

Personality and Social Psychology Bulletin

<http://psp.sagepub.com/>

Intention Invention and the Affect Misattribution Procedure : Reply to Bar-Anan and Nosek (2012)

B. Keith Payne, Jazmin Brown-Iannuzzi, Melissa Burkley, Nathan L. Arbuckle, Erin Cooley, C. Daryl Cameron and Kristjen B. Lundberg

Pers Soc Psychol Bull 2013 39: 375

DOI: 10.1177/0146167212475225

The online version of this article can be found at:

<http://psp.sagepub.com/content/39/3/375>

Published by:



<http://www.sagepublications.com>

On behalf of:



Society for Personality and Social Psychology

Additional services and information for *Personality and Social Psychology Bulletin* can be found at:

Email Alerts: <http://psp.sagepub.com/cgi/alerts>

Subscriptions: <http://psp.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Feb 11, 2013

[What is This?](#)

Intention Invention and the Affect Misattribution Procedure: Reply to Bar-Anan and Nosek (2012)

Personality and Social
Psychology Bulletin
39(3) 375–386
© 2013 by the Society for Personality
and Social Psychology, Inc
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146167212475225
pspb.sagepub.com


B. Keith Payne¹, Jazmin Brown-Iannuzzi¹, Melissa Burkley², Nathan L. Arbuckle³, Erin Cooley¹, C. Daryl Cameron¹, and Kristjen B. Lundberg¹

Abstract

A recent study of the affect misattribution procedure (AMP) found that participants who retrospectively reported that they intentionally rated the primes showed larger effect sizes and higher reliability. The study concluded that the AMP's validity depends on intentionally rating the primes. We evaluated this conclusion in three experiments. First, larger effect sizes and higher reliability were associated with (incoherent) retrospective reports of both (a) intentionally rating the primes and (b) being unintentionally influenced by the primes. A second experiment manipulated intentions to rate the primes versus targets and found that this manipulation produced systematically different effects. Experiment 3 found that giving participants an option to "pass" when they felt they were influenced by primes did not reduce priming. Experimental manipulations, rather than retrospective self-reports, suggested that participants make post hoc confabulations to explain their responses. There was no evidence that validity in the AMP depends on intentionally rating primes.

Keywords

affect misattribution procedure, implicit social cognition, implicit attitudes

Received August 23, 2012; revision accepted November 4, 2012

The affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) has been used increasingly to study implicit social cognition. The procedure works by presenting a series of prime images or words, each followed by an ambiguous target (such as a Chinese pictograph) in rapid succession. Participants are asked to rate the target items for pleasantness and are warned not to let the primes influence their ratings of the targets. Ratings of the targets nonetheless tend to be biased by the pleasantness of the primes. This unintentional biasing influence can be used as a measure of automatically activated attitudes toward the primes.

The AMP's growing popularity may be traced in part to its strong psychometric properties. Reliability is generally high, and the measure shows good predictive validity across a variety of domains. A meta-analysis of priming tasks (Cameron, Brown-Iannuzzi, & Payne, 2012) found that the average association between the AMP and behavioral measures was $r = .35$, an effect that was larger than other priming tasks and larger than the meta-analytic association between the Implicit Association Test (IAT) and behavioral measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). The AMP also showed good specificity in the meta-analysis. As predicted by dual process theories of attitudes, the correlation between the AMP

and explicit attitude measures was much stronger under conditions that discouraged deliberately controlled responding on explicit measures ($r = .36$) than under conditions that encouraged deliberate responding ($r = -.003$).

Recently, Bar-Anan and Nosek (2012) questioned the validity of the AMP based on a series of studies in which participants completed an AMP and were then asked whether they intentionally rated the primes. They found that the small percentage of participants who reported that they intentionally rated the primes showed larger effect sizes, higher reliability, and stronger associations with other attitude measures. The authors acknowledged that retrospective self-reports have many flaws as a measure of awareness or intentions. Nonetheless, based on these correlational data, Bar-Anan and Nosek concluded,

¹University of North Carolina at Chapel Hill, USA

²Oklahoma State University, Stillwater, USA

³Mind Research Network, Albuquerque, NM, USA

Corresponding Author:

Keith Payne, Department of Psychology, University of North Carolina at Chapel Hill, Campus Box 3270, Chapel Hill, NC 27599, USA
Email: Payne@unc.edu

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. (p. 13)

In addition, they found that retrospective reports about how much participants thought the primes influenced their ratings were associated with the AMP effect size and reliability. The authors concluded from this that "awareness is essential for the AMP's validity" (p. 13). Despite the correlational nature of these data, the conclusions drawn implied causality. If validity "relies heavily on" intent, and awareness is "essential" for validity, then removing intent and awareness would presumably decrease validity.

If Bar-Anan and Nosek are right that the strong psychometrics of the AMP are driven by participants who intentionally rate the primes, then this would indeed be an important limitation of the AMP as an implicit measure. But are they right? In this article, we report three experiments showing that the retrospective self-reports on which Bar-Anan and Nosek based their conclusions are not accurate reports about the causes of participants' responses. They are post hoc confabulations in which participants attempt to make sense of their own behavior. Because these confabulations result from—rather than cause—responses in the AMP, they pose no threat to the validity of the AMP as an implicit measure.

The Trouble With Retrospection

It is by now well understood that retrospective self-reports are not a reliable guide to the causes of behavior. People are influenced by factors that they do not recognize, and they report being influenced by factors that had no effect (Nisbett & Wilson, 1977). Their reports are not random, however. Nisbett and Wilson (1977) showed that subjects' self-reports reflect the same intuitive theories about their own behaviors that third-person observers use to explain behavior. Self-reports may therefore have above-chance accuracy while still revealing nothing about introspective insight, for the same reason that an impartial observer may accurately guess the causes of someone else's behavior (Bem, 1967).

If reporting on the cause of behavior is difficult in general, then reporting about whether conscious intentions caused the behavior is especially problematic. As Wegner and Wheatley (1999) argued,

the degree of conscious will that is experienced for an action is not a direct indication of any causal link between mind and action. Rather, our analysis suggests that conscious will results from a causal illusion that is the psychological equivalent of the third-variable problem in causal analysis. (p. 482)

People sometimes cause actions that they do not experience as willed action (automatisms, such as playing with an Ouija board, dowsing, or automatic writing; Wegner, 2002; Wegner, Fuller, & Sparrow, 2003). Other times, they report that they caused actions that were in fact caused by others, or that happened randomly (illusions of control; Langer & Roth, 1975; Pronin, Wegner, McCarthy, & Rodriguez, 2006; Wegner & Wheatley, 1999).

Judgments of intentional action are especially likely to be misleading when conditions are ambiguous or confusing (Kuhn & Brass, 2009; see also Aarts, Custers, & Wegner, 2005). One way to produce predictable confabulations is to subtly influence participants' behavior with priming manipulations (Oettingen, Grant, Smith, Skinner, & Gollwitzer, 2006; Parks-Stamm, Oettingen, & Gollwitzer, 2010). Perhaps the clearest example of confabulated motives following priming comes from a series of studies by Bar-Anan, Wilson, and Hassin (2010). In one study, men were primed with the goal to affiliate with women by reading about a man and woman who meet, dance together, and then go home together. Next they were given the choice between studying with a male or female tutor who each taught different subjects. Men primed to affiliate with women were more likely to choose whatever topic was taught by the woman. In retrospective reports, however, they attributed their choice to their intrinsic interest in the topic. Goals that were set in motion by priming (affiliating with women) were misattributed to unrelated personal motives (academic interests). Completing a sequential priming task such as the AMP is similar to the situation of participants who have been primed in other ways. The fast succession of primes and targets makes it difficult to know how to attribute the causes of each rating. As a result, participants are likely to latch onto whatever salient factor might plausibly explain their behavior.

Overview of the Present Research

In three experiments, we tested whether the validity of the AMP in fact depends on intent and awareness. In Experiment 1, we first replicated Bar-Anan and Nosek's finding that retrospective self-reports of intentionally rating the primes were associated with larger effect sizes and higher reliability in the AMP. We added a new condition, however, in which participants answered a different retrospective question. We asked participants whether the primes *unintentionally influenced* their ratings of the target symbols. The confabulation account and Bar-Anan and Nosek's veridical self-report account make opposing predictions for this question.

If responses to the question are a veridical report about the causal role of intent as Bar-Anan and Nosek assume, then "intentionally rating" the primes and being "unintentionally influenced" by them are conceptually near-opposites. That is, if intentional ratings of the primes caused large priming effects, then participants with the largest AMP effects

should agree that they intentionally rated the primes and disagree that they were unintentionally influenced.

By the confabulation account, however, responses to these two questions are expected to follow similar patterns. If reported intentions serve to explain or justify behavior after the fact, then respondents with the largest priming effects are likely to endorse either explanation. When asked whether they intentionally rated the primes, participants with large priming effects are likely to agree that they did. If, instead, they are asked whether the primes unintentionally influenced their responses, they are also likely to agree that they did. Although they are logically inconsistent, either intentionally rating the primes or being unintentionally influenced provides a plausible explanation for why a participant responded in prime-congruent ways. The confabulation account therefore predicts that larger and more reliable priming effects should be associated with both (a) reports of intentionally rating the primes and (b) reports of being unintentionally influenced by the primes.

In Experiment 2, we manipulated intent by asking participants to complete two versions of the AMP. One version was the standard AMP, in which participants were warned not to let the primes influence their judgments of the target symbols. The second version reversed this task and asked participants to ignore the symbols and rate the primes. Under Bar-Anan and Nosek's account, these tasks should be redundant because the systematic variance in the AMP is said to result from intentional ratings of the primes. We show that they are not redundant, and the standard AMP is the better predictor of racial bias when forming an impression of a new person.

In the third experiment, we address the question of awareness in the AMP. The confabulation account presupposes some level of accuracy in perceiving one's own behavior, because participants need to know what behavior needs explaining. Participants apparently notice with some degree of accuracy whether they have responded in prime-congruent ways. But is this kind of awareness relevant to the AMP's effectiveness as an implicit measure? In the original publication of the AMP, Payne et al. (2005) speculated,

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 . . . In short, we suggest that the misattribution effect was difficult to control because participants did not believe they were experiencing it. (p. 291)

Bar-Anan and Nosek quoted that passage in arguing that contrary to prior theorizing, participants appear aware of the priming effect based on retrospective self-report. The quoted statement, however, was not about retrospective judgments. It was about awareness on any given trial. That is, only real-time awareness that participants are being influenced as they

form an evaluation of the target would provide a basis for avoiding that influence.

To evaluate whether such real-time awareness exists, in this experiment, we gave participants a simple means to avoid the influence of primes based on awareness. One group completed the standard AMP, in which a pleasant or unpleasant rating was made on each trial. A second group had three response options. They were told only to respond pleasant or unpleasant to the qualities of the target symbol itself. If they believed that the prime was influencing their evaluation of a target, they were instructed to "pass" by pressing the space bar on that trial. This approach is critically different from retrospective report methodology because it is prospective; it gives participants the ability to modify their behavior based on their conscious experience before they respond. Awareness should eliminate the priming effect in the condition where passing is available.

Experiment 1

We tested whether reports of intentionally rating primes could be trusted by asking two versions of the retrospective intention question in a between-subjects design. The first version was taken from Bar-Anan and Nosek (2012) and asked whether participants intentionally rated the primes. The second version was designed to be opposite in meaning, and it asked whether participants' ratings of the symbols were unintentionally influenced by the primes. If self-reports are accurate about the causes of ratings, then these questions should show opposite effects. Those individuals who intentionally rated the primes should report that they did so on Bar-Anan and Nosek's original question, but should deny being unintentionally influenced by the primes. Therefore, large and reliable priming effects should be positively associated with the original question (replicating Bar-Anan & Nosek, 2012) and negatively associated with the reverse-worded item. In contrast, if reports of intentionally rating the primes are confabulations constructed to explain respondents' behavior during the AMP, then the two questions should show similar effects because both provide plausible explanations. Thus, large and reliable priming effects should be positively associated with reports of both intentionally rating the primes and being unintentionally influenced by them.

Method

Participants. Participants were recruited through Amazon Mechanical Turk and completed the study online for payment ($n = 313$). Participants were removed if they reported that they could read the Chinese symbols ($n = 24$) or if they completed the study twice ($n = 1$). The final sample included 288 participants (69% male) with a mean age of 26. The sample included 85% White, 4% Black, and 11% Other races.

Procedure. Participants first completed an AMP in which Black and White faces served as primes. The prime photos were matched for attractiveness and rated as equally prototypical of their respective racial groups (see Payne et al., 2005). Each of 48 trials briefly presented a photograph of the face of a White or Black man, followed by a Chinese pictograph. Each trial began with a fixation point, followed by a face presented for 75 ms, followed next by a blank screen for 125 ms, and then a pictograph for 100 ms. A black-and-white pattern mask then appeared until a response was registered. Respondents were instructed to judge whether each pictograph was pleasant or unpleasant by pressing one of two keys while avoiding influence from the photos.

Following the AMP, participants were randomly assigned to answer one of two questions. One question was the same as the question in Bar-Anan and Nosek (2012): “Did you intentionally rate the face pictures instead of the Chinese Symbols?”¹ The other question was “Did the face pictures unintentionally influence your ratings of the Chinese Symbols?” In addition, all participants were asked to estimate how much their responses were influenced by the primes (perceived priming) using the same item as in Bar-Anan and Nosek: “Do you think your ratings of the Chinese Symbols were influenced by the face pictures that appeared before the Chinese Symbols?” The order of the perceived priming question and intent questions was randomized and had no effect on results. Responses for all questions were made on a 5-point scale (*not at all; a little; sometimes; most of the time; almost always*). Following these questions, participants completed an explicit feeling thermometer in which they reported their feelings toward Blacks and Whites. Finally, they completed demographic questions and were debriefed.

Results

Reports of intentionally rating the prime were rare. The distribution of responses included, “not at all” (75.8%), “a little” (11.8%), “sometimes” (4.6%), “most of the time” (5.9%), and “almost always” (2%). These proportions are similar to those reported in Bar-Anan and Nosek (2012). Reports of being unintentionally influenced were also somewhat rare, with most respondents claiming “not at all,” (43.7%), followed by “a little” (30.4%), “sometimes” (23%), “most of the time” (2.2%), and “almost always” (0.7%).

Effect Size. The proportion of pleasant responses was significantly higher on White prime trials ($M = 0.58$, $SD = 0.16$) than Black prime trials ($M = 0.53$, $SD = 0.16$), $t = 4.07$, $p < .001$. Priming effects were computed by subtracting the proportion of pleasant responses on Black prime trials from pleasant responses on White prime trials. The effect size was then computed by taking the absolute value of this difference score (consistent with the method used in Bar-Anan & Nosek, 2012).

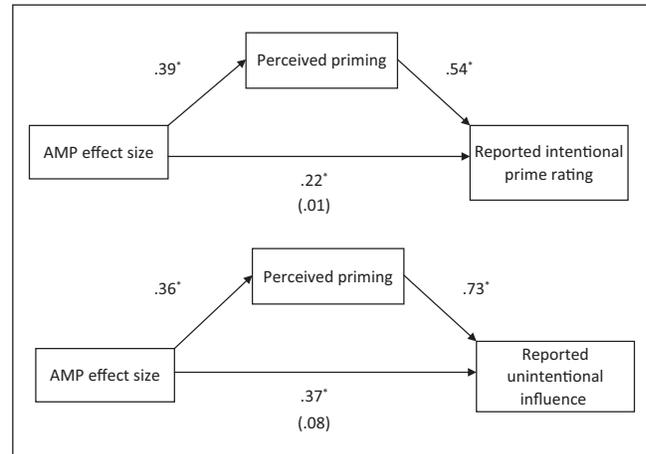


Figure 1. The associations between AMP effect size and reports of intentionally rating primes (top) and being unintentionally influenced by primes (bottom) were fully mediated by perceived priming

Note: AMP = affect misattribution procedure. Coefficients are unstandardized regression weights based on standardized variables.

* $p < .01$.

We tested the main hypothesis by conducting a regression analysis in which we predicted responses to the self-report question from the AMP effect size, the question wording (intentional rating vs. unintentional influence), and the interaction between these two factors. All variables were standardized before analysis. The AMP effect size was a significant predictor, $b = 0.29$, $t = 5.33$, $p < .001$. There was also an effect of question wording, indicating that respondents were more likely to agree that they were unintentionally influenced ($M = 1.86$, $SD = 0.90$) than to agree that they intentionally rated the primes ($M = 1.46$, $SD = 0.97$), $b = 0.20$, $t = 3.67$, $p < .001$. Most important, there was no interaction between question wording and AMP effect size, $b = 0.06$, $t = 1.16$, $p = .25$. Participants who showed the largest priming effects thus claimed both that they intentionally rated the primes ($b = 0.22$, $t = 2.67$, $p < .01$) and that they were unintentionally influenced by the primes ($b = 0.37$, $t = 4.96$, $p < .001$).

According to the confabulation account, both of these relationships result from participants perceiving their own behavior and explaining it based on whatever cues are suggested in the question. If so, then the associations with both questions should be mediated by self-reports about the size of the priming effect. We tested this hypothesis using an SPSS script for assessing mediation (Preacher & Hayes, 2008). For the original question, perceived priming was a significant mediator of the association between the AMP effect and reported intentional prime ratings, Sobel $z = 3.92$, $p < .001$ (see Figure 1, top panel). For the reverse-worded question, a parallel effect emerged. Perceived priming was a significant mediator of the association between the AMP effect size and reports of unintentional influence, Sobel $z = 4.84$, $p < .001$ (see Figure 1,

bottom panel). Moreover, after controlling for perceived priming, the estimate of the AMP effect size was no longer significant for either question. This suggests that perceived priming fully mediated the associations between the priming effect on one hand and reports of intentional prime rating or unintended influence on the other.

Reliability. Following the method described in Bar-Anan and Nosek (2012), we computed a split-half reliability by dividing AMP trials into two halves and taking the difference score between Black and White primes on each half. The overall Spearman–Brown reliability coefficient was .59. We next tested the moderating effects of self-reported intentions. Replicating Bar-Anan and Nosek, reports of intentionally rating the primes moderated the association between the first half and second half of the AMP, $b = 0.22$, $t = 2.62$, $p < .01$. The convention of testing simple slopes at one standard deviation from the mean was not appropriate here because the skewed distribution of the intent question resulted in point estimates (± 1 *SD*) that were outside the range of observed data. We therefore computed simple slopes for the maximum and minimum observed scores (“not at all” and “almost always”). Simple slopes were significant at both the minimum score, $b = 0.25$, $t = 3.15$, $p < .01$, and the maximum score, $b = 1.15$, $t = 3.62$, $p < .001$.

Next we tested the moderating effect of reporting unintentional influence from the primes. As expected, reports of unintentional influence moderated the reliability estimates, $b = 0.29$, $t = 4.49$, $p < .001$. The effect was in the same direction as for the original question. Simple slopes showed that the association was significant at the maximum scale rating of unintentional influence, $b = 1.29$, $t = 6.31$, $p < .001$, but not at the minimum, $b = 0.02$, $t = 0.12$, $p = .90$. Finally, we included both versions of the question in a single analysis and tested the three-way interaction with question wording. The three-way interaction was not significant, $b = 0.04$, $t = 0.80$, $p = .42$, indicating that greater reliability was associated with both the perception that participants had intentionally rated the primes and that they had been unintentionally influenced by the primes.

Association With Explicit Attitudes. To test whether reported intentions moderated the relationship between the AMP and explicit attitudes, we conducted a regression analysis predicting the feeling thermometer scores (the difference between warmth toward Whites and Blacks) from AMP scores, responses to the intention questions, and their interaction. We included a three-way interaction term to test whether question wording qualified the results. We found a main effect of AMP scores, $b = 0.20$, $t = 3.34$, $p = .001$. However, the AMP by intentions question interaction was nonsignificant, $b = 0.03$, $t = 0.55$, $p = .59$, and there was no three-way interaction with question wording, $b = 0.05$, $t = 0.77$, $p = .44$.

Although Bar-Anan and Nosek reported that the AMP was associated with other attitude measures (including

feeling thermometers) more strongly among respondents who reported intentionally rating the primes, they did not report significance tests for the interaction. It is therefore difficult to know whether the present findings should be considered a failure to replicate. To facilitate comparison, we examined the association as they did, between the AMP and feeling thermometers for respondents who reported that they never intentionally rated the primes (“not at all”; $n = 116$) and those who reported that they sometimes did (“a little” or more; $n = 37$). The implicit–explicit correlation among those who reported never rating the primes was smaller ($r = .08$, $p = .42$) than for those who reported intentionally rating the primes at least a little ($r = .33$, $p = .05$), although these correlations were not significantly different, $z = 1.34$, $p = .18$. Despite the lack of a significant difference, the results are consistent in direction with those reported by Bar-Anan and Nosek (2012). We repeated the same comparison for participants who reported never being unintentionally influenced ($n = 59$) and those who reported that they were unintentionally influenced at least a little ($n = 75$). Again, the implicit–explicit correlation was smaller among those who reported no unintentional influence ($r = .21$, $p = .10$) than those who reported some unintentional influence ($r = .31$, $p < .01$), although the differences were not significant, $z = .60$, $p = .55$. Critically, the relative magnitudes were in the same direction for reports of intentionally rating the primes and for reports of being unintentionally influenced.

Discussion

We found that participants who showed the largest and most reliable priming effects were the most likely to report that they had intentionally rated the primes. At the same time, participants with the largest and most reliable priming effects were also the most likely to claim that they were unintentionally influenced by the primes. If large priming effects were really caused by intentionally rating the primes, and reported intentions accurately reflected this, then we would have expected to see a very different pattern. Namely, participants with large priming effects would have claimed that they intentionally rated the primes, and they would have denied that they were unintentionally influenced.

If we follow Bar-Anan and Nosek (2012) and interpret these correlational results as causal evidence, we must conclude that (a) the AMP’s psychometric qualities are highly dependent on participants who intentionally rated the primes and (b) the AMP’s psychometric qualities are highly dependent on participants who were unintentionally influenced by the primes. How can both be true? This apparent paradox disappears once we recognize that both kinds of reports are post hoc constructions in which participants observe their own behavior and then try to make sense of it using whatever cues are readily available. Asking respondents whether they intentionally rated the primes suggests one plausible explanation, and asking them whether they were unintentionally

influenced suggests another. Such leading questions are a means to invent intentions where none existed. Neither confabulation is an accurate guide to how participants made their ratings in the AMP.

Participants may have based their retrospective reports on the number of prime-consistent responses (effect size) or the consistency of prime-consistent responses (reliability), or both. The fact that retrospective reports were associated in similar ways with both effect size and reliability is not surprising, because the effect size sets the upper limit for internal consistency in this task. That is, the most reliable possible pattern of responses is to respond in a prime-congruent manner on every trial or a prime-incongruent manner on every trial. Effect size (absolute value) and internal consistency will therefore tend to be correlated. Moreover, reliability sets the upper limit for how strongly a measure can be associated with any other measure. This may explain why the AMP was slightly (though not significantly) more strongly correlated with explicit attitudes among those whose retrospective reports indicated a larger/more reliable priming effect.

Experiment 2

Results of Experiment 1 suggest that retrospective self-reports cannot reveal the true causes of priming effects in the AMP. A more rigorous way to establish causality is to manipulate what participants intend to do. In Experiment 2, we asked participants to complete the AMP twice: once under standard instructions to rate the pictographs without influence from the primes (an indirect attitude test) and once under instructions to rate the primes without influence from the pictographs (a direct test). This method allows implicit and explicit measures to be compared directly, while holding constant methodological factors that usually differ in multiple ways (Payne, Burkley, & Stokes, 2008). Most important for present purposes, this approach manipulates intentions prospectively: In the direct test, participants intend to rate the primes, but in the indirect test, they intend not to rate the primes. If the systematic variance in AMP responses is driven by participants who intentionally rate the primes, then the two versions of the test should produce largely the same results. Specifically, the systematic variability in each should be the same, and they should differ only in their error variance. We tested this hypothesis in two ways. First, we examined the ability of each test to predict social judgments about a person who was described as either White or Black. Second, we examined whether intentional versus unintentional ratings were differentially related to motivations to appear unbiased.

Method

Participants. Forty-five undergraduate students (26 women) participated for partial course credit. The sample included White (80%), Black (11%), Asian (4.5%), and Hispanic (4.5%) respondents. No participants were excluded.

Design and Procedure. The experiment was a 2 (Black vs. White character) \times 2 (Direct vs. Indirect rating) design. The race of the character was manipulated between participants and the direct/indirect rating was manipulated within participants. Participants completed the impression formation task first, followed by the two AMP versions. We chose this order because it highlights a situation where implicit and explicit responses are likely to diverge. When asked to form an impression as the first task, participants were expected to rely, in part, on racial stereotypes. After making personality judgments about a Black individual, however, participants might be sensitized to race stereotypes on later measures. This reaction can be expected to have more impact on the direct test than the indirect test. Finally, participants completed the Internal Motivation to Control Prejudice Scale (Plant & Devine, 1998). After completing all materials, participants provided demographic information and were fully debriefed.

Materials

Impression formation task. Participants were told that in a previous study, students kept a diary of the events that happened to them throughout the day. The participants were then told that they would read a description summarized by the researchers based on a randomly selected excerpt from one of these diary entries. Participants were randomly assigned to read a summarized entry ostensibly from a Black or White student. Information about the author was presented as a completed demographic questionnaire, in which race was indicated along with sex (male), academic major (psychology), year in school (sophomore), expected graduation month (May 2009), and home town (Charlotte, North Carolina). The Black author's name was Tyrone and the White author's name was Eric. All participants were presented with the same one-paragraph summary, adapted from Lambert, Cronen, Chasteen, and Lickel (1996). It was designed to be ambiguous as to the author's aggressiveness or friendliness. After reading the description, participants rated the author on several filler traits (e.g., honest) and the two traits of interest: aggressive and friendly. Ratings were made on a 7-point scale ranging from 1 (*not at all*) to 7 (*extremely*).

Indirect and direct AMP. Participants were told the purpose of the task was to assess how well they could make judgments in the face of distractions. For each AMP trial, participants were presented 1 of 12 Black or 12 White faces as a prime. The prime appeared in the center of the screen for 100 ms, followed by a blank screen for 100 ms and then a randomly selected Chinese pictograph for 100 ms. We presented the prime and target items for the same durations (primes were presented for 100 ms rather than 75 ms as in Experiment 1) in this study because we wanted participants to be equally able to focus on either set of stimuli. After the pictograph appeared, it was followed by a black-and-white mask that remained on the screen until participants made their response. Indirect and direct ratings were completed in separate blocks, counterbalanced for order.

For indirect ratings participants were told, “some of the Chinese characters represent aggressive words and some represent friendly words.” They were then instructed to guess which Chinese characters appeared to reflect aggressive or friendly words based on their intuitive reactions to characters. Participants were instructed to not let their evaluation of the faces impact their ratings of the Chinese symbols. Below the mask, a rating scale displayed four choices: very aggressive, slightly aggressive, slightly friendly, and very friendly. Several studies have shown that priming effects in the AMP can reflect not only affective reactions but also semantic misattributions (Deutsch & Gawronski, 2009; Förderer & Unkelbach, 2011; Imhoff, Schmidt, Bernhardt, Dierksmeier, & Banse, 2011; Sava et al., 2012). We selected the friendly/unfriendly stereotype dimension because it most closely matched the personality dimension assessed in the impression formation task. We used a 4-point scale to facilitate comparison with direct ratings, which are more commonly assessed using continuous rather than binary ratings.

For direct ratings, participants were asked to rate how friendly or aggressive the person in the photo looked. Participants rated each of the 12 Black and 12 White faces once. They were instructed not to let the pictographs influence their ratings of the faces. Although the minimal content of the pictographs and their random pairing with photos means they are very unlikely to have any systematic effects on ratings, this instruction was used to keep the direct rating task parallel to the indirect task.

Results

Personality judgments were scored by subtracting friendliness judgments from aggressiveness judgments, with higher values reflecting judgments of greater aggressiveness. The indirect and direct AMP ratings were scored as the difference between aggressiveness ratings on White and Black prime trials, with higher values reflecting more aggressive judgments on Black trials than White trials. We conducted a regression analysis with personality judgments as the dependent variable. In the first step, the impression character’s race (coded as -1 for White and 1 for Black), indirect ratings, and direct ratings were entered as independent variables (all standardized). In the second step, the Indirect \times Target race and Direct \times Target race interactions were entered.

Results revealed a marginally significant main effect for the indirect test, but this was qualified by the expected Indirect \times Target race interaction (see Table 1). No other effects were significant. To clarify this effect, personality judgments were regressed on indirect and direct tests for the White impression character and Black impression character conditions separately. In the White character condition, neither direct ($\beta = .23$, $t = 0.87$, $p = .40$) nor indirect tests ($\beta = -.05$, $t = 0.20$, $p = .85$) were significantly related to personality judgments. But in the Black target condition, the indirect

Table 1. Regression Coefficients Predicting Personality Judgments, Experiment 2

| Independent variable | β | t | p |
|-------------------------------|---------|------|------|
| Target race | .06 | 0.38 | .70 |
| Indirect ratings | .28 | 1.73 | .09 |
| Direct ratings | -.02 | 0.12 | .91 |
| Indirect \times Target race | .46 | 2.81 | .008 |
| Direct \times Target race | -.18 | 1.08 | .29 |

test was significantly related to personality judgments ($\beta = .58$, $t = 3.20$, $p < .01$), whereas the direct test was not ($\beta = -.12$, $t = 0.67$, $p = .51$).²

As predicted, indirect AMP ratings predicted impressions independent of direct ratings. This pattern is inconsistent with the idea that systematic variability in the (indirect) AMP is driven by intentional ratings of the primes, as suggested by Bar-Anan and Nosek (2012). Our hypothesis, in contrast, was based on the assumption that participants would be more able to adjust their responses on the direct test than the indirect test. To test this possibility, direct ratings were compared in the White target and Black target conditions using an ANOVA. As expected, direct test scores were lower (indicating less stereotypical judgments) in the Black impression character condition ($M = 0.02$, $SD = 0.43$) than the White character condition ($M = 0.46$, $SD = 0.53$), $F(1, 43) = 9.51$, $p < .01$. No such difference was found for the indirect test scores, $F(1, 43) = 0.92$, $p = .34$.

To further examine whether direct and indirect tests were redundant, we examined the implicit–explicit correlations. Overall, indirect and direct ratings were moderately correlated, $r = .38$, $p = .01$. The correlation was significant in the White target condition, $r = .46$, $p < .05$, but not the Black target condition, $r = .18$, $p = .39$. The two correlations were not, however, significantly different from each other ($p = .33$). Furthermore, motivation to control prejudice was significantly associated with bias in direct ratings, $r = -.39$, $p < .01$, but not indirect ratings, $r = -.06$, $p = .69$. This is consistent with the idea that direct ratings were under intentional control to a greater degree than indirect ratings were. This finding is inconsistent with the idea that AMP priming (the indirect test) is driven by intentional prime ratings because intentional prime ratings (i.e., the direct test) showed systematically different patterns of associations.

Discussion

We found that intentionally rating the primes in the direct test produced systematically different results than trying to rate the pictographs. First, the two tests were differentially associated with impressions of a Black individual. Second, direct (intentional) ratings were more affected than indirect ratings by previously rating the personality of a Black person

in comparison with a White person. Third, direct (intentional) but not indirect ratings were associated with motivation to control prejudice. Together, these findings suggest that direct and indirect ratings were not redundant. Instead, they showed patterns consistent with typical dissociations between implicit and explicit measures (e.g., Cameron et al., 2012). If the validity of the AMP depended on intentional ratings of the primes, then all of the systematic variability would be reflected in the direct test. Thus, these dissociations provide additional evidence that intentional ratings are not responsible for the validity of the AMP. By manipulating intentions prospectively, we were able to avoid the problems of retrospective constructions.

Experiment 3

Our first experiment suggested that confabulations of intent—both intentionally rating primes and being unintentionally influenced by them—were based on subjective perceptions of whether the primes influenced responses. These perceptions of priming showed above-chance accuracy, similar to the finding that led Bar-Anan and Nosek (2012) to conclude that “awareness is essential for the AMP’s validity” (p. 13). But does the perception of the primes’ influence really contribute to priming effects? For the reasons already elaborated, retrospective self-reports are not a suitable method for finding out. Awareness can only provide a basis for controlling behavior that comes afterward. Reading about a horse race in the paper the next day does not help anyone to make money at the track. And inferring that one has probably been influenced by primes does not increase or decrease priming effects. Only prospective awareness can provide a basis for adjusting responses. If participants can detect that their response to a particular pictograph is being influenced by a prime before they press the key, then they have an opportunity to control how they respond.

In Experiment 3, we gave participants an easy opportunity to adjust their responses based on subjective experience. Namely, we allowed participants in one condition to choose whether to respond or “pass” on each trial. Past research in memory has shown that when subjective experience is well calibrated to accuracy, participants can take advantage of a pass option to selectively avoid responding when they are likely to be incorrect, thereby raising accuracy (Koriat & Goldsmith, 1996; Payne, Jacoby, & Lambert, 2004). However, when subjective experience is poorly calibrated, the pass option provides little benefit. In the present study, we reasoned that if a participant is aware when she is being influenced by a prime, then she can pass when she would otherwise display a priming effect. The trials on which he or she chooses to forego the pass option and evaluate the pictograph should therefore be free of influence from the primes. If subjective experiences of being influenced by the primes are well calibrated to actual influence, then the pass option should allow respondents to eliminate the priming effect. In this study, we

allowed one group to have a pass option on each trial, while a second group had to respond on every trial. By comparing the magnitude of priming in the two conditions, we can estimate how much awareness participants had about the influence of primes in real time as they formed their evaluations of the pictographs.

Method

Participants. Seventy-two undergraduate participants took part in this study in return for course credit. Gender and race were not recorded because of a software error. No participants were excluded from analysis.

Design and Procedure. Because the hypotheses in this experiment focused on mean responses rather than individual differences, we selected items that were consensually evaluated as pleasant, unpleasant, or neutral. The experimental design was 3 (Prime valence: pleasant, unpleasant, or neutral) \times 2 (Response options: passing allowed or not allowed). Prime valence was within-participants and response option was between-participants. All participants were told that the prime served as a warning signal for the Chinese character and were informed that they should not allow the pictures to influence their ratings of the pictographs. Participants in the no-pass condition were given standard instructions: They were told to press one of two keys to indicate whether the pictograph was more or less pleasant than the average Chinese pictograph. Those in the pass-option condition were given these two response options, along with a third option to pass on the trial by pressing the space bar whenever they thought their evaluations of the pictographs might be influenced by the prime. The instructions emphasized that participants should evaluate the pictograph only whenever they believed that their opinion reflected the qualities of the pictograph itself. Otherwise, they were instructed to skip the trial whenever they felt their evaluation of the pictograph was influenced by the prime.

During each trial of the priming task, a prime image appeared at the center of the screen for 75 ms, followed by a blank screen for 125 ms, then a Chinese pictograph for 100 ms, and then a black-and-white pattern mask. The next trial began as soon as the participants made a response. Participants completed a total of 72 randomly ordered trials, with 24 each of positive, negative, and neutral primes. The positive and negative primes were images selected from Lang, Bradley, and Cuthbert (1995) norms and were matched on arousal ratings. An image of a gray square served as the neutral prime.

Results

If participants were aware of the primes’ influence at the time they made their judgments, then they should pass when they are most likely to be influenced. This means that they should pass more frequently when the primes were pleasant

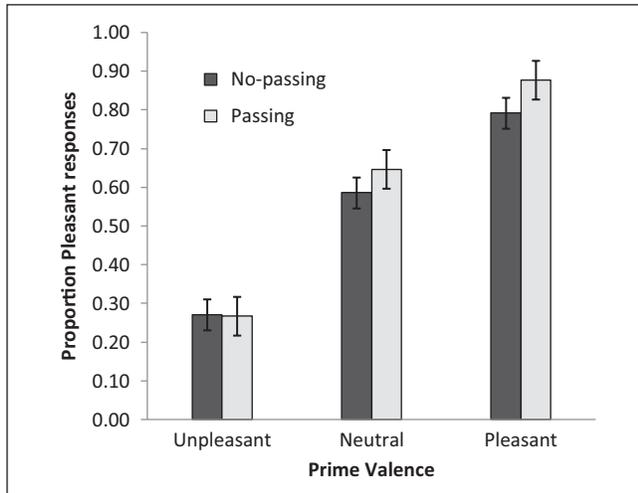


Figure 2. Proportion pleasant responses as a function of prime valence and passing-allowed versus no-passing conditions
Note: Error bars reflect 1 SE.

or unpleasant than when the primes were neutral. We first examined the proportion of “pass” responses in the pass-option condition as a function of the primes’ valence. Participants passed much less when the primes were pleasant ($M = 0.14$) or unpleasant ($M = 0.17$) than when the prime was neutral ($M = 0.54$), $F(2, 70) = 28.23$, $p < .001$. Passing rates on neutral trials were significantly higher than pleasant trials, $F(1, 35) = 34.0$, $p < .001$, or unpleasant trials, $F(1, 35) = 25.65$, $p < .001$, which were not significantly different from each other, $F(1, 35) = 1.71$, $p = .20$. This pattern suggests that the affective primes may have caused participants to like or dislike the pictograph, and therefore to rate it rather than pass. Ironically, the affective valence of the primes may have made participants least likely to pass when they were most likely to be influenced. Such deceptive phenomenology contradicts the idea that participants were aware of the influence of primes in real time. If participants were unaware of how the primes influenced each evaluation, then having the pass option should not reduce the priming effect.

We compared the influence of the primes on the proportion of pleasant responses in the pass and no-passing conditions. In the passing group, responses were computed as the proportion of pleasant responses out of the number of non-passed trials for each participant. The proportion of pleasant responses was analyzed using a 2 (Response option group) \times 3 (Prime valence) ANOVA. As shown in Figure 2, there was a strong main effect of priming, $F(2, 134) = 91.37$, $p < .001$. Critically, the main effect of prime valence was not qualified by the Prime valence \times Response option interaction, $F(2, 134) = 0.60$, $p = .55$. The priming effect was significant when the pass option was available, $F(2, 64) = 52.10$, $p < .001$, and when it was not available, $F(2, 70) = 39.72$, $p < .001$. These data contradict the idea that participants were aware that the primes influenced them before responding.

Discussion

Providing participants an option to pass offered them an opportunity to adjust their behavior based on subjective experience. Results indicated that they did not take advantage of this opportunity. This finding strongly suggests that subjective experiences were poorly calibrated to the actual influence of the primes. Whatever theories participants had about the primes’ influence, they did not base their responses on awareness of the primes’ true effects. Participants continued to evaluate pictographs on what they believed to be the qualities of the pictographs. Those perceived qualities, however, were driven by the affective primes. This pattern cannot be explained easily by any account which posits that participants are aware of the influence of primes as they are rating the pictographs. Moreover, this pattern is not easily explained by a lack of statistical power. First, the priming effect was slightly larger in the passing condition than in the no-passing condition. Second, a power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) indicated power greater than .85 to detect an interaction between the repeated measures and between-subjects factors, even assuming a small effect size. The pattern is easily explained, however, by a misattribution account in which participants mistake their affective reaction to the primes for their feelings about the pictographs (see Oikawa, Aarts, & Oikawa, 2011, for independent evidence for the misattribution account). The present data suggest that perceived priming is not simply a read out of subjective experiences at the time of their AMP ratings but rather a reconstruction afterward.

General Discussion

Summary

In three experiments, we tested the claim that reliable and valid measurement in the AMP depends on awareness and intentions to rate the primes rather than the targets. In short, it does not. Experiment 1 found that larger and more reliable AMP effects were associated with retrospective self-reports that participants had intentionally rated the primes *and* that they had been unintentionally influenced by the primes. This logically incoherent pattern of self-reports presents a puzzle if we take them as faithful descriptions of why participants responded as they did. But if we instead interpret these self-reports as confabulations constructed to explain responses, then the puzzle dissolves. Experiment 2 manipulated intention prospectively, and found that when participants intentionally rated the primes, their responses differed systematically from when they intended to rate the targets. Contrary to what would be predicted by Bar-Anan and Nosek’s (2012) account, intentional ratings of the primes were more reactive following judgments of a Black character and less predictive of biased judgments compared with unintentional priming effects. Moreover, intentional ratings

were associated with motivations to control prejudice, whereas the priming effect was not, suggesting that intentional ratings are susceptible to self-presentation strategies. This dissociation suggests that indirect ratings cannot not be driven by intentional ratings. Finally, Experiment 3 used a prospective manipulation of behavioral options that allowed participants to modify their responses when they were aware of influence from the primes. This opportunity for selective responding did not eliminate or even reduce the priming effects. This metacognitive failure suggests that participants lacked real-time insight into when their judgments were influenced by primes.

Theoretical Implications

Together, these studies suggest a new interpretation of the findings reported by Bar-Anan and Nosek (2012). Our findings suggest that participants do not have clear introspective access to the causes of their responses in the AMP. Instead, they observe their own responses and then tell plausible stories about how they might have come about. As has been documented in many areas of psychology, the stories people tell are shaped by the questions they are asked. As with eyewitness testimony, asking leading questions can lead to the invention of intentions where none existed before (Loftus, 1975). In the hindsight bias, people learn about an outcome and then are asked to judge what they knew beforehand. People tend to claim that they “knew it all along” (Fischhoff, 1975; Nestler, Egloff, Küfner, & Back, 2012). However, they cannot make predictions about what will happen next based on their retrospective constructions. Retrospective constructions such as these produce the feeling that we know exactly what we are doing. But only real-time experience allows us to truly control what will happen next, and only prospective manipulations can test the effects of real-time experience.

Bar-Anan and Nosek (2012) anticipated this objection and argued that “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” (p. 13). However, this is mistaken because post hoc confabulations cannot identify whether an individual’s attitude has been validly measured by the AMP. To see why, let us consider how such confabulations might come about.

Under the confabulation account, participants are assumed to try to follow the instructions. Their ratings of the Chinese symbols may be influenced by the primes to varying degrees, based on factors such as how strong their attitudes toward the primes are and how well they can keep attention focused on the targets rather than the primes (see Payne, Hall, Cameron, & Bishara, 2010, for a formal model of this process). Some individuals have strong automatic evaluations of the primes, which result (all else being equal) in larger priming effects. Others are indifferent toward the primes, and

because the pictographs are randomly paired with primes and presented in a random order, there will be very little that is systematic about these participants’ responses. Their priming effects will be close to zero, and there will be little internal consistency because they are simply rating a random set of unrelated items. Later, when asked to indicate whether they intentionally rated the primes, both groups reconstruct memories of their own responses and agree with explanations that seem plausible.

Notice that one cannot be agnostic about whether the retrospective reports are illusory and still attempt to use those reports as diagnostic information about the AMP’s validity. If the confabulation account is true, then reports of intent are illusory and they result from validly measured preferences (otherwise they would not be associated with actual priming effect sizes). Thus, individual differences in illusory reports of intent cannot provide evidence about whose attitudes are validly measured and whose are not. Retrospective reports of intent could only provide diagnostic evidence about validity if they had accurately reflected the causal processes producing priming effects.

The confabulation account supported here argues that reports of intent follow from self-perceived priming effects. As an anonymous reviewer noted, this suggests an additional test of the confabulation account against Bar-Anan and Nosek’s interpretation. Under the confabulation account (but not the veridical report account), manipulating attitude strength should affect reports of intentionally rating the primes. In fact, we did not need to conduct this study because Bar-Anan and Nosek (2012, Study 2) reported just such a study. They experimentally created strong or weak attitudes toward a fictional character, then administered the AMP, and finally subjects reported whether they intentionally rated the primes. As predicted by the confabulation account, stronger attitudes caused participants to report greater intentional rating of the primes. But rather than interpreting this finding as evidence that reports of intention were illusory, the authors instead split respondents by how often they reported intentional prime ratings, and then interpreted the size of the priming effects within each self-report group. They found significant priming effects (reflecting the attitude strength manipulation) only among respondents who expressed some degree of intentional prime ratings. This is not surprising because intention reports were affected by the manipulation; conditionalizing on these reports guarantees that the effect of the manipulation will be weaker among those who reported no intentional prime ratings.

The inferential approach taken in Bar-Anan and Nosek (2012) can be used to make any experimental effect appear to be invalid among certain subgroups. First, take any experimental effect, E (i.e., a difference between two experimental conditions), that has systematic variability across persons. Second, find a variable, V , that is positively correlated with the size of the experimental effect (for any reason, spurious or meaningful). Third, divide the sample into subjects who

are high versus low on V, and select the subgroup that is low. Finally, interpret the experimental effect among the low-V subgroup. It is bound to be small because it has been selected to be so. All else equal, the stronger the correlation between E and V, the weaker the experimental effect will appear among the low-V group. Substantive interpretations of the experimental effect in this group will be biased to conclude that the experimental effect E is invalid.

This problem is essentially the same problem identified by Vul, Harris, Winkielman, and Pashler (2009) of puzzlingly high correlations in some brain imaging studies (popularly discussed as “voodoo correlations”). In that case, selecting brain regions that are significantly related to a behavior, and then interpreting the association between behavior and brain activation within those regions leads to overestimated effects. This problem of nonindependence arises whenever the same criteria are used to select a sample (or a subsample) and also to evaluate statistical results based on that sample. In the studies of Bar-Anan and Nosek (2012), the initial selection is based on the association between retrospective self-reports and effect size (or internal consistency) in the AMP. Effect sizes and internal consistency are then interpreted within each subgroup, raising the problem of nonindependence. The problem is the same: Interpretations of statistics based on nonindependent observations will tend to be biased.

It might be objected that Bar-Anan and Nosek (2012) also used correlations with other attitude measures as an independent criterion. However, correlations with other attitude measures do not solve the problem because these too are affected by effect size and internal consistency. Effect size in implicit attitude measures reflects attitude strength, and correlations among attitude measures are larger for strong attitudes (Nosek, 2007). Internal consistency reflects the ratio of true score variance to error variance, which sets the upper limit on the expected value of correlations with other measures. Therefore, differences in correlations with other attitude measures fall out naturally from differences in effect size and reliability.

How can the nonindependence problem be avoided? The simplest way is to use random assignment to experimental conditions rather than correlations with self-report. The experimental method not only allows a prospective means to establish causality as we have discussed, but it also provides independent observations. In conclusion, we found that retrospective self-reports of intention produced incoherent patterns as participants searched for plausible ways to explain their implicit responses. In contrast, experimental manipulations yielded no evidence that the AMP’s validity depended on awareness or intentions. Intentional ratings did not cause AMP effects; AMP effects caused reports of intentional ratings. In evaluating the validity of implicit measures, it is critical to distinguish between the real causal effects of intentions and intentions that are invented afterward.

Acknowledgment

We thank Sara Algoe for thoughtful comments on an earlier draft of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Science Foundation Grant 0924252.

Notes

1. The only difference in wording for the items was that Bar-Anan and Nosek referred to the Chinese pictographs as “drawings” whereas we referred to them as “symbols.”
2. The finding that racial attitudes were more strongly related to judgments of the Black than White character is consistent with previous research suggesting that attitudes toward majority groups are often less meaningful and less predictive than attitudes toward minorities (Lambert et al., 1996; Lambert, Payne, Ramsey, & Shaffer, 2004).

References

- Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition: An International Journal*, *14*, 439-458.
- 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.
- Bar-Anan, Y., Wilson, T. D., & Hassin, R. R. (2010). Inaccurate self-knowledge formation as a result of automatic behavior. *Journal of Experimental Social Psychology*, *46*, 884-894.
- Bem, D. J. (1967). Self perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*, 183-200.
- Cameron, C. D., Brown-Iannuzzi, J., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behaviors and explicit attitudes. *Personality and Social Psychology Review*, *4*, 330-350.
- Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on “automatic evaluations” as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, *45*, 101-114.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.
- Förderer, S., & Unkelbach, C. (2011). Beyond evaluative conditioning! Evidence for transfer of non-evaluative attributes. *Social Psychological and Personality Science*, 2, 479-486.
- Greenwald, A., Poehlman, T., Uhlmann, E., & Banaji, M. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17-41.
- Imhoff, R., Schmidt, A. F., Bernhardt, J., Dierksmeier, A., & Banse, R. (2011). An inkblot for sexual preference: A semantic variant of the affect misattribution procedure. *Cognition & Emotion*, 25, 676-690.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490-517.
- Kuhn, S., & Brass, M. (2009). Retrospective construction of the judgment of free choice. *Consciousness and Cognition: An International Journal*, 18, 12-21.
- Lambert, A. J., Cronen, S., Chasteen, A. L., & Lickel, B. (1996). Private vs. public expressions of racial prejudice. *Journal of Experimental Social Psychology*, 32, 437-459.
- Lambert, A. J., Payne, B. K., Ramsey, S., & Shaffer, L. M. (2004). On the predictive validity of implicit attitude measures: The moderating effect of perceived group variability. *Journal of Experimental Social Psychology*, 40, 708-722.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. (1995). *International affective picture system*. Gainesville: Center for Research in Psychology, University of Florida.
- Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, 32, 951-955.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7, 560-572.
- Nestler, S., Egloff, B., Küfner, A. P., & Back, M. D. (2012). An integrative lens model approach to bias and accuracy in human inferences: Hindsight effects and knowledge updating in personality judgments. *Journal of Personality and Social Psychology*, 103, 689-717. doi:10.1037/a0029461
- 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. (2007). Implicit-explicit relations. *Current Directions in Psychological Science*, 16, 65-69.
- Oettingen, G., Grant, H., Smith, P. K., Skinner, M., & Gollwitzer, P. M. (2006). Nonconscious goal pursuit: Acting in an explanatory vacuum. *Journal of Experimental Social Psychology*, 42, 668-675.
- 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.
- Parks-Stamm, E. J., Oettingen, G., & Gollwitzer, P. M. (2010). Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving. *Journal of Experimental Social Psychology*, 46, 531-542.
- 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., Hall, D., Cameron, C. D., & Bishara, A. J. (2010). A process model of affect misattribution. *Personality and Social Psychological Bulletin*, 36, 1397-1408.
- Payne, B. K., Jacoby, L. L., & Lambert, A. J. (2004). Memory monitoring and the control of stereotype distortion. *Journal of Experimental Social Psychology*, 40, 52-64.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879-891.
- 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.
- Sava, F. A., Maricuțoiu, L. P., Rusu, S., Macsinga, I., Virgă, D., Cheng, C. M., & Payne, B. K. (2012). An inkblot for the implicit assessment of personality: The semantic misattribution procedure. *European Journal of Personality*, 26, 613-628.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4, 274-290.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, 85, 5-19.
- Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480-492.