (race vs. age).

	Positive Target		Negative Target	
	Young Prime	Old Prime	Young Prime	Old Prime
Response Latencies				
Attention to Race				
White Prime	561 ms (103)	561 ms (99)	577 ms (89)	581 ms (101)
Black Prime	576 ms (92)	577 ms (97)	578 ms (93)	574 ms (93)
Attention to Age				
White Prime	553 ms (98)	557 ms (127)	564 ms (124)	542 ms (123)
Black Prime	544 ms (130)	571 ms (132)	551 ms (106)	558 ms (118)
Error Rates				
Attention to Race				
White Prime	8.1% (27.2)	7.9% (27.0)	11.3% (31.6)	10.4% (30.5)
Black Prime	10.5% (30.7)	11.8% (32.3)	8.6% (28.0)	10.8% (31.0)
Attention to Age				
White Prime	11.0% (31.3)	13.0% (33.7)	12.4% (33.0)	12.0% (32.5)
Black Prime	11.9% (32.4)	14.6% (35.3)	10.4% (30.5)	10.5% (30.7)

Note: Standard deviations are printed in parentheses.

Table 2

Means and standard deviations of proportions of “more pleasant” responses to neutral chinese ideographs in an affective priming measure based on affect misattribution (Payne et al., 2005) as a function of race of face prime (white vs. black), age of face prime (young vs. old), and attention to prime feature (race vs. age).

	Attention to Race		Attention to Age	
	Young Prime	Old Prime	Young Prime	Old Prime
White Prime	.57 (.17)	.51 (.19)	.56 (.20)	.52 (.22)
Black Prime	.54 (.18)	.48 (.19)	.52 (.23)	.49 (.22)

Note: Standard deviations are printed in parentheses.

Table 3

Overall effect sizes (Cohen's d), split-half correlations, and correlations to explicit preference scores of priming scores reflecting implicit preferences for whites over blacks (racism) and for young over old (ageism) derived from affective priming measures based on response interference (Fazio et al., 1995) and affect misattribution (Payne et al., 2005) as a function of attention to race versus age of face primes.

		Response Interference		Affect Misattribution	
		Racism	Ageism	Racism	Ageism
Effect size	Attention to race	.28	-.02	.14	.24
	Attention to age	.04	.24	.27	.16
Split-half correlation	Attention to race	.25	-.29	.58	.82
	Attention to age	-.13	.27	.57	.79
Correlation to explicit measure	Attention to race	.26	-.10	.30	.31
	Attention to age	-.05	.39	.29	.29

Note: Effect sizes with negative signs indicate priming effects in the opposite direction of the respective preference scores.