Personality and Social Psychology Bulletin
http://psp.sagepub.com/

Reporting Intentional Rating of the Primes Predicts Priming Effects in the Affective Misattribution Procedure

Yoav Bar-Anan and Brian A. Nosek

Pers Soc Psychol Bull published online 18 May 2012
DOI: 10.1177/0146167212446835

Personality and Social
Psychology Bulletin
XX(X) 1–15

DOI: 10.1177/0146167212446835
http://pspb.sagepub.com

Reporting Intentional Rating of the Primes Predicts Priming Effects in the Affective Misattribution Procedure

Yoav Bar-Anan[1] and Brian A. Nosek[2]

Abstract

In the affective misattribution procedure (AMP), pairs of prime and target stimuli appear rapidly in succession. Attitudes toward the prime influence the evaluation of the target despite instructions to avoid this influence. Because this priming effect presumably happens without people's knowledge, the AMP is used to study automatic evaluation. Participants in four studies performed the AMP and reported their perception of the priming effect. The authors found that the priming reflected reliable and valid attitudes toward the primes mostly among participants who reported that the priming occurred and that they intentionally rated the primes instead of the targets. The authors conclude that the AMP hardly captures attitude effects that escape people's knowledge. The AMP's good psychometric qualities as an attitude measure rely mainly on a small subset of participants who believe that they intentionally caused the attitude effect.

Keywords

automatic evaluation, implicit attitudes, implicit measures, the affective misattribution procedure, illusory control

Received October 22, 2011; revision accepted February 20, 2012

Attitudes can influence behavior without intention (Bargh, Chaiken, Govender, & Pratto, 1992; Duckworth, Bargh, Garcia, & Chaiken, 2002; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), and it appears that this can happen without awareness (Bargh & Pietromonaco, 1982; Berridge & Winkielman, 2003; Dovidio, Kawakami, & Gaertner, 2002; Gawronski, Geschke, & Banse, 2003; for reviews, see Gawronski, Hofmann, & Wilbur, 2006; Uhlmann, Pizarro, & Bloom, 2008). A number of tasks, collectively called *implicit measures*, are designed to measure unintentional attitude effects (Wittenbrink & Schwarz, 2007). There is little evidence of whether people know that unintentional evaluations influenced their performance in any of these tasks (for an exception, see Monteith, Voils, & Ashburn-Nardo, 2001). The present research investigated people's perception of the attitude effects in the affective misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005).

In the AMP, participants observe a prime (e.g., a picture of Hillary Clinton), quickly followed by an unfamiliar target (a Chinese pictograph for non-Chinese readers). Participants are instructed to ignore the prime and evaluate the target as pleasant or unpleasant. Despite the instructions, Payne et al. (2005) found that primes nonetheless influenced judgments of the targets. People rated targets that appeared after pleasant stimuli (e.g., a baby) as more pleasant than targets that appeared after unpleasant stimuli (e.g., a snake). The presumed mechanism of this effect was that the primes automatically activate a positive or negative reaction that is misattributed to the target.

The AMP is gaining widespread use as an implicit measure (Nosek, Hawkins, & Frazier, 2011; for example, Gawronski & Walther, 2008; Hofmann, Friese, & Roefs, 2009; Miller, Brewer, & Arbuckle, 2009; Oikawa, Aarts, & Oikawa, 2011; Rydell, McConnell, & Mackie, 2008; Shariff & Tracy, 2009), probably because it shows good internal consistency and because it seems to have straightforward procedural validity: Attitudes affect performance despite instructions to prevent it. Payne et al. (2005) hypothesized that participants failed to prevent the effect because they were unaware of it:

[1]Ben-Gurion University of the Negev, Be'er Sheva, Israel
[2]University of Virginia, Charlottesville, USA

Corresponding Author:
Yoav Bar-Anan, Department of Psychology, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel
Email: baranany@bgu.ac.il

> We suspect that if participants recognized that their judgment on any given trial was being influenced by the prime, they would be able to correct by simply giving the opposite response . . . the effect was difficult to control because participants did not believe they were experiencing it. (p. 291)

Following Payne et al.'s (2005) hypothesis, the present research tests people's beliefs about the priming effect.

If people are unaware of attitudes affecting their AMP performance, it would be a unique feature in comparison with other popular implicit measures. For example, although some evidence suggests that people cannot predict in advance their performance on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) for race-relevant topics (Mitchell, Nosek, & Banaji, 2003), during task performance many participants have a palpable experience of having more difficulty associating good words with Black faces compared with White faces. One study found that about 60% of American participants were able to report about the effect of their attitudes on their IAT performance (Monteith et al., 2001). An advantage of the AMP might therefore be as a reliable assessment after which participants are still unaware of having demonstrated attitude effects. Furthermore, Payne et al.'s (2005) quote above suggests an important theoretical implication—that the effect of attitudes in the AMP *depends* on people's unawareness of that effect.

The present research used retrospective self-reports to assess whether people think that the primes influenced the evaluation of the targets and whether they think that this influence was unintentional. People's retrospective self-report about psychological effects is one of the most frequent measures of awareness. For example, Nisbett and Wilson (1977) used inaccurate retrospective reports as evidence that people lack introspection to many high-level mental processes. In research of automatic social behavior (e.g., automatic goal pursuit; Aarts, Gollwitzer, & Hassin, 2004; Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trotschel, 2001; Hassin, Bargh, & Zimerman, 2009), retrospective reports are used to support the researchers' claims that these effects occur without people's awareness. Research of the roles of awareness in learning has also used retrospective self-reports of the learned rules as one of the main measures to assess whether learning involved awareness (e.g., Field, 2000; Olson & Fazio, 2001; Stahl & Unkelbach, 2009; Walther, 2002).

Despite its widespread use, retrospective self-report is not a perfect measure of awareness (Reingold & Merikle, 1988). Self-reports might fail to detect awareness of a psychological process if the awareness was forgotten, if people pretend to be unaware, or if people fail to verbalize their awareness (e.g., misunderstand the questions' meaning; Ericsson & Simon, 1984). Self-reports might erroneously detect awareness of a psychological process despite unawareness if people pretend to have awareness, or if awareness was gained after the process had ended, even because of asking about awareness (Gawronski & Walther, 2012; Lovibond & Shanks, 2002). An example for the limitation of self-report as a measure of awareness comes from research on evaluative conditioning—the effect of pairing neutral stimuli with affective stimuli on the evaluation of the neutral stimuli (De Houwer, Thomas, & Baeyens, 2001). The common measure of awareness is self-report about the pairing, collected retrospectively. However, people's reports about the pairing are sometimes the result of inference from their own attitudes rather than an actual awareness of the pairing during the procedure (Bar-Anan, De Houwer, & Nosek, 2010; see Gawronski & Walther, 2012, for more discussion).

Retrospective self-report is also not a perfect measure of intentional behavior. Reporting about an intentional act suggests that the reporter remembers intending to perform the act and also believes that the intention caused the act (Wegner, 2004, p. 4). Remembering intention suffers from the limitations of false memory, forgetting, and lying. And, even if the participant's memory of intention is accurate, this does not guarantee that the intention caused the effect (e.g., Kühn & Brass, 2009; Pronin, Wegner, McCarthy, & Rodriguez, 2006; Wegner & Wheatley, 1999).

Notwithstanding the flaws of self-report as a measure of awareness and intention, self-report is a reasonable measure of people's beliefs. In the context of automatic evaluation, beliefs are important because they have implications. If people think that automatic evaluation biased their judgment, they can try to compensate for this bias later on, or to justify it (Monteith, 1993). If people think that they intentionally caused the bias, then they are more likely to take responsibility of their actions and try to correct them (Mueller & Dweck, 1998; Vohs & Schooler, 2008). Furthermore, self-reports about whether automatic evaluation biased judgment can also shed light on whether people have accurate knowledge regarding their biases. If people tend to report the priming after the priming actually occurred, then it means that people hold accurate knowledge about the automatic evaluation effects.

Finally, if people's retrospective reports are linked to the actual priming effect, then it means that people's beliefs regarding the priming effect moderate the effect. Uncovering moderators can help in identifying populations that are more sensitive to an effect than other populations. For instance, had research found that the AMP works mostly on people younger than 20 years, it would have suggested an important limitation to the procedure as a measure of attitudes and as a procedure to investigate automatic evaluation. To anticipate our findings, the present research found such limitation: The AMP works mostly on people who believe that priming occurred, and mainly on those who believe that they intentionally caused the priming.

Overview of the Studies

Studies 1A and 1B recorded people's beliefs about the priming effect in an AMP that measure attitudes toward racial groups (Study 1A) or political parties (Study 1B). These studies found wide variability in people's belief about whether biased evaluation occurred despite instructions to avoid it. The studies also examined whether these beliefs are related to the actual priming effect and found that people generally have an accurate knowledge of their biased evaluation. Finally, Studies 1A and 1B found that the AMP is a reliable and valid attitude measure mainly in the subset of participants who believe that, at least sometimes, they intentionally rated the primes instead of the targets. Study 2 replicated those findings with newly formed attitudes.

To investigate the source of people's knowledge, Study 3 manipulated the strength of the priming effect without changing people's attitudes by varying the duration of the target's presentation. Longer target presentation duration decreased the priming effect and also decreased people's reports that they intentionally caused it. This suggests that the perception of intentionally rating the primes is based on experience during the AMP, and not on an inference from people's knowledge of their own attitudes, separate from the AMP itself.

Study 1A

Study 1A measured race attitudes with self-reported thermometer ratings, the AMP, and the IAT (see Nosek, Greenwald, & Banaji, 2007, for a review). After performing the AMP, participants reported what they thought influenced their evaluation of the targets. Previous research showed positive relationships between the AMP and self-reported race attitudes (e.g., r = .58, Payne et al., 2005), and between the AMP and the IAT (r = .22, Greenwald, Smith, Sriram, Bar-Anan, & Nosek, 2009; r = .27, Bar-Anan & Nosek, 2012), which are both considered implicit measures that are sensitive to automatic evaluations. We investigated whether participant's self-reported beliefs of the causes of the AMP effect moderated its internal consistency and its relationship with self-reported attitudes and the IAT.

Method

Participants. The participants in all four studies came from the volunteer participant pool at the Project Implicit research website (https://implicit.harvard.edu; see Nosek, 2005, for more information). A total of 403 participants (68% women, 31% men, 1% unknown; 71% White, 6% Black, 15% other, 8% unknown; M age = 26.3, SD = 12.1) completed the study. We removed participants who always responded unpleasant (n = 1) or pleasant (24) in the AMP, reported reading Chinese (20), and/or responded rapidly (under 300 ms) to more than 10% of the IAT trials (6), leaving 354 (88%) participants in the analyses.

Measures

Stimuli. The AMP and the IAT used the same stimuli. The stimuli were three White males and three Black males (a subset of the stimuli used by Gawronski, Cunningham, LeBel, & Deutsch, 2010). The evaluative terms in the IAT were words clearly denoting good or bad meaning, and the categories were *Black people* and *White people*.

AMP. The instructions explicitly stated that the task is to ignore the first picture of every trial, concentrate on the Chinese drawing, and decide whether it is more pleasant or more unpleasant than the average Chinese drawing. After three practice trials, the instructions appeared again and the task began. The prime stimulus appeared for 100 ms, followed by a blank screen for 100 ms, and then the target stimulus appeared for 100 ms. A mask stimulus appeared at the end of the trial until participants evaluated the target stimulus as pleasant or unpleasant. The intertrial interval (ITI) was 250 ms. We used two blocks of 36 trials. Each block presented 12 primes of White people, 12 primes of Black people, and 12 gray rectangle primes.

IAT. The IAT measured the implicit preference between White and Black people. We used a standard seven-block IAT, identical to the one described in Nosek, Smyth, et al. (2007). We scored the IAT following recommendations of Greenwald, Nosek, and Banaji (2003).

Self-report. A thermometer rating probed participants' feelings toward each social group on scale from 0 = *the coldest* to 8 = *the warmest*. Participants also reported their liking of each person who appeared as an item in the tasks, on a 6-point scale, ranging from *dislike strongly* to *like strongly*.

The AMP experience. Following the AMP, participants completed a questionnaire with the question, "Do you think your ratings of the drawings were influenced by the face pictures that appeared before the drawings?" with the response options [No], [I'm not sure, but I think they did not], [I'm not sure, but I think they did], and [Yes], coded, respectively, from −1.5 to 1.5, with 1 point difference between responses. Another question tapped perceived intentional primes rating: "Did you intentionally rate the face pictures instead of the drawings?" with the response options [Not at all, I rated the drawings], [Usually no], [Sometimes, but not always], [Usually yes], and [Yes, I rated the pictures]. The response options were coded from −2 to 2 with 1 point difference between responses.

Procedure

Participants completed the three attitude measures in a random order. The AMP experience questionnaire always followed

Table 1. Studies 1A and 1B: Means and Intercorrelation of Attitude Measures

	AMP	IAT	Category rating	Item rating
Study 1A				
Reliability	.59	.55		.50
M (*SD*)	−0.02* (0.20)	0.28*** (0.40)	0.49*** (1.65)	−0.37*** (0.88)
AMP		.22***	.34***	.40***
IAT			.27***	.21***
Category rating				.43***
Study 1B				
Reliability	.69	.83		.89
M (*SD*)	0.09*** (0.28)	0.30*** (0.58)	2.55*** (4.43)	1.30*** (2.24)
AMP		.39***	.50***	.53***
IAT			.65***	.74***
Category rating				.85***
Study 3				
Reliability	.63	.59		
M (*SD*)	−0.04*** (0.22)	0.28*** (0.40)	0.57*** (1.87)	
AMP		.26***	.32***	
IAT			.31***	

Note: AMP = affective misattribution procedure; IAT = Implicit Association Test. Positive values mean more pro-White or pro-Democrats attitudes; reliability was the Cronbach's alpha computed of two halves for the AMP and the IAT, and three difference scores in the item rating of Studies 1A and 1B.
*p < .10. **p < .05. ***p < .01.

the AMP. In all the studies reported here, the measures order did not moderate the results.

Results and Discussion

We computed the AMP preference score by subtracting the pleasant response rate after Black primes from the pleasant response rate after White primes. The preference score is larger when the priming had a larger effect. However, priming can occur even if the preference score is zero (e.g., an 80% pleasant response rate after Black and White primes, and a 20% rate after neutral primes). Therefore, we computed a score for the size of the priming effect as the average absolute difference between the rate of pleasant judgments after neutral (gray rectangles) primes versus each of the other two categories. This measure can detect a scenario in which primes of both categories cause the same priming effect (e.g., more pleasant ratings) than the baseline priming effect, measured in neutral trials. Unless stated otherwise, the statistical significance of all the reported relationships and the differences have $p < .05$. The means of all measures and their intercorrelations are presented in Table 1. The AMP showed a small but significant overall preference for Black people over White people ($d = 0.10$). This was compatible with an explicit preference for the Black stimuli over the White stimuli in the item rating measures ($d = 0.42$). The pro-Black AMP effect could be interpreted as deliberate alteration because of motivation to appear unprejudiced. However, this seems unlikely because the participants explicitly reported a pro-White preference when rating the social groups ($d = 0.30$). The Black stimuli may just have had more favorable features than the White stimuli.

Reported priming effect (RPE). The RPE is the participants' response to the question, "Do you think your ratings of the drawings were influenced by the face pictures that appeared before the drawings?" The findings with RPE are presented in Table 2. RPE predicted the priming effect size, $r(354) = .34$, suggesting that people showed some accuracy in detecting the priming. People who reported no effect at all (36% of the sample) still showed a significant priming effect, which supports the possibility that priming sometimes occurred without awareness. The most important findings that Table 2 presents are that the AMP score of people who reported no priming effect showed very poor psychometric qualities: no significant split-half correlation and no relationship with any of the other attitude measures. These psychometric qualities improved among people who reported more confidence that priming occurred; the participants who reported that priming occurred (the highest levels of RPE) showed significantly stronger priming effect, split-half correlation, and relationship with other attitude measures than participants who reported that priming did not occur.

Reported intention. Participants answered, "Did you intentionally rate the face pictures instead of the drawings?" As detailed in Table 3, a little more than half of the participants reported that they never intentionally rated the primes instead of the targets. Reporting intentions was related to stronger priming effects, $r(354) = .46$. As Table 3 details, participants who reported no intentional prime rating showed very little evidence that the AMP measured their attitudes. Their AMP

Table 2. Studies 1 and 2: The Relationship Between RPE and Various Measures

"Do you think your ratings of the drawings were influenced by the faces?"				
Response options	No	No (unsure)	Yes (unsure)	Yes
Study 1A				
Frequency	127 (36%)	81 (23%)	82 (23%)	63 (18%)
AMP White preference	$-.03_{a}$***	$-.04_{a}$***	$.03_{b}$	$-.02_{a}$
Priming effect	$.13_{a}$***	$.13_{a}$***	$.17_{b}$***	$.30_{c}$***
AMP split-half correlation	$.08_{a}$	$.25_{ab}$**	$.46_{bc}$***	$.67_{c}$***
AMP/category rating correlation	$.00_{a}$	$.22_{ab}$**	$.37_{bc}$***	$.58_{c}$***
AMP/item rating correlation	$.15_{a}$	$.08_{a}$	$.56_{b}$***	$.54_{b}$***
AMP/IAT correlation	$.10_{a}$	$.16_{ab}$	$.24_{ab}$**	$.39_{b}$***
IAT/category rating correlation	$.28_{a}$***	$.23_{a}$**	$.31_{a}$***	$.30_{a}$**
Study 1B				
Frequency	87 (23%)	98 (26%)	80 (21%)	117 (31%)
AMP Democrats preference	$.03_{a}$***	$.04_{a}$***	$.09_{ab}$***	$.17_{b}$***
Priming effect	$.14_{a}$***	$.14_{a}$***	$.19_{b}$***	$.28_{c}$***
AMP split-half correlation	$.30_{a}$***	$.22_{a}$**	$.38_{a}$***	$.66_{b}$***
AMP/category rating correlation	$.36_{ab}$***	$.13_{a}$	$.50_{bc}$***	$.66_{c}$***
AMP/item rating correlation	$.36_{ab}$***	$.17_{a}$	$.53_{bc}$***	$.71_{c}$***
AMP/IAT correlation	$.31_{ab}$***	$.16_{a}$	$.42_{ab}$***	$.53_{b}$***
IAT/category rating correlation	$.69_{b}$***	$.41_{a}$***	$.68_{b}$***	$.72_{b}$***
Study 2				
Frequency	170 (32%)	194 (36%)	104 (20%)	65 (12%)
AMP: Preference for the positive individual	$.01_{a}$ (n = 117)	$.00_{a}$ (n = 125)	$.01_{a}$ (n = 64)	$.33_{b}$*** (n = 56)
Self-reported preference for the positive individual	2.48_{a}***	2.28_{a}***	2.52_{a}***	3.76_{b}***
Priming effect	$.12_{a}$***	$.12_{a}$***	$.20_{b}$***	$.39_{c}$***
AMP split-half correlation	$.11_{a}$	$.15_{a}$**	$.65_{b}$***	$.80_{c}$**
AMP/explicit correlation	$.07_{a}$	$.11_{a}$	$.45_{b}$***	$.66_{c}$***

Note: RPE = reported priming effect; AMP = affective misattribution procedure; IAT = Implicit Association Test. The response options referred to the question: "Do you think your ratings of the drawings were influenced by the faces?" The drawings were the targets and the faces were the primes. The preference scores in Study 2 did not include participants from the no-difference induction condition. The significance test for the extremity scores was difference from zero (i.e., no preference). The priming effect is the average difference between responses in the neutral trials and trials of each of the attitude objects. In each row, values that do not share subscripts are significantly different from each other.
**p < .05. ***p < .01.

scores showed nonsignificant split-half correlation or relationship with other attitude measures.

It is possible that people who report no intentions have weak attitudes that fail to elicit effects with any measures—that is, the AMP could be ineffective because the attitudes had very low variability, not because the AMP did not detect them. However, other measures detected meaningful variability in these people's attitudes. The self-reported attitudes and the IAT scores of people who reported no intention at all still showed variability and a moderate positive intercorrelation, no different than the correlation found for the rest of the sample (Table 3). Although neither the IAT nor the self-report on its own could be taken as gold standard of revealing attitudes, their correlation shows that there is meaningful variability in the attitudes of people who reported no intention to rate the primes, but that variability was not detected by the AMP.

The rest of the sample—people who reported some intentional rating of the primes—showed much better evidence that the AMP measured their attitudes. Their priming effect, split-half correlation, and relationship with other measures were significantly stronger, compared with participants who reported no intentional rating of the primes.

Table 4 illustrates the most important finding of this study. In all, 34% percent of the sample—those who reported that they at least occasionally rated the primes and/or that they noticed the primes—were the main contributors of the AMP's psychometric qualities. Without them, the AMP showed little to no indication that the priming effect was related to the participants' attitudes.

Study 1B

Study 1B was identical to Study 1A, only measuring political attitudes instead of racial attitudes. The results generally replicated Study 1A's results. However, the study differs from the other studies in this article because it found better evidence that the AMP measured attitudes even when participants reported that priming did not occur.

Table 3. Study 1: The Relationship Between Reporting Intentional Rating and Various Measures

"Did you intentionally rate the face pictures instead of the drawings?"

Response option	Never	Usually no	Sometimes	Usually yes	Yes
Study 1A					
Frequency	180 (51%)	93 (26%)	55 (16%)	15 (4%)	11 (3%)
AMP White preference	−.03$_{a}$***	−.02$_{a}$	.02$_{b}$	−.05$_{ab}$	−.09$_{ab}$
Priming effect	.11$_{a}$***	.18$_{b}$***	.21$_{c}$***	.40$_{d}$***	.38$_{d}$***
AMP split-half correlation	.11$_{a}$	.44$_{b}$***	.50$_{b}$***	.70$_{b}$***	.83$_{b}$***
AMP/category rating correlation	.04$_{a}$	.44$_{b}$***	.25$_{ab}$	.76$_{b}$***	.76$_{b}$***
AMP/item rating correlation	.11$_{a}$	.50$_{b}$***	.52$_{b}$***	.23$_{ab}$***	.85$_{b}$***
AMP/IAT correlation	.04$_{a}$	.31$_{b}$***	.19$_{ab}$	.77$_{c}$***	.66$_{bc}$**
IAT/category rating correlation	.28$_{a}$***	.32$_{a}$***	.13$_{a}$	.37$_{a}$	.45$_{a}$
Study 1B					
Frequency	155 (41%)	98 (27%)	73 (19%)	22 (6%)	34 (9%)
AMP Democrats preference	.04$_{a}$***	.06$_{a}$***	.15$_{b}$***	.14$_{b}$	.21$_{b}$**
Priming effect	.13$_{a}$***	.18$_{b}$***	.22$_{b}$***	.22$_{ab}$***	.41$_{c}$***
AMP split-half correlation	.24$_{a}$***	.34$_{a}$***	.63$_{b}$***	.37$_{ab}$	.84$_{c}$**
AMP/category rating correlation	.16$_{a}$	.52$_{b}$***	.56$_{b}$***	.68$_{ab}$***	.74$_{b}$***
AMP/item rating correlation	.18$_{a}$**	.61$_{b}$***	.65$_{b}$***	.71$_{b}$***	.74$_{b}$***
AMP/IAT correlation	.19$_{a}$**	.50$_{b}$***	.38$_{ab}$***	.55$_{b}$***	.60$_{b}$**
IAT/category rating correlation	.61$_{a}$***	.71$_{a}$***	.65$_{a}$***	.66$_{a}$***	.68$_{a}$***
Study 2					
Frequency	333 (62%)	117 (22%)	47 (9%)	16 (3%)	20 (4%)
AMP: Preference for the positive individual	.00$_{a}$ (n = 213)	.01$_{a}$ (n = 79)	.13$_{a}$ (n = 34)	.53$_{b}$*** (n = 14)	.34$_{b}$** (n = 20)
Self-report: Preference for the positive individual	2.22$_{a}$***	2.87$_{b}$***	3.12$_{bc}$***	4.43$_{c}$***	3.55$_{bc}$***
Priming effect	.13$_{a}$***	.16$_{b}$***	.23$_{c}$***	.35$_{d}$***	.32$_{d}$***
AMP split-half correlation	.16$_{a}$***	.49$_{b}$***	.63$_{b}$***	.90$_{c}$***	.90$_{c}$***
AMP/explicit correlation	.08$_{a}$	.31$_{b}$***	.63$_{c}$***	.87$_{c}$***	.70$_{c}$***
Study 3					
Frequency	333 (47%)	191 (27%)	119 (17%)	28 (4%)	31 (4%)
AMP White preference	−.04$_{a}$***	−.04$_{a}$***	−.01$_{a}$	−.02	−.06$_{a}$
Priming effect	.12$_{a}$***	.17$_{b}$***	.18$_{b}$***	.30$_{c}$***	.35$_{c}$***
AMP split-half correlation	.13$_{a}$**	.49$_{b}$***	.62$_{bc}$***	.80$_{c}$***	.69$_{bc}$***
AMP/category rating correlation	.15$_{a}$***	.40$_{b}$***	.46$_{b}$***	.29$_{ab}$	.27$_{ab}$
AMP/IAT correlation	.20$_{a}$***	.34$_{a}$***	.33$_{a}$***	.07$_{a}$	.31$_{a}$
IAT/category rating correlation	.28$_{a}$***	.31$_{a}$***	.33$_{a}$***	.42$_{a}$***	.43$_{a}$***

Note: AMP = affective misattribution procedure; IAT = Implicit Association Test. The response options referred to the question: "Did you intentionally rate the face pictures instead of the drawings?" The drawings were the targets and the faces were the primes. The preference score in Study 2 did not include participants from the no-difference induction condition. The significance test for the extremity scores was difference from zero (i.e., no preference). The priming effect is the average difference between responses in the neutral trials and trials of each of the attitude objects. Different subscripts indicate significant difference in the same row.
**$p < .05$. ***$p < .01$.

Method

Participants. A total of 436 participants (69% women, 30% men, 1% unknown; M age = 26.3, SD = 10.7) completed the study. We removed participants who always responded unpleasant (n = 1) or pleasant (18), or reported reading Chinese (26), or responded rapidly to too many IAT trials (10), leaving 382 (88%) participants in the analyses.

Measures. The measure procedures were identical to Study 1A, only with different stimuli: images of American politicians, three from each main political party. The category labels in the IAT were *Democrats* and *Republicans*.

Results

The means of all measures and the intercorrelations between them are presented in Table 1. The relationship between RPE and the priming effect was $r(382) = .37$, suggesting, again, that people have some knowledge about the priming effect. Table 2 presents the relationship between RPE and the

Table 4. Studies 1 to 3: The Contribution of Participants Who Reported Noticing or Causing the Priming to the AMP's Overall Psychometric Qualities

Response options	All participants	Reported no awareness or no intention	Reported awareness and intention
Study 1A			
Sample size	354	235 (66%)	119 (34%)
AMP split-half correlation	.41***	.16**	.57$_a$***
AMP/category rating correlation	.34***	.08	.51$_a$***
AMP/item rating correlation	.37***	.10	.54$_a$***
AMP/IAT correlation	.22***	.10	.37$_a$***
IAT/category rating correlation	.27***	.25***	.31***
Study 1B			
Frequency	382	207 (54%)	175 (44%)
AMP split-half correlation	.53***	.32***	.60$_a$***
AMP/category rating correlation	.50***	.27***	.62$_a$***
AMP/item rating correlation	.53***	.30***	.67$_a$***
AMP/IAT correlation	.39***	.23***	.51$_a$***
IAT/category rating correlation	.65***	.60***	.71***
Study 2			
Frequency	533	271 (51%)	262 (49%)
AMP: Preference for the positive individual	0.05***	0.00	0.20$_a$***
Explicit preference for the positive individual	2.61***	2.36***	3.37$_a$***
AMP split-half correlation	.54***	.17***	.81$_a$***
AMP/explicit correlation	.36***	.12**	.62$_a$***
Study 3: Short target duration			
Sample size	365	209 (57%)	156 (54%)
AMP split-half correlation	.52***	.24***	.59$_a$***
AMP/category rating correlation	.32***	.12	.38$_a$***
AMP/IAT correlation	.25***	.16**	.30***
IAT/category rating correlation	.34***	.36***	.31***
Study 3: Long target duration			
Sample size	337	230 (68%)	107 (32%)
AMP split-half correlation	.37***	.25***	.47$_a$***
AMP/category rating correlation	.30***	.20***	.32***
AMP/IAT correlation	.27***	.17**	.32***
IAT/category rating correlation	.27***	.20***	.32***

Note: AMP = affective misattribution procedure; IAT = Implicit Association Test. *Participants who reported no awareness or no intention* are participants who responded "No" or "No (unsure)" to the question whether the primes influenced the target evaluation *or* responded "Not at all" to the question whether they intentionally rated the primes instead of the targets. *Participants who reported awareness and intention* are the rest of the sample. On the right column, the subscript "a" indicates significant difference than the parallel value in the "no-awareness or no-intention" subsample.
**$p < .05$. ***$p < .01$.

psychometric qualities of the AMP. Reporting priming was more common than in Study 1A, and even people who reported no priming at all showed some internal consistency and a moderate relationship between their AMP score and other attitude measures. Still, Table 2 shows that the AMP scores of participants who were certain that priming occurred exhibited significantly stronger split-half correlation and relationship with other measures than the AMP scores of participants who believed that priming did not occur.

Table 3 shows that we found that among people who reported never intentionally rating the primes (41% of the sample), there was only weak evidence that AMP effects were related to people's attitudes: a weak split-half correlation (r = .24) and weak relationships between the AMP and other attitude measures (all r < .20), contrasted with strong correlations for those reporting some intention to rate the primes. And, again, the weak effects for people reporting no intention to rate the primes could not be explained by them having no or weak attitudes. The correlation between the IAT and self-reported attitudes remained strong for that group (r = .61).

As Table 4 demonstrates, the psychometric qualities of the AMP were overall acceptable, but mostly because of the contribution of participants who reported that they caused or at least noticed the priming. Without these participants, the psychometric qualities of the AMP were low compared with the expected qualities when measuring political attitudes.

Study 2

Studies 1A and 1B showed that priming effects reflected attitudes with acceptable reliability and validity only when participants reported that they sometimes intentionally rated the primes instead of the targets. Study 2 was designed to examine whether these results would replicate when the measured attitudes were created experimentally. In addition, Study 2 manipulated the extremity of the difference between two fictional characters. This provided another measure of the sensitivity of the AMP to attitudes. In particular, we tested whether the AMP's sensitivity to the evaluative difference between the primes would be moderated by participants' reports of priming and of intentional rating of the primes.

Method

Participants. A total of 561 participants (57% women, 42% men, 1% unknown; *M* age = 31.2, *SD* = 11.5) completed the study. We removed participants who always responded unpleasant (1) or pleasant (22), or reported reading Chinese (5), leaving 533 (95%) participants.

Procedure and materials. The experiment always started with the induction task, followed by the AMP and self-reported attitude measures (presented in a random order). The AMP experience questionnaire always followed the AMP. This study did not include the IAT for simplification because we did not expect any discrepancy between implicit and explicit attitudes in the present context.

Attitude induction. Participants read 36 descriptions of behaviors each attributed to male individuals named Reemolap and Vabbenif (adapted from Ranganath & Nosek, 2008, and Gregg, Seibt, & Banaji, 2006). In all, 18 of the behaviors were positive (e.g., Reemolap helped an elderly man who dropped some packages) and 18 were negative (e.g., Vabbenif insulted his assistant at work). Each behavior was presented on the screen for 5 s with a drawing of the individual.

In the *extreme-difference* condition, one individual performed 16 positive and 2 negative behaviors, and the other individual performed 16 negative and 2 positive behaviors. In the *moderate-difference* condition, the ratio of behaviors was 12:6. In the *no-difference* condition, the ratio was 9:9 for both individuals. The order of the presentation of the behaviors was selected randomly for each participant, but the actual behaviors were not selected randomly. For example, in the 12:6 condition, the 12 positive and 6 negative behaviors of the positively portrayed individual were always the same behaviors. This was also true in the 9:9 condition: The same 9 positive and 9 negative behaviors were used for one individual and another 18 behaviors were used for the other individual (there were two 9:9 conditions, different in the assignment of each of the individuals to each set of 18 specific behaviors). The individual behaviors were not equated for extremity, which may result in some preference for one individual over the other even in the 9:9 condition.

Measures. The AMP followed the same procedure as in Study 1, but with photos of Reemolap and Vabbenif instead of White and Black people, and without the 100-ms blank screen between the prime and the target in the AMP (the target followed the prime immediately). The questionnaire that followed the AMP was the same as in Study 1. In the self-report questionnaire, participants rated their feelings toward each individual on a thermometer item.

Design. The overall design included 12 between-participant conditions: 3 (positive–negative behaviors ratio: 16:2, 12:9, and 9:9) × 2 (individual/behaviors match: Reemolap and Vabbenif switched behaviors) × 2 (measure order: explicit first, explicit second).

Results and Discussion

The study used two AMP scores to measure its sensitivity to the induction. One was the size of the priming effect, computed in the same way as in Study 1. The other was the preference for the positively portrayed character, computed as the difference in pleasant rating between the two characters.

Manipulation check. As expected, the self-reported preference extremity (absolute explicit preference) was larger for the more extreme induction conditions (Table 5). However, the variation across conditions was not gradual: The difference between the no-difference and moderate-difference inductions was much smaller, $t(350) = 4.06$, $d = 0.43$, than the difference between the moderate-difference and the extreme-difference inductions, $t(358) = 10.40$, $d = 1.09$.

The AMP measure was less sensitive to the extremity induction (Table 5). The average size of the priming effect in the no-difference and moderate-difference inductions were similar, and only the extreme-difference induction generated a larger priming effect than the two other induction conditions, $t(531) = 4.12$, $d = 0.36$.

Reported priming. People were most likely to report that the primes did not affect their evaluation of the targets (Table 2). The AMP could not detect the attitude induction in this subgroup. People who reported no priming showed the same priming effect in the no-difference ($M = 0.13$, $SD = 0.09$, $n = 53$) and the extreme-difference conditions ($M = 0.13$, $SD = 0.13$, $n = 54$). In addition, among people who reported no priming, the AMP did not detect any preference for the positively portrayed character, in the moderate-difference ($M = 0.004$, $SD = 0.13$, $t < 1$) and the extreme-difference ($M = 0.01$, $SD = 0.16$, $t < 1$) conditions. In fact, as detailed in Table 2, the AMP detected this preference only among participants who reported certainty that priming occurred. This suggests that the AMP measure showed no sensitivity to the attitude induction procedure unless people were certain that priming occurred.

Table 5. Study 2: Induced Attitudes and Perceived Awareness and Control

Induced preference extremity	9:9	12:6	16:2
AMP extremity	0.12_a (0.08)	0.11_a (0.10)	0.15_b (0.13)
Explicit extremity	1.27_a (1.35)	1.94_b (1.71)	3.97_c (1.98)
RPE	-0.45_a (0.89)	-0.48_a (0.96)	-0.21_b (1.1)
Reported intentional primes rating	-1.60_a (0.68)	-1.40_b (1.04)	-1.11_c (1.20)

Note: AMP = affective misattribution procedure; RPE = reported priming effect. A preference extremity of 12:6 indicates that the positively portrayed individual performed 12 positive behaviors and 6 negative behaviors, whereas the negatively portrayed bad individual performed 12 negative behaviors and 6 positive behaviors. Within each row, different subscripts indicate significant difference.

Replicating the previous findings, RPE was related to a larger priming effect, $r(533) = .43$. In addition, the preference for the positively portrayed character in the AMP was larger when RPE was higher, in the moderate-difference, $r(179) = .27$, and the extreme-difference conditions, $r(181) = .33$. The split-half correlation and the AMP/explicit relationship were weak and nonsignificant when participants reported no priming, and were significantly stronger when participants reported more and more certainty that priming occurred (Table 2).

Reporting intention. Most participants (62%) reported that they did not rate the primes at all, and only 4% reported always rating the primes intentionally (Table 3). Reported intention to rate the primes was positively related to the size of the priming effect, $r(533) = .52$. The AMP score of participants who reported never intentionally rating a prime instead of the target showed very weak split-half correlation and weak relationship with self-reported attitudes—both significantly smaller than the correlations among the rest of the sample.

The AMP scores of the large "no intention" subgroup showed no evidence of the attitude induction. The size of their priming effect was almost identical in the no-difference ($M = 0.13$, $SD = 0.08$) and extreme-difference conditions ($M = 0.14$, $SD = 0.11$), $t < 1$. In addition, the AMP did not detect any preference for the positively portrayed character, in the moderate-difference ($M = -0.01$, $SD = 0.13$) and the extreme-difference ($M = 0.01$, $SD = 0.16$) conditions, $|ts| < 1$. This failure to detect the induction was not because the induction did not influence the attitudes of people in that subgroup. The extremity of their self-reported attitudes was much smaller in the no-difference condition ($M = 1.3$, $SD = 1.35$) than in the extreme-difference condition ($M = 3.7$, $SD = 2.04$), $d = 1.38$. The self-reported attitudes of people who reported never intentionally rating the primes showed variance in line with the experimental induction, but the AMP did not capture this variance.

The effect of induced preference extremity. There was no difference of RPE between the no-difference and the moderate-difference induction levels (Table 5). But, when these two conditions were merged together, their RPE level was smaller than in the extreme-difference induction, $t(531) = 2.80$, $d = 0.25$. The induced difference extremity affected reported intentional primes rating in all three conditions, such that a stronger induction predicted more reports of intentional primes ratings. People who underwent the moderate-difference induction reported more intentional primes rating than people in the no-difference condition, $t(350) = 2.06$, $d = 0.23$, and less intentional primes rating than people in the extreme-difference condition, $t(358) = 2.51$, $d = 0.26$. These results suggest that experimentally inducing more extreme preference causes stronger tendency to report that the primes influenced judgment and to report intentional primes rating.

Summary. The main findings of the previous studies were replicated with experimentally manipulated attitudes. People who reported more awareness and control of the priming effect showed stronger priming effects in the AMP, and were the main contributors to its psychometric qualities. In addition, the AMP showed no sensitivity to the attitude induction procedure when participants reported that they never intentionally rated the primes. We also found that extreme preferences increase perceptions that priming occurred, as well as the tendency to report that the priming was intentional.

Study 3

In Study 2, when more extreme attitudes were induced experimentally, the priming strengthened, and so did reports that the primes were rated intentionally. One possible reason for this finding is that when asked about the priming effect, people with more extreme attitudes are more likely to suspect that their attitudes influenced the evaluation of the targets. They may even retrospectively infer that they intentionally rated the primes instead of the targets, an alternative to them actually intentionally rating the primes. Before turning to our experimental test of this possibility, we conducted correlational analyses using data from the previous studies.

If explicit attitudes were information that participants used in their inference about the AMP priming effect, then there should be a positive relationship between the extremity of people's explicit attitudes and reports about awareness and intention. The extremity of people's explicit attitudes was correlated to RPE, $r(353, 382) = .21, .22$, in Studies 1A and 1B, respectively, and with reported intentional primes rating, $r(353, 382) = .17, .27$. These correlations provide some support to the inference account, but without any evidence of causal relationships. Study 3 tested the account experimentally.

If people inferred awareness and intention retrospectively from their feelings toward the primes and not from experiencing the priming effect during the task, then a manipulation that changes the priming effect without changing explicit attitudes toward the primes should have no influence on reports of awareness and intention. Study 3 tested this possibility by manipulating the duration of the target presentation. Payne et al. (2005, Study 4) found that when the target duration is longer, the priming effect decreases. If this manipulation would moderate the priming effect but not people's reports of awareness and control, then it would support the possibility that people infer the awareness and intention from their feelings toward the primes.

Method

Participants. A total of 780 participants (71% women, 28% men, 1% unknown; M age = 26.4, SD = 12) completed the study. In all, 702 (90%) were included in the analyses, after removing participants who responded unpleasant (n = 1) or pleasant (52) in all the AMP trials, reported reading Chinese (23), and/or performed too many fast trials in the IAT (2).

Procedure and materials. The procedure and measures were identical to those used in Study 1A, with the following exceptions: We added nine White males and nine Black males to the six stimuli used in Study 1. In the *long target duration* condition, the target in the AMP was presented for 2,000 ms instead of 100 ms. To increase the sensitivity of the RPE measure, there were six response options to the question about the influence of the primes on the targets: [Yes, all of the time], [Yes, sometimes], [Yes, very few times], [I'm not sure, but I think not], [I'm almost sure that not], and [I'm sure not] (coded from 2.5 to −2.5, respectively, with 1 point difference between responses).

Another important addition in Study 3 were two questions that probed for knowledge about specific priming effects: "When black[white] faces appeared before the Chinese symbols, what was their influence on your judgment of the Chinese symbols?" The response options were as follows: "I was strongly [moderately] [slightly] more likely to judge the Chinese symbols as [un]pleasant," coded from 3 = *strongly pleasant* to −3 = *strongly unpleasant*, with 1 point difference between responses. The response option "I was not more likely to judge the Chinese symbols as pleasant or unpleasant" was coded as 0. These questions may add information about the accuracy of people's knowledge about the priming because, unlike the prior questions about the priming effect, they refer specifically to the prime categories and the specific priming effect (a positive or negative bias) and not to the prime stimuli and the priming effect in general.

Results and Discussion

The means and the intercorrelations (Table 1) were generally similar to those observed in Study 1A.

The effect of target duration. Replicating previous research, the priming effect was stronger (M = 0.18, SD = 0.17) when the target duration was short than when it was long (M = 0.14, SD = 0.11), t(646, unequal variance) = 3.92, d = 0.29. The main question of this study was whether this manipulation would also affect people's perception of the priming effect. We found that people in the *short-duration* condition were more likely than people in the *long-duration* condition to report that the primes influenced their judgment of the targets, t(700) = 4.17, d = 0.32. Similarly, people reported more intentional primes rating in the short-duration condition (M = −0.86, SD = 1.2) than in the long-duration condition (M = −1.34, SD = 0.91), t(675, unequal variance) = 5.85, d = 0.44. These results indicate that people's perceptions were sensitive to the manipulation of the target duration and, thus, that the perceptions are based on the experience during the task and not on the extremity of their attitudes.

Perceptions of the priming. The priming effect was related to RPE, r(702) = .21, and to reported intentional primes rating, r(702) = .27. The RPE results also replicated previous studies. The AMP scores of people who responded with one of the three negative responses to the RPE question (54% of the sample) showed significantly weaker split-half correlation (r = .28), and correlation with self-report (r = .18) and the IAT (r = .13), than the rest of the sample (parallel rs = .57, .41, and .37). People who reported no intentional primes rating (47% of the sample) showed very weak split-half correlation and very weak correlation with other attitude measures (Table 3). Again, these relationships were significantly stronger among participants who reported more frequent intentional rating of the primes. In the short-duration condition, the overall AMP's psychometric qualities were much worse without participants who reported priming or intentional rating of the primes (Table 4). This decline was not as drastic in the long-duration condition, probably because the overall psychometric qualities of the AMP in that condition were weaker to begin with.

Specific perceptions of the priming. On average, participants believed that White primes increased the likelihood that the evaluation of the targets would be positive (M = 0.15, SD = 0.92), t(701) = 4.3, d = 0.16, whereas the average presumed effect of Black primes (M = 0.04, SD = 1.00) was not different than zero, t(702) = 1.06, p = .29. The estimate regarding the influence of Black primes correlated with the actual rate of pleasant responses after Black primes, r(702) = .41, significantly more than it correlated with the rate of pleasant responses after White primes, r(702) = .05, z (for the comparison) = 7.48, and also more than it correlated with self-reported attitudes toward Black people, r(702) = .22, z = 4.31. Similarly, the estimate regarding the influence of White primes correlated with the rate of pleasant responses after White primes, $r(702) = .30$, more than it correlated with the rate of pleasant responses after Black primes, $r(702) = -.01$, $z = 6.60$, and also more than it correlated with self-reported attitudes toward Black people, $r(702) = .10$, $z = 4.10$. These results suggest that people had good knowledge of the specific priming effect that was caused by each prime category and that this knowledge was not based on their self-reported attitudes.

General Discussion

Four studies investigated people's perception of the impact of attitudes toward the primes on their evaluation of the targets in the AMP. We found that (a) even when people report that the primes did not influence their evaluation of the targets, priming occurs; (b) when the priming effect is strong, with good internal consistency and good relationship to other attitude measures, people tend to accurately report that the effect has happened and even think that they intentionally caused the priming; and (c) people's perception of the priming effect is sensitive to changes in that effect that are caused by manipulating the extremity of the attitudes (Study 2) and by manipulating the duration of the target's presentation (Study 3).

Priming Without Knowledge

Perhaps the most promising aspect of the AMP as a measure of automatic evaluation is that attitudes affect the performance despite instructions to avoid this influence. Assuming that participants follow the instructions, the priming effect must be unintentional and perhaps escape people's awareness. In support of that assumption, even participants who reported no awareness or control of the priming effect still evaluated targets differently when they followed different prime categories. The AMP score showed significant preference for Black over White people (Studies 1A and 3) and for Democrats over Republicans (Study 1B) among participants who reported that no priming occurred.

Although retrospective reports have been a common tool in studies of awareness (e.g., Nisbett & Wilson, 1977; automatic social behavior, Bargh & Ferguson, 2000; evaluative conditioning, De Houwer, 2007), we consider those reports as a measure that is affected by awareness, but not exclusively or exhaustively (Reingold & Merikle, 1988). Participants might have forgotten that they noticed the priming or participants might have chosen to hide that they did not follow the instructions. Therefore, this finding supports the possibility that the priming sometimes occurred without awareness, but it is not definitive.

Priming With Knowledge

Even if the reported unawareness is the result of true unawareness of the priming effect, our studies suggest that this ignorance does not characterize the participants when the AMP reflects attitudes reliably. We found that in addition to being the largest group in all studies, participants who reported no priming were poor contributors to the AMP's psychometric qualities as an attitude measure. The smaller subsets of participants—those who reported that the priming occurred and that they occasionally caused the priming intentionally—were the main contributors of the AMP's good internal consistency and relationship with other attitude measures. We therefore conclude that when attitudes have a reliable effect on performance in the AMP, this effect usually does not escape people's knowledge.

One may contend that when people reported no awareness or intention, their performance in the AMP was influenced mostly by their implicit attitudes, whereas when people reported awareness and intention, their performance in the AMP was influenced mostly by their explicit attitudes. Alternatively, perhaps participants who were motivated to seem unbiased (especially in Studies 1A and 3 that measured socially sensitive attitudes) were more likely to report that they were unaffected by the primes and also to alter their self-reported attitudes away from their actual attitudes that were still captured by the AMP. These two alternatives could surely explain why the relationship between the AMP and self-reported attitudes was almost nonexistent when people reported no priming, and much stronger when they reported awareness and intention. However, relationship with explicit attitudes was not the only psychometric criteria in our studies. In Studies 1 and 3, the AMP score of people who reported no awareness and no intention showed weaker relationship with the IAT, although the relationship between the IAT and self-reported attitudes was unaffected. In addition, in Study 2, there was little reason to assume any discrepancy between explicit and implicit attitudes because the attitude objects were unknown individuals who clearly behaved immorally or morally. In addition, in Study 2, the attitudes were manipulated experimentally, adding the measurement criterion of sensitivity of the AMP to evaluative information about the attitude objects. We found sensitivity to the attitude induction only when participants reported that priming occurred, and only when people reported that they usually intentionally rated the primes instead of the targets. Finally, in all studies, the internal consistency of the AMP was strongly related to reported priming and intentions. As such, we believe that the most parsimonious explanation is that the AMP is simply less effective at assessing attitudes among those people who report no awareness or intention for rating the primes.

Because we did not collect data about participants' perception of the attitude effect in the IAT, we cannot directly compare the awareness of the attitude effect in the AMP with the awareness of the attitude effect in the IAT. Monteith et al. (2001) found a correlation of $r(78) = .39$ between people's retrospective reports on the direction of the attitude effect in a Black–White IAT and the actual effect. This retrospective report is similar to the questions asked in Study 3 about the specific priming effects after each of the two groups of primes. For each social group (Black and White faces), participants estimated whether primes of that group led to more pleasant or more unpleasant rating of the targets. We computed a difference score from participants' estimation of their tendency to rate the targets as positive after White faces versus Black faces. That difference score correlated with the actual preference score in the AMP, $r(702) = .50$. Future research should compare people's awareness of the attitude effect in the IAT and the AMP in the same study.

The Source of Knowledge

Inference. One possible reason for our failure to find evidence for a strong and reliable attitude effect in the AMP without people's knowledge is that people inferred this effect retrospectively, after it had occurred unintentionally and nonconsciously during task performance. What information might have such inference used? One possible source of information is people's awareness of the strength of their own attitudes. People with weak attitudes toward the primes could easily infer that priming did not occur in the AMP and would also show much weaker signs that priming has happened (i.e., weak psychometric qualities). In support of that assumption, our studies found weak but significant relationships between the extremity of people's explicit attitudes and reports that priming has occurred. However, in Studies 1A, 1B, and 3, even participants who reported no priming effect and showed very small AMP/IAT and AMP/self-report relationships still showed IAT/self-report correlations consistent with the whole sample. This suggests that the attitudes of people who reported no priming in the AMP had enough variability to show reasonable psychometric qualities, and only the AMP showed little sensitivity to that variability. Similarly, in Study 2, participants who reported no priming and showed hardly any indication of the priming in their AMP scores still reported explicit attitudes that were sensitive to the experimental attitude induction.

There is more evidence against the possibility that people inferred the priming from the extremity of their explicit attitudes. In Study 3, the manipulation of the target exposure duration influenced the priming effects without influencing people's attitudes (self-reported or measured with the IAT), and yet it also influenced people's reports whether priming has occurred. This could not have occurred had people's reports of the priming effect relied only on their attitudes. In addition, when participants specifically reported how each prime category influenced their performance (Study 3), these reports were correlated significantly better with the actual priming effect of that category in the AMP than with the extremity of the explicit attitudes toward that category. All that evidence suggests that people's reports whether the priming has occurred were not based merely on whether they had any attitudes toward the attitude objects. Also, people's reports were not based solely on their knowledge of the extremity of their attitudes.

It follows that if people inferred the priming effect after it had already happened, this was probably based on them noticing the priming during the task, maybe immediately after it happened. For instance, perhaps people noticed that their evaluation of the targets was often compatible with their attitude toward the primes. This account fits the main finding of Study 3: Prolonging the target's exposure decreased the priming effect and also decreased people's tendency to report that priming occurred. This result supports the assumption that people based their judgment whether priming occurred on the actual experience during the task. Perhaps in the long-duration condition in Study 3, participants were less likely to notice compatibility between their evaluation of the target and their evaluation of the primes, and therefore they were (accurately) less likely to infer the priming occurred.

For an inference account to survive, it should also explain how retrospectively inferring that the priming occurred caused people to believe that they sometimes rated the primes intentionally. One possible explanation is that when people link their internal states (e.g., attitudes toward the primes) to an effect (e.g., evaluation of the targets), they are more likely to attribute the effect to their agency (Wegner, 2005) and believe that they caused it (see Kühn & Brass, 2009, for a convincing demonstration of illusory perception of intention after unintentional behavior).

Was the priming intentional? In our view, the most surprising and interesting findings of this research are summarized in Tables 3 and 4. The AMP's good psychometric qualities rely on the minority of the participants who reported that, at least sometimes, they intentionally rated the primes. The second column of Table 3 shows that almost half of the participants in each study reported never intentionally rating the primes. And, these participants showed very little evidence that attitudes influenced their performance. The first column of Table 4—presenting the psychometric qualities of the AMP in the whole sample—suggests that the acceptable psychometric qualities of the AMP were contributed mainly by the rest of the sample that reported at least some intentional rating of the primes.

This brings us to the final possible reason for our findings: people cause the priming effect intentionally. For instance, perhaps the task of rating unfamiliar Chinese characters that flash quickly on the screen is annoying and frustrating, so when a person does not see it clearly, they rate the prime instead. Although participants reported intentionally rating the primes, these reports cannot be accepted uncritically. People's accounts may sometimes fail to provide the accurate explanation of their behavior (Nisbett & Wilson, 1977). In addition, previous research has found evidence to suggest that the attitude effect in the AMP is not intentional. In one study, the motivation to hide one's preference between White and Black people was manipulated between participants (Payne, Burkley, & Stokes, 2008, Study 4; see Payne, Govorun, & Arbuckle, 2008, Study 4, for a conceptual replication). The correlation between the AMP and a direct rating of the primes was lower when participants were motivated to hide their preference than when they were motivated to tell the truth. This could suggest that the manipulation influenced people's direct ratings but did not influence the AMP. If that was the case, then this would support the assumption that the AMP was influenced by unintentional priming effect. However, an alternative account is that people with high motivation to conceal their preferences were less likely to intentionally rate the primes in the AMP. Therefore, their AMP scores were less likely to reflect their attitudes and had lower relationship with the direct ratings.

More evidence to support the possibility that the AMP reflects unintentional attitude effects comes from a study that found an effect of subliminal conditioning on the AMP, but not on self-reported attitudes (Rydell et al., 2008). In that study, people's self-reported attitudes were influenced by explicit information about a novel attitude object and insensitive to subliminal conditioning of the attitude object with affective stimuli; in contrast, the AMP score was sensitive to subliminal conditioning and insensitive to explicit evaluative information about the attitude object. In summary, some previous findings provide evidence that, at least sometimes, the AMP is sensitive to unintentional attitude effects. Therefore, although the possibility that people's reported intention was an illusory perception is somewhat puzzling, we think that the present evidence increases the plausibility that the AMP's properties are mostly a consequence of intentional ratings of the primes, but is not definitive.

Investigations of intentionality in other implicit measures—the IAT and evaluative priming (EP; Fazio et al., 1995)—showed that, under some conditions, people can control the measurement outcome (for IAT: Kim, 2003; Steffens, 2004; for EP: Degner, 2008; Teige-Mocigemba & Klauer, 2008). Similarly, it is obvious, that if instructed to do so, people can control the measurement outcome of the AMP. It is not surprising that people who were directly instructed to rate the primes instead of the targets were successful in following these instructions (Payne, Burkley, et al., 2008). However, the fact that a measurement outcome *can be* controlled does not inform us whether it is controlled in ordinary measurement conditions. The present research found that a reliable attitude influence in the AMP is usually accompanied by a perception of intentionally causing this effect. It would be very interesting if this finding is the result of an illusory perception of intentionally causing an unintentional effect. However, because the AMP is gaining much popularity as a measurement tool in the field of attitude research, the alternative possibility—that it is heavily influenced by intentional primes rating—cannot be ignored. Further research might clarify whether perception of intentional primes rating stems from an actual or a retrospectively constructed intention.

Conclusion

The present research found that when the attitude toward the primes influences the evaluation of the targets in the AMP, people tend to report that this attitude effect occurred. It seems that like in the IAT (Monteith et al., 2001), people are often aware of the attitude effect in the AMP. Furthermore, for the AMP, without people who report that the attitude effect occurred, the psychometric qualities were very weak. This suggests that awareness is essential for the AMP's validity.

In addition, we found that retrospective reports about occasional intentional rating of the primes instead of the targets during the task were essential for good reliability and validity of the AMP score. This is alarming because if these reports are accurate, then the AMP's psychometric qualities may rely heavily on intentional evaluation. Even if most of the intentionality reports only reflect illusory intention, they help to identify a sizable subset of the participants—those who reported no intentional primes rating at all—whose automatic evaluations are hardly measured by the AMP. As Table 3 details, this "AMP insensitive" population was 41% to 62% of the participants in our studies. Therefore, even the most cautious conclusion from the present results is that the AMP's psychometric qualities are highly dependent on only a subset of the participants and that many people are unaffected by the procedure. Further research should investigate whether this is a unique feature of the AMP or a common feature shared with other measures of automatic evaluation.

Acknowledgment

The authors thank Keith Payne for extensive feedback and ideas about this research.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Nosek is an officer of Project Implicit, Inc., a non-profit organization that includes in its mission "To develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender or other factors.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by grants from the National Institute of Mental Health (R01 MH68447) to Brian Nosek, and by grants from the European Union (PIRG06-GA-2009-256467) to Yoav Bar-Anan.

References

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology, 87*, 23-37.

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology, 63*, 2313-2335.

Bar-Anan, Y., & Nosek, B. A. (2012). *Multi-method investigation of implicit measures*. Unpublished manuscript.

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

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

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

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology, 43*, 437-449.

Berridge, K. C., & Winkielman, P. (2003). What is an unconscious emotion: The case for unconscious "liking." *Cognition & Emotion, 17*, 181-211.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology, 10*, 230-241.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869.

Degner, J. (2008). On the (un)controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition & Emotion, 28*, 327-354.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 13*, 513-519.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

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.

Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition, 9*, 13-36.

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

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology, 33*, 573-589.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition, 15*, 485-499.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin, 34*, 1276-1289.

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning. *Journal of Experimental Social Psychology, 48*, 617-623.

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.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Race attitude measures predicted vote in the 2008 U.S. presidential election. *Analysis of Social Issues and Public Policy, 9*, 241-253.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier made than undone: The asymmetric malleability of automatic preferences. *Journal of Personality and Social Psychology, 90*, 1-20.

Hassin, R. R., Bargh, J. A., & Zimerman, S. (2009). Automatic and flexible: The case of non-conscious goal pursuit. *Social Cognition, 27*, 20-36.

Hofmann, W., Friese, M., & Roefs, A. (2009). Three ways to resist temptation: The independent contributions of executive attention, inhibitory control, and affect regulation on the impulse control of eating behavior. *Journal of Experimental Social Psychology, 45*, 431-435.

Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly, 66*, 83-96.

Kühn, S., & Brass, M. (2009). Retrospective reconstruction of intentionality. *Consciousness and Cognition, 18*, 12-21.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 3-26. doi:10.1037//0097-7403.28.1.3

Miller, K. P., Brewer, M. B., & Arbuckle, N. L. (2009). Social identity complexity: Its correlates and antecedents. *Group Processes & Intergroup Relations, 12*, 79-94.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455-469.

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

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial bias. *Social Cognition, 19*, 395-417.

Mueller, C. M., & Dweck, C. S. (1998). Intelligence praise can undermine motivation and performance. *Journal of Personality and Social Psychology, 75*, 33-52.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565-584.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265-292). New York: Psychology Press.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15*, 152-159.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-88.

Oikawa, M., Aarts, H., & Oikawa, H. (2011). There is a fire burning in my heart: The role of causal attribution in affect transfer. *Cognition & Emotion, 25*, 156-163.

Olson, M., & Fazio, R. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 5*, 413-417.

Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16-31.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.

Payne, B. K., Govorun, O., & Arbuckle, N. L. (2008). Automatic attitudes and alcohol: Does implicit liking predict drinking? *Cognition & Emotion, 22*, 238-271.

Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: The role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology, 91*, 218-231.

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately, explicit attitude generalization takes time. *Psychological Science, 19*, 249-254.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 34*, 65-71.

Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology, 44*, 1526-1532.

Shariff, A. F., & Tracy, J. L. (2009). Knowing who's boss: Implicit perceptions of status from the nonverbal expression of pride. *Emotion, 9*, 631-639.

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 286-291.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*, 165-179.

Teige-Mocigemba, S., & Klauer, K. C. (2008). "Automatic" evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology, 44*, 1414-1417.

Uhlmann, E., Pizarro, D. A., & Bloom, P. (2008). Varieties of unconscious social cognition. *American Psychologist, 52*, 613-629.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19*, 49-54.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82*, 919-934.

Wegner, D. M. (2004). Precis of the illusion of conscious will. *Behavioral and Brain Sciences, 27*, 649-692.

Wegner, D. M. (2005). Who is the controller of controlled processes? In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 19-36). New York, NY: Oxford University Press.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*, 481-492.

Wittenbrink, B., & Schwarz, N. (2007). *Implicit measures of attitudes*. New York, NY: Guilford.