Prevention of intention invention in the affect misattribution procedure

Bertram Gawronski & Yang Ye

The University of Western Ontario

Please address correspondence to:

Bertram Gawronski
Department of Psychology
The University of Western Ontario
Social Science Centre
London, Ontario N6A 5C2
Canada

Email: bgawrons@uwo.ca
Phone: +1 (519) 661-2111 Ext. 87672

Abstract

The affect misattribution procedure (AMP) is one of the most promising indirect measures, showing high reliability and large effect sizes. However, the AMP has recently been criticized for being susceptible to explicit influences, in that priming effects tend to be larger and more reliable among participants who report that they intentionally responded to the primes instead of the targets. Consistent with interpretations of these effects in terms of retrospective confabulation, two experiments obtained reliable priming effects when (a) participants lacked meta-cognitive knowledge about their responses to the primes and (b) participants' attention was directed away from response-eliciting features of the primes. Under either of these conditions, priming effects were unrelated to self-reported intentionality, although self-reported intentionality was positively related to priming effects under control conditions. The findings highlight the contribution of meta-cognitive inferences to retrospective self-reports of intentionality and suggest an effective procedure to rule out explicit influences in the AMP.

The notion that attitudes can be activated unintentionally is one of the most significant ideas in the history of attitude research (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). This groundbreaking insight not only helped to explain varying degrees of attitude-behavior consistency (Fazio, 2007); it also served as the foundation for the development of a new class of indirect measures that assess attitudes by virtue of their unintentional effects on overt responses (for a review, see Gawronski & De Houwer, 2014). The two most prominent examples are the evaluative priming task (EPT; Fazio, Jackson, Dunton, & Williams, 1995) and the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998). According to De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009), the measurement outcomes of these instruments can be described as implicit to the extent that the to-be-measured psychological attribute influences measurement outcomes in an automatic fashion. Although the term *automatic* subsumes multiple distinct features (Bargh, 1994; Moors & De Houwer, 2006), one of the most central features in these tasks is intentionality, in that attitudes are assumed to influence measurement outcomes in a manner that does not require intention to evaluate the attitude object.

Despite the widespread use of the EPT and the IAT, either measure has been the target of criticism. Whereas the IAT has the advantage of showing large effect sizes and high reliability, its task structure makes it susceptible to various sources of systematic measurement error (Teige-Mocigemba, Klauer, & Sherman, 2010). Conversely, the EPT has the advantage of being based on the well-understood notion of sequential priming, but its effect sizes and reliability tend to be rather low (Wentura & Degner, 2010). At this point, one of the most promising alternatives to the IAT and the EPT is the affective misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) which combines the advantages of both measures. Similar to the IAT, AMP scores typically show high reliability and large effect sizes. At the same time, the AMP is based on the established notion of sequential priming, which makes it less susceptible to task-related

criticism than the IAT (Payne & Gawronski, 2010). Further support for the usefulness of the AMP comes from a recent meta-analysis confirming its validity in predicting various kinds of behaviors and real-world outcomes (Cameron, Brown-Iannuzzi, & Payne, 2012).

Despite these promising characteristics, the AMP has recently been criticized for being prone to explicit influences that could potentially undermine the implicit nature of its measurement outcomes (Bar-Anan & Nosek, 2012). To illustrate this criticism, it is useful to first explain the basic structure of the task. On a typical AMP trial, participants are briefly presented with a prime stimulus which is followed by a neutral target stimulus—usually a Chinese ideograph. After a short delay, the target stimulus is replaced by a black-and-white pattern mask and participants are asked to indicate if they consider the target stimulus visually more pleasant or visually less pleasant than the average Chinese ideograph. The modal finding is that the targets are evaluated more favorably when participants have been primed with a positive stimulus than when they have been primed with a negative stimulus. The most common interpretation of such priming effects is that the primes activate affective feelings or semantic concepts in memory, which are mistakenly attributed to the targets instead of the primes (Gawronski & Ye, 2014). From this perspective, the measurement outcomes of the AMP may be described as implicit in the sense that attitudes toward the primes unintentionally influence participants' responses to the targets.

Challenging this interpretation, Bar-Anan and Nosek (2012) recently presented a series of studies showing that priming effects in the AMP were larger and more reliable among participants who reported that they had intentionally rated the primes instead of the targets. These findings suggest that priming effects in the AMP may not result from unintentional influences of the primes, but instead reflect the extent to which participants intentionally use the primes in

judging the neutral targets. Needless to say, if this interpretation is correct, it would pose a serious challenge to the implicit nature of the AMP.

In response to Bar-Anan and Nosek's (2012) findings, Payne et al. (2013) reported a series of studies suggesting that the obtained relations between AMP effects and self-reported intentionality reflect retrospective confabulations rather than genuine effects of intentional processes. Specifically, Payne et al. found that AMP effects were related to incoherent self-reports of both intentional and unintentional influences of the primes. Moreover, giving participants the option to skip a target judgment when they felt that their judgment would be influenced by the prime failed to reduce priming effects. Taken together, these results suggest that relations between AMP effects and self-reported intentionality reflect retrospective confabulations of intentionality rather than genuine effects of intentional processes.

The current research expands on this debate by investigating the contribution of meta-cognitive inferences to retrospective self-reports of intentionality (Wegner & Wheatley, 1999). The central assumption underlying this research is that confabulations of intentionality depend on participants' naive assumptions about evaluative properties of the primes (Wilson & Brekke, 1994). Thus, although retrospective confabulations of intentionality may be positively related to priming effects when participants' naive assumptions about the primes are accurate, they may be unrelated to priming effects when (a) participants lack meta-cognitive knowledge about their actual responses to the primes and (b) their attention is directed away from response-eliciting features of the primes. Although either of these conditions should reduce the relation between AMP effects and self-reported intentionality, the size and construct validity of AMP effects should be unaffected. Evidence for these predictions would not only highlight the contribution of meta-cognitive inferences to retrospective self-reports of intentionality; it would also provide the basis for methodological refinements to rule out intentional responses to the primes in the AMP.

Experiment 1

Experiment 1 investigated the relation between priming effects in the AMP and self-reported intentionality when participants lack meta-cognitive knowledge about their actual responses to the primes. Toward this end, we used a mere exposure manipulation to influence evaluative responses toward unfamiliar stimuli (Zajonc, 1968). A well-replicated finding in research on mere exposure is that previously encountered stimuli elicit a favorable affective response due to the enhanced fluency in processing these stimuli (Winkielman, Huber, Kavanagh, & Schwarz, 2012). Importantly, although participants are typically able to verbally report their positive feelings when they are presented with a previously encountered stimulus, they tend to be unaware that their feelings are influenced by prior exposure to that stimulus (for a meta-analysis, see Bornstein, 1989). Because retrospective confabulations of intentionality occur in the absence of the relevant stimuli, participants have to rely on their meta-cognitive knowledge rather than momentarily experienced feelings when drawing inferences of intentionality. Thus, corresponding confabulations of intentionality should be undermined for evaluative responses resulting from mere exposure, thereby eliminating the relation between AMP effects and self-reported intentionality. To test this hypothesis, participants were presented with meaningless artificial words before they completed an AMP that included these words as prime stimuli. Participants in a control condition were provided with information about the positive meaning of the artificial words. Drawing on earlier findings by Gawronski and Ye (2014), we predicted that priming effects in the AMP should be sensitive to both mere exposure and positive information about the artificial words. However, self-reported intentionality should be related only to priming effects resulting from positive information (i.e., when participants have meta-cognitive knowledge about their actual responses), but not to priming effects resulting from mere exposure (i.e., when participants do not have meta-cognitive knowledge about their actual responses).

Method

Participants and design. A total of 100 undergraduates (77 women, 23 men) at the University of Western Ontario were recruited for a study entitled "First Impressions, Language, and Memory." The study was part of a one-hour session that included the current study and two additional studies on unrelated topics. Participants received research credit for an introductory psychology course. The study included a 2 (Word Type: presented vs. not presented) × 2 (Presentation Context: mere exposure vs. positive information) mixed-model design with the first variable as within-subjects factor and the second one as between-subjects factor.

Procedure. The main experimental manipulation was adopted from Gawronski and Ye (2014), involving presentations of artificial words as part of a language learning task. For half of the participants, the artificial words appeared individually on the screen (mere exposure condition). For the remaining half, the artificial words were presented together with positive English words that ostensibly described the meaning of the artificial words (positive information condition). Participants in both conditions were asked to memorize the artificial words. The presentations included five artificial words, each of which was presented 10 times for 1000ms slightly above the center of the screen. For participants in the positive information condition, a positive English word was simultaneously presented slightly below the center of the screen. The inter-trial interval was 2000ms. Order of trials was randomized individually for each participant. For the artificial words, we used two sets of five words. The artificial words of the first set were: *nijaron*, *kadirga*, *felkani*, *lokanta*, *safmeri*; the artificial words of the second set were: *vikesta*, *tunbalo*, *latipor*, *belnica*, *gorikas*. The artificial words of one set were presented as the target stimuli in the language learning task; the artificial words of the other set were used as baseline primes in the AMP without prior presentation. The use of the two sets as target stimuli versus

baseline primes was counterbalanced across participants. The English words in the positive information condition were: *love*, *friend*, *happiness*, *holiday*, *summer*.

After completion of the language learning task, participants were asked to complete the AMP, which was introduced as a concentration test. On each trial of the task, participants were first presented with a fixation cross for 500ms, which was replaced by one of the artificial words as a prime stimulus for 100ms. The presentation of the prime was followed by a blank screen for 100ms, after which a Chinese ideograph appeared for 100ms. The Chinese ideograph was then replaced by a black-and-white pattern mask, and participants were asked to make their response. Participants' task was to indicate if they considered the Chinese ideograph visually more pleasant or visually less pleasant than the average Chinese ideograph. The pattern mask remained on the screen until participants gave their response. The next trial started after an inter-trial interval of 500ms. Participants were asked to press a right-hand key (*Numpad 5*) when they considered the Chinese ideograph visually more pleasant than average and a left-hand key (*A*) when they considered the Chinese ideograph visually less pleasant than average. The task included 12 presentations of the five artificial words that were presented during the language learning task and 12 presentations of the five artificial words that were not presented before, summing up to a total of 120 trials. Order of trials was randomized by the computer for each participant. Following the instructions by Payne et al. (2005), participants were told that the artificial words can sometimes bias people's responses to the Chinese ideographs, and that they should try their absolute best not to let the words bias their judgments of the Chinese ideographs in any possible way. After participants had completed the AMP, they were asked if they intentionally rated the words instead of the Chinese ideographs when they completed the task. Self-reported intentionality was measured with Bar-Anan and Nosek's (2012) 5-point scale using the response

options (1) *not at all, I rated the ideographs*, (2) *usually no*, (3) *sometimes, but not always*, (4) *usually yes*, and (5) *yes, I rated the words*.

Results

Participants' responses on the AMP were aggregated by calculating the proportion of *more pleasant* responses for each type of prime stimuli (i.e., presented vs. not presented). Higher values on these scores indicate higher levels of positivity in response to the respective type of prime stimuli. Submitted to a 2 (Word Type) × 2 (Presentation Context) mixed-model ANOVA, these scores revealed a significant main effect of Word Type, $F(1, 98) = 25.52$, $p < .001$, $\eta_p^2 = .207$, indicating that artificial words that had been presented before elicited more favorable evaluations of the Chinese ideographs than artificial words that had not been presented before. This main effect was qualified by a significant two-way interaction of Word Type and Presentation Context, $F(1, 98) = 5.57$, $p = .02$, $\eta_p^2 = .054$ (see Figure 1), indicating that the obtained difference between artificial words that had been presented before and those that had not been presented before was more pronounced in the positive information condition, $F(1, 48) = 18.65$, $p < .001$, $\eta_p^2 = .280$, compared with the mere exposure condition, $F(1, 50) = 6.44$, $p = .01$, $\eta_p^2 = .114$. However, the effect of Word Type was statistically significant in both Presentation Context conditions, supporting the emergence of a mere exposure effect.

To investigate the relation between priming effects in the AMP and self-reported intentionality, we subtracted the mean proportion of *pleasant* responses to artificial words that had not been presented before from the mean proportion of artificial words that had been presented before. Higher values on this score indicate stronger priming effects in the AMP. Replicating the pattern obtained by Bar-Anan and Nosek (2012), priming effects were positively related to self-reported intentionality across the two experimental conditions ($r = .32$, $p = .001$), indicating that priming effects were more pronounced among participants who reported rating the

artificial words instead of the Chinese ideographs. However, this relation was primarily driven by a positive correlation between priming effects and self-reported intentionality in the positive information condition ($r = .51$, $p < .001$); there was no significant relation between priming effects and self-reported intentionality in the mere exposure condition ($r = .04$, $p = .78$). The difference between the two correlations was statistically significant, $Z = 2.50$, $p = .006$. Mean levels of self-reported intentionality did not significantly differ across the two conditions, $F(1, 98) = 1.05$, $p = .31$, $\eta_p^2 = .011$.

Following Bar-Anan and Nosek's (2012) approach, we also investigated potential differences in the reliability of the obtained priming effects. Toward this end, we calculated two separate priming scores, one using the first half of all priming trials and one using the second half. Across the two experimental conditions, AMP scores showed high internal consistency with a Cronbach's α of .84. Internal consistency did not differ as a function of whether priming effects were due to mere exposure (α = .81) or positive information (α = .85).

Discussion

The results of Experiment 1 provide further support for the hypothesis that relations between AMP effects and self-reported intentionality reflect retrospective confabulations rather than genuine effects of intentional processes. Consistent with this hypothesis, we found reliable priming effects when participants lacked meta-cognitive knowledge about their actual responses to the primes (i.e., favorable responses resulting from mere exposure of artificial words). Priming effects obtained under these conditions were unrelated to self-reported intentionality, although self-reported intentionality was positively related to priming effects when participants did have

meta-cognitive knowledge about their responses to the primes (i.e., favorable responses resulting from positive meaning of artificial words).[1]

Experiment 2

The main goal of Experiment 2 was to investigate confabulations of intentionality for individual differences in responses to social stimuli instead of experimentally created differences in responses to artificial stimuli. In addition, we aimed to extend our focus to participants' attention to response-eliciting features of the primes. Although attention to response-eliciting features is essential for reliable priming effects in the EPT, priming effects in the AMP have been shown to be independent of attention to response-eliciting features (Gawronski, Cunningham, LeBel, & Deutsch, 2010). In the current study, we utilized this characteristic to investigate differential relations of AMP effects to self-reported intentionality as a function of participants' attention to response-eliciting features of the primes. Our main hypothesis was that AMP effects should be unrelated to self-reported intentionality when participants' attention is directed away from response-eliciting features of the primes. Toward this end, participants were presented with primes showing faces of black and white men of either young or old age. Half of the participants were instructed to pay attention to the race of the face primes; the remaining half was instructed to pay attention to the age of the face primes (cf. Olson & Fazio, 2003). Drawing on earlier findings by Gawronski et al. (2010), we expected reliable priming effects of both race (i.e., preference for whites over blacks) and age (i.e., preference for young over old) regardless of attention instructions. Yet, priming effects of race were expected to be positively related to self-

[1] One reviewer correctly pointed out that the current study did not include a manipulation check of meta-cognitive knowledge. Although we agree that a manipulation check would further strengthen the findings of Experiment 1, the reviewer's concern implies another interesting possibility to test our hypothesis by informing participants' about the attitudinal effects of mere exposure. Whereas AMP effects and self-reported intentionality should be unrelated when participants do not have meta-cognitive knowledge about their actual responses to the primes, their relation should increase when participants are informed about the effects of mere exposure.

reported intentionality only when participants paid attention to race, but not when they paid attention to age. Conversely, priming effects of age were expected to be positively related to self-reported intentionality only when participants paid attention to age, but not when they paid attention to race.

Method

Participants and design. A total of 105 undergraduates at the University of Western Ontario (67 female, 38 male) were recruited for a study entitled "How Do We Perceive Unfamiliar Faces and Objects?". The study was part of a one-hour session that included the current study and two additional studies on unrelated topics. Participants received research credit for an introductory psychology course. The study included a 2 (Race: white vs. black) × 2 (Age: young vs. old) × 2 (Attention: race vs. age) mixed-model design with the first two variables as within-subjects factors and the last one as between-subjects factor.

Procedure. The materials and procedural details of the AMP were adopted from Gawronski et al. (2010). Each trial began with the presentation of a fixation cross for 500ms, which was replaced by a face prime for 75ms. The presentation of the prime was followed by a blank screen for 125ms, after which a Chinese ideograph appeared for 100ms. The Chinese ideograph was then replaced by a black-and-white pattern mask, and participants were asked to make their response. Participants' task was to indicate if they consider the Chinese ideograph more pleasant or less pleasant than the average Chinese ideograph, using a right-hand key (*Numpad 5*) for positive responses and a left-hand key (*A*) for negative responses. The inter-trial interval was 1000ms. The prime stimuli included 40 head-and-shoulder photographs of men, 10 for each of the four prime categories (i.e., young-white, old-white, young-black, old-black). Each of the 40 face primes was presented four times, summing up to a total of 160 randomized trials.

Participants were told that the faces can sometimes bias people's responses to the Chinese ideographs, and that they should try their absolute best not to let the faces bias their judgments of the Chinese ideographs in any possible way. To manipulate participants' attention in the AMP, half of the participants were instructed to keep a mental tally of how many black and white faces were presented over the course of the task; the remaining half were instructed to keep a mental tally of how many young and old faces were presented (cf. Olson & Fazio, 2003). After completion of the AMP, participants were asked to rate their feelings toward black people, white people, young people, and elderly people on 7-point scales ranging from 1 (*very cold*) to 7 (*very warm*). In addition, we asked participants to rate their gut reactions toward each of the 40 faces that were used as primes in the AMP. Self-reported gut reactions were assessed with 7-point scales ranging from 1 (*very negative*) to 7 (*very positive*). Self-reported intentionality was again measured with Bar-Anan and Nosek's (2012) 5-point scale using the response options (1) *not at all, I rated the ideographs*, (2) *usually no*, (3) *sometimes, but not always*, (4) *usually yes*, and (5) *yes, I rated the faces*.

Results

Participants' responses on the AMP were aggregated by calculating the proportion of *more pleasant* responses for each of the four prime categories (i.e., young-white, old-white, young-black, old-black). Higher values on these scores indicate higher levels of positivity in response to a given prime category. Following the data analytic procedure by Gawronski et al. (2010), an index of implicit preference for whites over blacks (*implicit racism*) was calculated by subtracting the mean positivity scores for black face primes from the mean positivity scores for white face primes. In addition, we calculated an index of implicit preference for young over old people (*implicit ageism*) by subtracting the mean positivity scores for old face primes from the mean positivity scores for young face primes. Scores of implicit racism and implicit ageism were

uncorrelated (r = -.09, p = .34). Corresponding preference scores were calculated for self-reported category evaluations by subtracting participants' ratings of black people from their ratings of white people and by subtracting their ratings of elderly people from their ratings of young people. Self-reported exemplar evaluations were aggregated by subtracting participants' average ratings of black faces from their average ratings of white faces and by subtracting their average ratings of elderly faces from their average ratings of young faces.

Submitted to a 2 (Type of Bias: implicit racism vs. implicit ageism) × 2 (Attention: race vs. age) mixed-model ANOVA, implicit preference scores revealed a statistically significant intercept, indicating a prejudice-related priming effect across experimental conditions, $F(1, 103) = 5.42$, $p = .02$, $\eta_p^2 = .050$. More importantly, the two-way interaction of Type of Bias and Attention was far from statistical significance, $F(1, 103) = 0.19$, $p = .67$, $\eta^2 = .002$, replicating Gawronski et al.'s (2010) finding that priming effects in the AMP do not require attention to response-eliciting features of the primes (see Figure 2). If anything, implicit ageism scores tended to be somewhat larger when participants paid attention to race than when they paid attention to age. However, the effect of Attention failed to reach statistical significance for both implicit racism, $F(1, 103) = 0.02$, $p = .90$, $\eta^2 < .001$, and implicit ageism, $F(1, 103) = 0.35$, $p = .56$, $\eta^2 = .003$.

Although the mean effect sizes of the two implicit preference scores were relatively small, they showed substantial correlations with their explicit counterparts regardless of whether explicit preferences were assessed at the category-level or exemplar-level (see Table 1).[2] Importantly, these correlations were also unaffected by attention instructions in the AMP. Implicit racism

[2] Note that the small effect sizes at the mean level do not necessarily indicate low reliability of the task, but may instead reflect lower average levels of implicit prejudice in our sample. The latter interpretation is supported by the high correlations to corresponding self-report measures and the high internal consistencies of the two preference scores.

scores were significantly correlated with explicit racism scores regardless of whether participants paid attention to race or age. The same was true for implicit ageism scores, which were significantly correlated with explicit ageism scores regardless of whether participants paid attention to age or race. A markedly different pattern emerged for self-reported intentionality, which was significantly correlated with implicit preference scores only when participants paid attention to the corresponding category (see Table 1). Specifically, implicit racism scores were significantly correlated with self-reported intentionality only when participants paid attention to race, but not when they paid attention to age. Conversely, implicit ageism scores were significantly correlated with self-reported intentionality when participants paid attention to age, but not when they paid attention to race. The difference between correlations was statistically significant for implicit racism, $Z = 1.84$, $p = .03$, and marginally significant for implicit ageism, $Z = 1.64$, $p = .05$. Mean levels of self-reported intentionality did not significantly differ across the two attention conditions, $F(1, 103) = 0.38$, $p = .54$, $\eta_p^2 = .004$.

To investigate the reliability of the obtained priming effects, we calculated two preference scores for each of the two types of bias, one using the first half of all priming trials and one using the second half. Across the two attention conditions, AMP scores showed moderate to high internal consistencies with Cronbach's α values of .71 for implicit racism and .88 for implicit ageism. More importantly, internal consistency of the two preference scores did not differ as a function of attention. Implicit racism scores showed acceptable internal consistencies regardless of whether participants paid to race ($\alpha = .74$) or age ($\alpha = .67$). Similarly, implicit ageism scores showed high internal consistency regardless of whether participants paid to age ($\alpha = .92$) or race ($\alpha = .81$).

Discussion

Experiment 2 investigated the relation between AMP effects and self-reported intentionality when participants' attention is directed away from response-eliciting features of the primes. Our results showed reliable priming effects of a given prime feature regardless of whether participants did or did not pay attention to that feature while completing the task. Yet, priming effects of a given feature were positively related to self-reported intentionality only when participants paid attention that feature, but not when their attention was directed toward an alternative feature. Together with the results of Experiment 1, these findings corroborate the hypothesis that relations between AMP effects and self-reported intentionality reflect retrospective confabulations rather than genuine effects of intentional processes.[3]

General Discussion

Expanding on the debate about the potential role of intentional processes in the AMP (Bar-Anan & Nosek, 2012; Payne et al., 2013), the current research investigated the relation between AMP effects and self-reported intentionality when (a) participants lacked meta-cognitive knowledge about their actual responses to the primes and (b) participants' attention was directed away from response-eliciting features of the primes. Across two experiments, we found reliable priming effects under either of these conditions. Although AMP effects were positively related to self-reported intentionality under control conditions, their relation was attenuated under conditions of absent meta-cognitive knowledge and lack of attention. Taken together, these results support Payne et al.'s (2013) conclusion that relations between AMP effects and self-

[3] A potential concern is that our attention manipulation influenced participants' interpretation of the intentionality question (i.e., which features of the primes are referred to in the self-report measure), implying the possibility that the measure was insensitive in capturing actual intentional processes for unattended category cues (cf. Shanks & St. John, 1994). We are currently engaged in follow-up research to rule out this concern.

reported intentionality reflect retrospective confabulations of intentionality rather than genuine effects of intentional processes.

A potential objection is that our findings rule out a causal role of intentional processes only for conditions of absent meta-cognitive knowledge and lack of attention. However, it is still possible that intentional processes contributed to priming effects when participants did have meta-cognitive knowledge about their actual responses to the primes (Experiment 1) and when they paid attention to response-eliciting features of the primes (Experiment 2). This assumption is consistent with the findings of Experiment 1 showing larger priming effects when participants did have meta-cognitive knowledge about their actual responses to the primes.[4] However, it is difficult to reconcile with the findings of Experiment 2 in which the overall size of priming effects did not differ as a function of attention (see also Gawronski et al., 2010).

Although the current data are insufficient to rule out a causal role of intentional processes when AMP effects are significantly related to self-reported intentionality, they do offer a simple, yet highly effective, procedure to address potential concerns in this regard. Specifically, our findings indicate that intentional ratings of the primes are less likely when participants' attention is directed away from the response-eliciting features of the primes. For example, research using the AMP to investigate racial attitudes may cross the manipulation of race with a manipulation of a race-unrelated category (e.g., gender, age) and direct participants' attention toward the race-unrelated category (e.g., by asking participants to keep a mental tally of male and female faces while completing the AMP). The current findings suggest that AMP scores reliably reflect racial attitudes even participants' attention is directed away from the racial category membership of the primes (see also Gawronski et al., 2010). More importantly, race-related priming effects were

[4] Note that larger priming effects in the positive translation condition could also be due to stronger evaluative responses to the primes compared to those in the mere exposure condition.

unrelated to self-reported intentionality under such conditions, suggesting that intentional processes did not play a causal role in the obtained priming effects. Thus, in addition to highlighting the contribution of meta-cognitive inferences to retrospective self-reports of intentionality, our results suggest a simple, yet highly effective, procedure to rule out intentional responses to the primes in the AMP. Hence, counter to concerns that AMP scores may be contaminated by intentional processes, we believe that the AMP still represents one of the most promising alternatives to the EPT and the IAT.

References

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the Affective Misattribution Procedure. *Personality and Social Psychology Bulletin, 38,* 1194-1208.

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* (pp. 1-40). Hillsdale, NJ: Erlbaum.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research. *Psychological Bulletin, 106,* 265-289.

Cameron, C. D., Brown-Iannuzzi, J., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review, 16,* 330-350.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135,* 347–368.

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., 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.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion, 24,* 1008-1025.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition, pp. 283-310). New York: Cambridge University Press.

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin*, *40,* 3-15.

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.

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin, 132,* 297-326.

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., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the Affect Misattribution Procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin, 39,* 375-386.

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.

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1-15). New York: Guilford Press.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367-447.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to the Implicit Association Test and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117-139). New York: Guilford Press.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54,* 480-492.

Wentura, D. & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95 - 116). New York: Guilford Press.

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

Winkielman, P., Huber, D. E., Kavanagh, L., & Schwarz, N. (2012). Fluency of consistency: When thoughts fit nicely and flow smoothly. In B. Gawronski, & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 89–111). New York: Guilford Press.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9,* 1-27.

Table 1. Correlations of implicit preference for whites over blacks (implicit racism) and implicit preference for young over old (implicit ageism) to corresponding explicit preferences and self-reported intentionality as a function of attention to race versus age in the AMP, Experiment 2.

		Implicit Racism	Implicit Ageism
Explicit category preference	Attention to race	.34*	.33*
	Attention to age	.36*	.31*
Explicit exemplar preference	Attention to race	.58***	.42**
	Attention to age	.42**	.48***
Self-reported intentionality	Attention to race	.49***	.19
	Attention to age	.16	.48***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

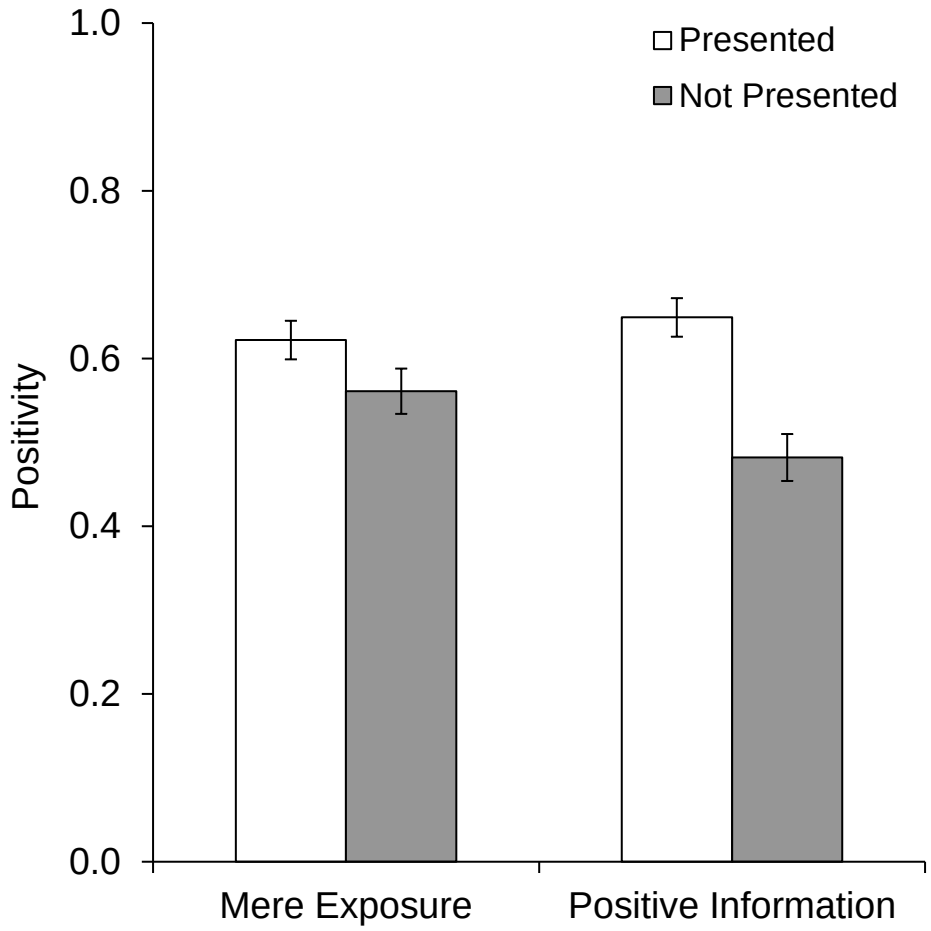

Figure 1. Priming effects of artificial words as a function of prior presentation (presented vs. not presented) and context during prior presentation (mere exposure vs. positive information), Experiment 1. Higher values indicate higher proportions of positive responses. Error bars depict standard errors.

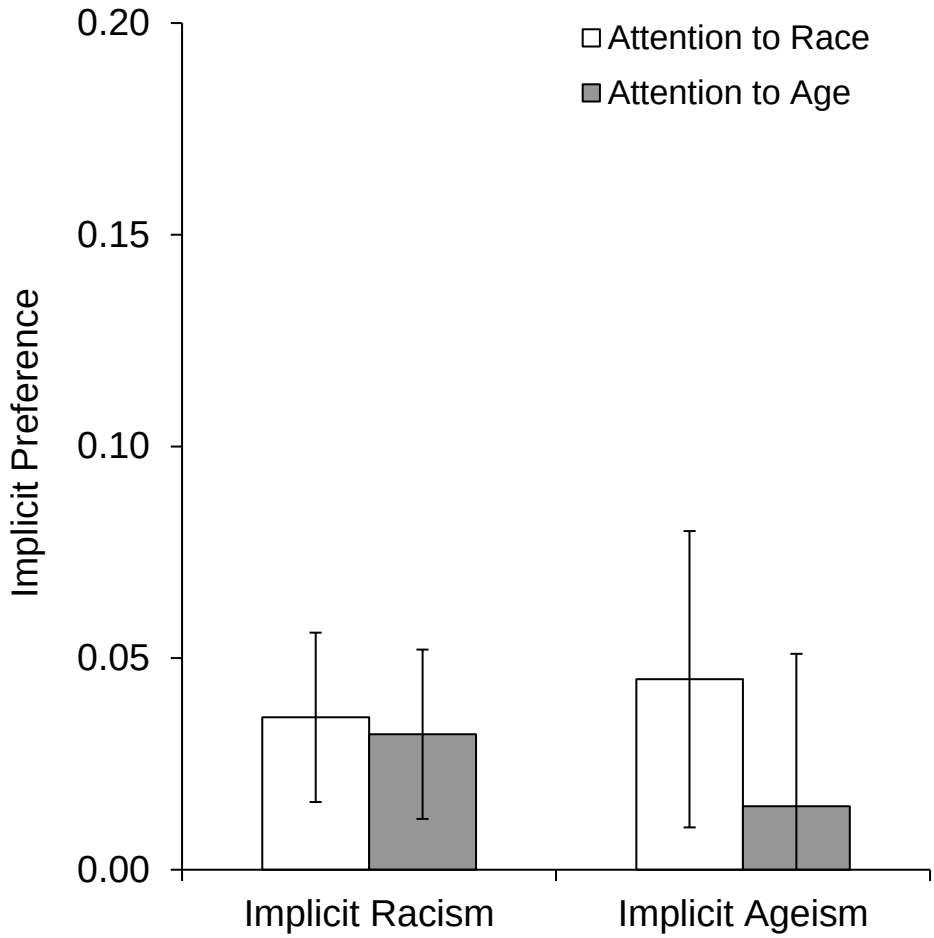

Figure 2. Priming effects reflecting implicit preference for whites over blacks (implicit racism) and implicit preference for young over old (implicit ageism) as a function of attention to race versus age of face primes, Experiment 2. Error bars depict standard errors.