

# Effects on the Affect Misattribution Procedure are Strongly Moderated by Influence Awareness

Sean Hughes, Jamie Cummins, & Ian Hussey

The Affect Misattribution Procedure (AMP) is used in many areas of psychological science based on the assumption that it not only taps into attitudes and biases but does so without a person's awareness. Across eight preregistered studies (N = 1603) plus meta-analyses we reexamined the 'implicitness' of AMP effects, and in particular, the idea that people are unaware of the prime's influence on their evaluations. Results indicated that AMP effects and their predictive validity are primarily moderated by a subset of influence aware trials (within individuals), and high rates of influence awareness (between individuals). Interestingly, an individual's influence awareness rate on one AMP predicted how they performed on an earlier AMP, even when the two assessed different attitude domains. Taken together, our results suggest that AMP effects are not implicit in the way that has been claimed, a finding that has implications for the procedure, past findings, and theory. All materials and data are available at [osf.io/gv7cm](https://osf.io/gv7cm)

Over the past twenty years research on implicit cognition has grown from a relatively niche field into, what is today, one of the most prolific and widely examined topics in psychological science. The idea that our automatic thoughts, feelings, and actions shape downstream behavior drives research, theory, and application throughout the discipline, especially in social and personality psychology, neuroscience, health, cognitive, and clinical psychology (for a book length treatment see Gawronski & Payne, 2010).

The success of the topic has been due in large part to the development and widespread use of tasks known as *indirect measurement procedures*. In contrast to *direct measurement procedures*, which simply ask people to report on their thoughts, feelings, and actions, indirect procedures seek to probe the mind by interpreting performance (e.g., speed and/or accuracy) on experimental paradigms. Notable examples include the Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998), evaluative priming tasks (Hermans, De Houwer, & Eelen, 1994), and approach-avoidance tasks (Rinck & Becker, 2007; for a review see Gawronski & De Houwer, 2014). The outcomes of these procedures are commonly referred to as *implicit measures* (e.g., the IAT effect, priming effects; for more see De Houwer, 2006).

Indirect procedures are often deployed under the assumption that they limit a person's ability to control how they respond, or their need for introspective access and/or conscious awareness of the content under investigation (i.e., that they operate under the conditions of automaticity). As a result, these tasks are typically used when researchers want to gain insight into content that people may be unwilling or unable to report (see Greenwald et al., 1998; Hahn & Gawronski, 2019). Although debate continues about what implicit measures actually reflect (Brownstein, Madva, & Gawronski, 2019; Corneille & Hutter, 2020; Schimmack, 2021), a vast and ever-increasing number of studies continue to rely on indirect procedures and their effects to provide insights that other (self-report) procedures cannot.

## The Affect Misattribution Procedure

The Affect Misattribution Procedure (AMP) has emerged as one of the more popular indirect procedures (see Payne & Lundberg, 2014). At its core, the AMP consists of trials made up of three elements: (a) a prime stimulus (e.g., an image of a social in-group member) which is first flashed on screen for a brief period of time, followed quickly by (b) a target stimulus (usually a neutral Chinese pictograph), which is subsequently masked by (c) a white noise image. The AMP requires

participants to subjectively evaluate how visually pleasing the target stimulus is, while ignoring the prime that preceded it. Despite being explicitly told to disregard the prime when evaluating the target, people nonetheless evaluate the latter in ways that are consistent with the valence of the former. For instance, when a neutral Chinese pictograph is preceded by a social in-group member, people are more likely to evaluate it as pleasant, compared to when it is preceded by a social out-group member (Payne, Cheng, Govorun, & Stewart, 2005).

Since its creation, the AMP has attracted considerable attention. It is commonly used in social psychology to assess attitudes in domains such as race (Payne et al., 2005; Ditonto, Lau, & Sears, 2013; although see Teige-Mocigemba, Becker, Sherman, Reichardt, & Klauer, 2017), gender (Ye & Gawronski, 2018), sexuality (Imhoff, Schmidt, Bernhardt, Dierksmeier, & Banse, 2011), and politics (Payne et al., 2005; Kalmoe & Piston, 2013). It has been used to investigate the origins of attitudes and stereotypes (Dunham & Emory, 2014; Mann et al., 2019; Van Dessel, Mertens, Smith, & De Houwer, 2017), and to assess the effectiveness of attitude change interventions (Mann & Ferguson, 2017). In clinical psychology, it is used to assess, and sometimes provide prospective prediction of, maladaptive behaviors such as eating disorders, non-suicidal self-injury, alcoholism, anxiety, depressive symptoms, and physical abuse (Fox et al., 2018; Görgen, Joormann, Hiller, & Withöft, 2015; Jasper & Withöft, 2013; McCarthy, Skowronski, Crouch, & Milner, 2017; Smith, Forrest, Velkoff, Ribeiro, & Franklin, 2018; Zerhouni, Bègue, Comiran, & Wiers, 2018). Some clinical researchers also use the task as an outcome measure to benchmark the effectiveness of psychological interventions (Chapman et al., 2018; Schreiber, Withöft, Neng, & Weck, 2016).

### **Two Competing Accounts of the AMP Effect**

Two distinct perspectives have emerged to explain the aforementioned effects: an *implicit* account and an *explicit* account.<sup>1</sup> Both start from the position that AMP effects represent a valid measure of attitudes and bias. However, they differ in how “implicit” or “automatic” those effects are said to be. On the one hand, the implicit account argues that AMP effects reflect evaluations captured under certain conditions of automaticity (i.e., specifically, in the absence of

intention and awareness; Payne et al., 2005; Payne et al., 2013). On the other hand, the explicit account rejects this idea and argues that participants are aware of the prime’s influence on their evaluations, and exert intentional control over their behavior in order to respond in-line with those primes (e.g., Bar-Anan & Nosek, 2012; Mann et al., 2019). In what follows we briefly consider research which has examined the issues of awareness and intention of AMP effects.

### ***Intentionality in the AMP***

The idea that participants intentionally rely on the prime when evaluating the target can be argued on the basis of several studies. For instance, Bar-Anan and Nosek (2012) asked participants to first complete an AMP and then retrospectively indicate if they had intentionally based their evaluations on the prime rather than the target. They found that AMP effects were moderated by those who retrospectively reported intentionally rating the primes. Across a series of studies, Weil and colleagues (2017) exposed participants to either a bogus subliminal presentation task that falsely claimed to show certain targets (Chinese ideographs), or a genuine supraliminal presentation task that actually did so. Thereafter they completed an AMP. Half of the sample were told to evaluate the valence of the targets whereas the other half were asked to evaluate them in terms of their familiarity. Once the AMP was complete participants completed similar post-hoc intention questions as in Bar-Anan and Nosek (2012). Replicating Bar-Anan and Nosek (2012), AMP scores were positively correlated with intention ratings in the valence-judgment condition (but not in the familiarity condition) (also see Mann et al., 2019, for additional evidence on the role of intentionality in AMP effects).<sup>2</sup>

Proponents of the implicit account conducted several experiments which rejected these claims. For instance, Payne and colleagues (2013) found that the relationship between intentionality ratings and AMP effects was similar when people had to indicate if they were intentionally or unintentionally influenced by the prime. Drawing on this finding they claimed that people may be able to identify *that* they acted in a particular way, but they are unable to say *why* they acted in this way. In a second experiment, participants were asked to complete the AMP twice: once where they had to evaluate the target instead of prime (standard ‘unintentional’ AMP) and once where they had to evaluate the prime instead of the target (an ‘intentional’ AMP).<sup>3</sup> The authors found that the relationship between standard AMP and personality judgements of a

<sup>1</sup> The term ‘implicit’ does not represent an all-or-nothing concept, but rather is an umbrella term which refers to a set of automaticity conditions under which mental processes are said to operate (see Moors & De Houwer, 2006). The effects obtained from an indirect procedure are assumed to occur under one or more of these automaticity conditions. Thus to describe a measure or effect as implicit requires that one is clear about the exact automaticity conditions relevant to that effect. For those looking for an extensive debate about the meaning and usefulness of the term implicit we recommend Corneille and Hütter (2020).

<sup>2</sup> Note that Weil and colleagues offer an alternative (non-intentional) explanation for the pattern of empirical findings they obtained (see Weil et al., [2017], p.384).

<sup>3</sup> Payne et al. (2013) referred to these as these ‘indirect’ and ‘direct’ AMPs, respectively.

Black person were different to the relationship between the intentional AMP and that same personality judgement. These results and others (e.g., Gawronski & Ye, 2014) have been advanced in support of the idea that AMP effects are unintentional in nature.

In short, a number of studies investigating intentionality within the AMP have been carried out and debates continues as to whether AMP effects qualify as unintentional in nature.

### **Awareness in the AMP**

The implicit and explicit accounts also differ in how awareness and AMP effects are thought to be related to one another. Although both acknowledge that people can be aware that the prime has influenced their response to the target, they differ in the causal vs. correlational role that awareness is assumed to play. Proponents of the implicit account argue that awareness may be correlated with, but is not causally required to demonstrate, AMP effects, whereas proponents of the explicit account argue that participants are aware of the influence of the primes on their responses, and that it is possible this awareness is causally related to those effects.

For instance, Bar-Anan and Nosek (2012) asked participants in a self-reported (post-hoc) manner if they were aware of the prime's influence on their target evaluations. Across multiple studies they found that post-hoc, self-reported awareness predicted AMP effect sizes. That is, participants who reported no priming effect produced AMP effects that were psychometrically poor (no significant split-half correlation and no relationship with any of the other attitude measures) whereas those who reported being aware that priming had occurred showed stronger priming effects, split-half correlations, and relationships with other attitude measures.

Payne and colleagues (2013) responded to the above work. In one study (Experiment 3) they divided participants into two groups: the first completed a standard AMP, whereas the second completed a 'skip' AMP. During the latter AMP participants were given the option to respond in one of three ways: they could either indicate that the target stimulus was pleasant, unpleasant, or choose to 'skip' that trial entirely if they felt that their evaluation would have been influenced by the prime. The authors argued that if AMP effects were due to responding on trials where participants were aware of the prime's influence on their evaluations, then removing such trials "should eliminate the priming effect" (p. 377). When they compared skip-AMP effects (where influence aware trials had been removed) to standard AMP effects they found that the former did not significantly differ from the latter.

### **Awareness Revisited**

The aforementioned study by Payne et al. (2013) is often viewed as strong evidence in favor of the idea that AMP effects occur without awareness and are therefore implicit along this dimension. We believe this assumption may be premature for several reasons.

### **Methodological Issues**

On the surface the 'skip' AMP developed by Payne et al. (2013) appears to provide an *in vivo* measure of awareness insofar as participants are provided with an option to signal that the prime has influenced their evaluations. Such an approach is certainly more favorable than a single post-hoc self-report measure administered once the task is complete. However, even this task has its issues. Foremost amongst these is that it requires people to make an either/or decision: *either* provide an evaluative response *or* indicate that they were aware of the prime's influence. But it never allows one to do both (i.e., respond to the target *and* indicate this was a 'contaminated' response). This makes it impossible to *directly* compare performance on trials where people indicated that they were influence aware to those trials where they were non-influence aware. Without both pieces of information, it is difficult to determine whether trial-by-trial influence awareness moderates the magnitude of the AMP effect, or how AMP effects calculated from only influence aware trials compare to those calculated from non-influence aware trials.

### **Statistical Issues**

Payne et al. (2013; Experiment 3) also argued that effects on the standard AMP did not differ from those on the 'skip' AMP, and used this as evidence in support of the implicit account. Yet this conclusion is also problematic given that non-significant statistical differences between two means does not necessarily imply that they are statistically equivalent (Lakens, Scheel, & Isager, 2018; Quertemont, 2011). As such the original inference drawn was not supported by the analyses that were conducted in the paper.

### **Conceptual Issues**

Finally, the authors noted that participants tended to skip trials more frequently on trials with neutral compared to valenced primes, and suggested that such a pattern could be explained by the implicit but not the explicit account (i.e., that if people were aware they should skip when confronted with valenced primes and not with neutral primes). But an explicit account can be formulated which assumes that AMP effects arise because a subset of participants, on a subset of trials, intentionally and with awareness, use the prime's valence to determine their response to the target. In cases where the prime is neutral there is no evaluative information available which one can use to guide their response to the target. Thus it follows that they will

skip more on such trials. The opposite is true on valenced prime trials and thus skipping should occur less frequently.

### **Conclusion**

Given these conflicting accounts of the role of awareness in the AMP and a combination of methodological (absence of information about influence aware trial performance), statistical (conflation of statistical non-significance and statistical equivalence), and conceptual issues (equally plausible explanation of findings by the explicit account) within certain studies, it seems reasonable to continue examining whether participants really are aware of the prime's influence on their evaluations.

### **The Current Research**

Across eight preregistered studies (1 replication and 7 novel studies) we re-examined the implicitness of AMP effects and, in particular, the assertion that people are unaware of the prime's influence on their evaluations. In Experiment 1 we conducted a high-powered, preregistered study that sought to replicate the finding that standard and 'skip' AMPs are no different to one another. As we previously noted, the finding that 'skip' AMP effects are no different to standard AMP effects is viewed as strong support for the implicit account. To briefly preface what is to come, we found that scores on the standard AMP were significantly larger than those on the 'skip' AMP, a finding inconsistent with previously formulated implicit accounts.

In Experiment 2 we sought to address a key limitation of the original 'skip' AMP - namely - that it forces people to either skip or evaluate the target and thus only provides partial data. We developed an influence aware (IA-) AMP that had participants rate the target (provide evaluative information) and then indicate if evaluations had been influenced by the prime (provide influence information). We found that AMP effects were moderated at the trial-by-trial level by influence awareness, as well as by at the group level by inter-individual differences in influence awareness.

In Experiments 3 and 4 we controlled for the possibility that by probing for influence awareness on each trial of the IA-AMP we artificially altered the relationship between awareness and AMP effects. Participants now completed a standard AMP at Time 1

and an IA-AMP at Time 2, either from the same (Experiment 3, i.e., both generic valence) or different attitude domains (Experiment 4, i.e., one generic valence and one politics). In both cases influence awareness during an IA-AMP at Time 2 were related to the magnitude of standard AMP effects at Time 1, suggesting that influence awareness is a stable (within-participant) pattern of responding that holds within and between content domains.

In Experiment 5 we had two groups of participants (Democrats and Republicans) first complete a political IA-AMP and then an IA-AMP with generic valenced primes. We found that the AMP's ability to correctly classify a person as a Democrat or Republican was superior when effects were based solely on influence aware trials and inferior when based solely on non-influence aware trials. Experiment 6 had participants first complete a newly developed version of the AMP that purportedly reduces subset effects within the AMP (the Mann et al. [2019] modifications to the AMP) followed by a Mann et al. IA-AMP. Once again, the same pattern of findings emerged as outlined above, even within a variant of the task designed to optimize the implicitness of the AMP.

In our final two studies we modified the IA-AMP so that influence awareness was measured *prospectively*, either before the target was evaluated (Experiment 7) or before the target stimulus was even presented (Experiment 8). In this way influence awareness was measured before an overt evaluation took place or a covert evaluation could even be formed. In both studies the same pattern of findings emerged as before (see Figure 1). Taken together, our studies extends beyond previous work in this area by investigating awareness in both retrospective and prospective ways. They do so using multiple measures (single post-hoc self-reports, trial-by-trial online measures), versions of the AMP (standard, Influence Aware version, Mann et al. version), and attitude domains (political, positive vs. negative). They explore the bidirectional relationship between performance on one AMP and that of another (something that has not been examined in prior awareness work), and examine how awareness plays out at the trial, individual, and group levels. The result is a level of generalizability about awareness in the AMP that extends beyond that which has come before.

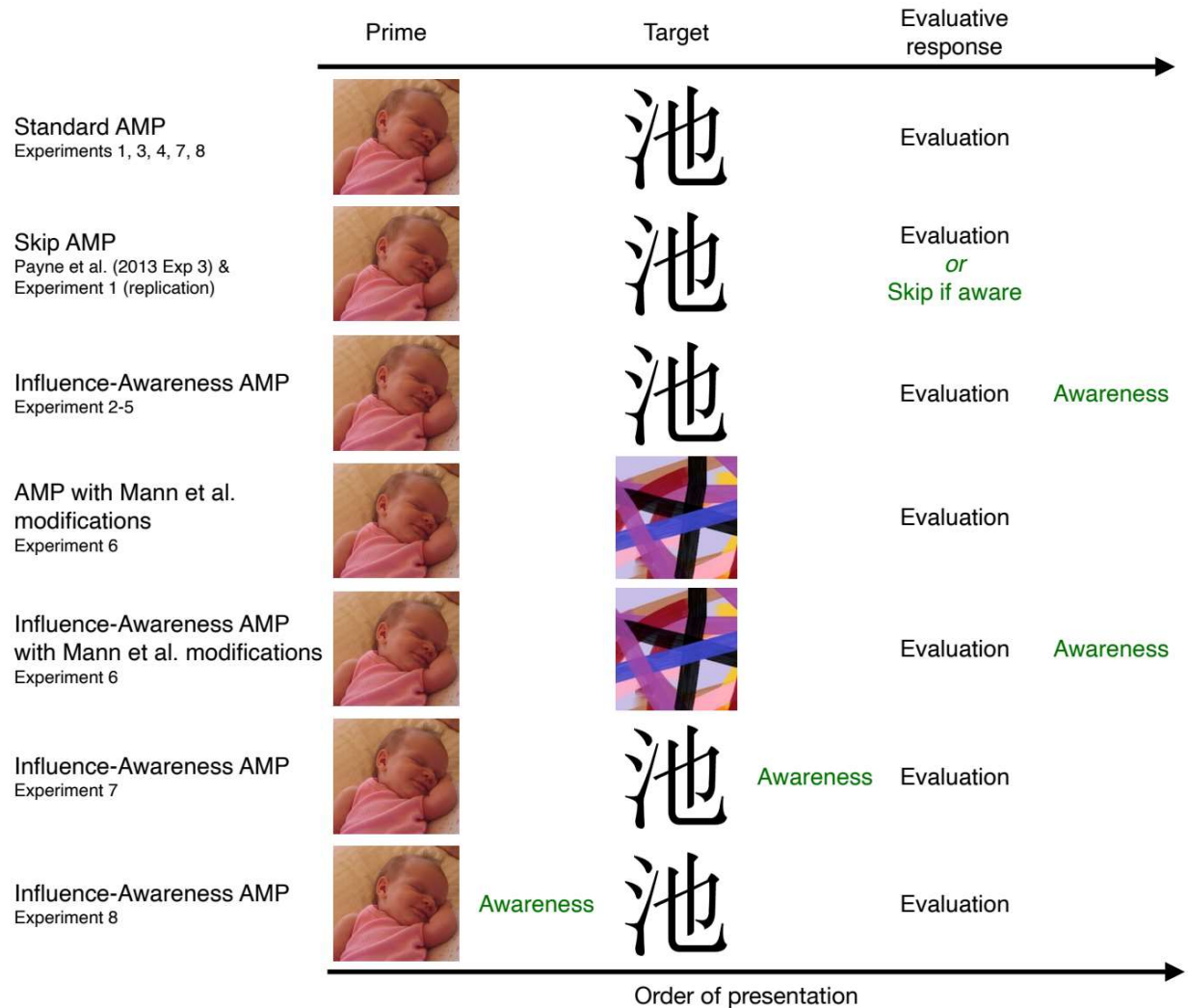


Figure 1. A schematic overview of the AMP variants used in Experiments 1-8. The point within the trial at which awareness of influence of the prime on evaluations was systematically varied between studies.

### Experiment 1: Skip AMP effects Differ from Standard AMP effects

In their original paper, Payne et al. 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.” (p.382). In other words, if awareness of influence of the prime is central to AMP effects then effects on the standard version of the task should be larger than effects on the skip variant. The authors found no statistically significant difference between the two task variants and concluded that “these data contradict the idea that

participants were aware that the primes influenced them before responding” (p.383).

In Experiment 1 we examined if this claim (that ‘skip’ AMP effects do not differ from standard AMP effects) is replicable. We first carried out power analyses (detailed below) to ensure that we had sufficient power to detect even small effects (something that may have presented a problem in the original study).<sup>4</sup> We then administered similar standard and

<sup>4</sup> The sample size used in the authors original study was relatively small ( $N = 36$  per cell). Although they argued that this sample would provide power “greater than .85 to detect an interaction between the repeated measures and between-subjects factors, even assuming a small effect size” (p. 383), they did not specify what effect size they qualified as “small”, nor what they specified as the correlation between the different within-subject measurements (i.e., the correlation between evaluations of the positive and negative primes). This unspecified correlation can make such a power analysis vary widely. Combining this with the fact that their sample size was relatively small, it is very possible that a true difference between the

skip-AMPs as used by Payne et al. (2013). Whereas the original authors relied on a between- subjects manipulation we opted to administer both AMP variants to all participants in order to improve our statistical power, as well as to better compare effect sizes within rather than between individuals. We also carried out statistical comparisons that allowed us to test the original key claim (e.g., we used a partially-overlapping *t*-test to account for participants who skip either all trials or no trials). In addition to the confirmatory analysis, we also asked an exploratory question: is one's awareness of the prime's influence on their evaluations (as indexed by skip rates in the skip AMP) related to the magnitude of their effect in a standard AMP? If so, then this would suggest that influence awareness may play more of a role in standard AMP effects than previously thought.

## Method

Materials for all experiments can be found at [osf.io/gv7cm](https://osf.io/gv7cm). This includes details of the designs, experimental scripts, raw and processed data, preregistrations, analytic plans, and all R code for data processing and analyses. We also report how sample sizes, data exclusions (if any), manipulations, and measures were determined in each study. We report only the key effects that serve to test our hypotheses. All other results of the models can be found on OSF.

## Sample Selection Strategy

Power analyses indicated that 147 participants would be required to detect a Cohen's *d* effect size of 0.3 in a paired-samples *t*-test at the conventional alpha level (.05, two-sided) with 95% power. 289 participants would be required to detect such an effect size in a two-sample *t*-test with otherwise identical parameters. Given that a partially-overlapping *t*-test's power typically falls somewhere between a paired-samples *t*-test and a two-sample *t*-test (Derrick, Toher & White, 2017), 289 participants would provide (a) at least 95% power to detect a small effect size in such a test and (b) power to detect a very small Cohen's *f* effect size (i.e., 0.045) in a linear regression with one dependent variable and one independent variable (i.e., the analysis used to investigate our second question).

## Participants and Design

316 individuals were recruited via Prolific (prolific.co) and took part in exchange for a monetary reward. We initially recruited 290 participants, but a number provided incomplete or partial data or did not meet our preregistered exclusion criteria. Recruitment was continued in batches of 10 until analyzable data was available for at least 290 individuals (final  $n = 295$ ; 160 men), who ranged in age from 18 to 61 ( $M = 29.8$ ,  $SD = 10.3$ ). A 2 (*Task Type*: standard vs. 'skip' AMP) x

2 (*Prime Type*: positive vs. negative) design was employed with both factors manipulated within participants. Ratings of the target stimuli (positive and negative images) served as the dependent variable.

## Ethical Approval

Approval for all studies was provided by the Ethical Committee of the Faculty of Psychology and Educational Sciences at Ghent University (approval numbers 2015/13, 2016/63, and 2016/80).

## Materials

Materials were programmed in Inquisit 4.0 and administered via the Inquisit Web Player. Both versions of the AMP contained three types of stimuli: primes, targets, and a mask. Prime stimuli consisted of 12 positive and 12 negative images taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997). Target stimuli consisted of 72 Chinese pictographs and the mask consisted of a white noise image (see [osf.io/p6e3c](https://osf.io/p6e3c) for the valenced items and Chinese pictographs used in this study).

## Procedure

Participants initially provided informed consent and demographic information (age and gender) and then completed a standard AMP followed by a skip-AMP.

**Standard AMP.** Prior to the task participants were told that the study would examine how people make simple judgements. Pictures would appear one after another on the screen. The first would be a real-life image and the second a Chinese symbol. Their task was to judge the visual pleasantness of the Chinese symbol using the E (pleasant) and I (unpleasant) keys while trying their best to not let the real-life images bias those judgements. Overall, the task consisted of 10 practice trials followed by 72 test trials. Each trial began with the presentation of a fixation cross for 500ms, after which a positive or negative prime stimulus was presented for 100ms, followed a blank screen for 100ms and then a target image (i.e., a Chinese pictograph). The target remained onscreen for 100ms, after which, a white noise image appeared and replaced the Chinese pictograph. This mask remained onscreen until the target stimulus was rated as positive or negative using the E or I keys respectively.

**Skip AMP.** The skip AMP was similar to the standard AMP. Participants were informed that they would complete a similar task once again but with one important difference: that they could now respond in a third way - namely - to 'skip' a trial by pressing the spacebar whenever they thought their evaluations of the pictographs might be influenced by the prime. Instructions emphasized that they should *only* evaluate the pictograph whenever their opinion reflected the qualities of the pictograph itself. If their judgement was influenced by the prime then they should skip that trial.

---

AMP types exists, but the authors simply did not achieve sufficient power to detect such a difference.

The AMPs used in our conceptual replication were similar to those used by Payne et al. (2013) with two exceptions. First we use 72 rather than 120 test trials in order to make online completion of two AMPs manageable for participants. Second, whereas Payne et al. used valenced and neutral primes we only used valenced primes as - in the majority of AMP studies - only valenced primes are used.<sup>5</sup>

## Results

### Data Exclusion

Participants were excluded if they completed the experiment too quickly (i.e., in under three minutes) or provided incomplete data on any of the measures ( $n = 21$ ).

### Data Preparation

AMP effects were computed by subtracting the proportion of 'pleasant' responses emitted on trials with an unpleasant prime from the proportion of 'pleasant' responses emitted on trials with a pleasant prime (Payne et al., 2005). Scores were calculated from all trials in the standard AMP and exclusively from the non-skipped trials in the skip-AMP.

### Analytic Strategy

We carried out a partially-overlapping  $t$ -test to examine our first question (i.e., do AMP effects differ as a function of *Task Type* [standard vs. skip]). We opted for this test for the following reason: it may be that some participants recognized the prime's influence on their evaluations and therefore skipped all trials in the skip-AMP. If so, then these individuals produced no AMP scores on this version of the task. One could simply exclude such participants in order to run a paired-sampled  $t$ -test between skip and standard AMP scores. Yet skip-AMP effects are not missing at random and are instead missing for a very important reason (i.e., people are highly influenced). Excluding such individuals would undermine the inferences we ultimately want to make.

The partially overlapping  $t$ -test is a variant of the  $t$ -test which overcomes this issue (Derrick et al., 2017). It is neither a dependent nor independent  $t$ -test but rather a mixed  $t$ -test containing independent and dependent data. Given that participants with an influence awareness rate close to 100% had no skip AMP effect and therefore had no data for the skip AMP, their standard AMP effects were entered as independent data. Those with standard *and* skip AMP effects were entered as dependent data.

We also investigated a second preregistered exploratory question: whether influence awareness rates

in the skip AMP predict the magnitude of effects in the standard AMP. To answer this question, we carried out a linear regression analysis with rate of skipping in the skip-AMP as the independent variable and effects in the standard AMP as the dependent variable. All reported analyses were preregistered in this and all other studies unless otherwise noted.

## Hypothesis Testing

### Do Effects On The Standard AMP Differ From Those On The Skip-AMP?

Results indicated that effects in the standard AMP ( $M = 0.17$ ,  $SD = 0.33$ ) were significantly different from that of the skip-AMP ( $M = -0.12$ ,  $SD = 0.26$ ),  $t(294) = 9.36$ ,  $p < .001$ ,  $M_{diff} = 0.29$ . A between-subjects Cohen's  $d$  was also calculated for familiarity, although this should be interpreted with caution as it does not acknowledge the partial dependence within the data:  $d = 0.96$ , 95% CI [0.79, 1.13].

### Does The Rate of Skipping in The Skip AMP Predict Standard AMP Effects?

Regression analyses indicated that rates of skipping in the skip AMP predicted the magnitude of effects in the standard AMP,  $B = 0.30$ , 95% CI [0.07, 0.53],  $\beta = .15$ , 95% CI [0.04; 0.26],  $p = .010$ .

## Discussion

We did not replicate the finding that standard and skip AMP effects are similar to one another. Whereas Payne et al. (2013; Experiment 3) found no difference between a skip- and standard AMP we found a difference between the two, such that scores on the standard AMP were significantly larger than those on the skip AMP (which were also in the opposite direction to what the original authors found). Thus it seems that people's subjective experiences during the skip-AMP were well-calibrated to the actual influence of the primes on their responses, and this allowed them to significantly reduce the priming effect. Even more interestingly, we found that a given participant's awareness of the prime's influence on their evaluations (during the skip-AMP at Time 2) strongly predicted the magnitude of their effects in the standard AMP at Time 1.<sup>6</sup> This suggests that awareness of the prime's influence on evaluations may play a role in the standard AMP as well.

<sup>6</sup> Throughout this article we employ the word 'predict' in its statistical sense, i.e., the estimate scores on one task based on scores on another task. Our claims are agnostic to the temporal order of the tasks, which are based exclusively on the statistical relationship among scores on the tasks. Indeed, the temporal order of the tasks in our experiments were specifically chosen so that participants complete a standard and unaltered AMP before other tasks so as to exclude the possibility that performance on the standard AMP was altered or perturbed in any way by the other tasks. Elsewhere, when the dependent variable is completed prior to the independent variable, this is occasionally referred to as 'postdiction' rather than 'prediction', however we employ the latter for familiarity.

<sup>5</sup> As we previously noted, we adopted a within participant design (participants first completed a standard AMP followed by a 'skip' AMP) whereas Payne et al. (2013) adopted a between participant design (participants completed *either* a standard AMP or a 'skip' AMP).

## **Experiment 2: AMP Effects are Strongly Related to Awareness of the Prime's Influence on Evaluations**

Although the skip-AMP is an improvement on post-hoc awareness measures it is not without its issues. By forcing people to either skip or evaluate the target stimulus the task can only provide partial data (i.e., either evaluations or indications of influence). A better task would be one where participants provide their evaluation of the target *and* indicate if that evaluation was influenced by the prime. Such a task would provide evaluative and prime-influence information for every participant on all trials and enable performance on the influence aware trials to be directly compared to performance on the non-influence aware trials, within rather than between participants. With this information one can then quantify what contribution influence aware vs. non-influence aware trials make to AMP effects.

With this in mind, we designed a task known as the Influence-Awareness AMP or IA-AMP to investigate the following questions. First, would we observe an AMP effect for generic valenced primes? Second, and at the trial-by-trial level, are those effects moderated by a subset of trials, namely trials in which a participant is aware of the prime's influence on their evaluations? Third, are AMP effects at the group level moderated by inter-individual differences in awareness of influence of the primes? Fourth, does the "on-line" measure of awareness provided by the IA-AMP correlate with the post-hoc self-report awareness measures typically used in this literature? Finally, does influence awareness on the IA-AMP predict AMP effects better than post-hoc self-report measures?

### **Method**

#### **Sample Selection Strategy**

Based on power analyses (i.e., 95% power to observe a medium effect size [ $f^2 = 0.15$ ] in a linear regression analysis with a single predictor at the 0.05 alpha level), our *a priori* sample size after exclusions was 150 participants. Our sampling strategy involved recruiting 150 participants and then excluding those with incomplete data or who failed to meet inclusion criteria. Recruitment continued in batches of 10 until analyzable data was available for at least 150 participants.

#### **Participants and Design**

In total 214 participants took part, and of those, 147 (82 female) ranging in age from 18 to 65 years ( $M$  age = 34.9,  $SD = 11.7$ ) provided complete and analyzable data. A single factor (*Prime Valence*: positive vs. negative) was manipulated within participants, and two dependent variables were

assessed: influence awareness ratings and target stimulus evaluations on each trial.

### **Materials**

The IA-AMP consisted of 12 positively and 12 negatively valence IAPS images (primes), a random selection of 120 of 200 possible Chinese pictographs (targets), and a white noise image (mask).

### **Procedure**

Participants first provided informed consent and demographic information (age and gender), and then completed an IA-AMP. Thereafter they completed a post-hoc self-report measure of awareness and exploratory questions.

**IA-AMP.** Prior to the task participants received a similar set of instructions as in the skip-AMP outlined in Experiment 1. These instructions emphasized that the first image (prime) that appears onscreen can sometimes bias judgements of the Chinese pictograph that follows, and that they should try their absolute best not to let this happen. After each trial they would be able to indicate if their judgement of the Chinese pictograph had been influenced by the first image that appeared by pressing the spacebar. If not, then they simply had to wait for the following trial.

The IA-AMP consisted of 10 practice trials followed by 120 critical trials with similar parameters to the AMPs used in Experiment 1 (i.e., a fixation cross for 500ms, a valenced prime stimulus for 100ms, followed by a blank screen for 100ms, a target Chinese pictograph for 100ms, after which, a white noise image appeared and replaced that pictograph).

At the end of each trial, participants were given the opportunity to press the spacebar to indicate that their evaluation had been influenced by the prime on that trial. Specifically, a cue to "press spacebar if you felt you were influenced by the picture" was presented after each evaluation was made. This cue remained onscreen for 2000ms during which the spacebar could be pressed, followed by a 200ms inter-trial interval and the next trial (note: this response window was fixed regardless of whether a response was emitted or not).

**Self-Report Awareness Measure.** This measure was identical to that used in Payne et al. (2013; Experiment 1) and asked participants: "to what extent were your ratings of the Chinese symbols influenced by the pictures that appeared immediately before those symbols?". They could respond using on a 7-point Likert scale ranging from "Never" to "Almost always".<sup>7</sup>

**Exploratory Questions.** Three exploratory questions were asked. The first centered on *prime*

<sup>7</sup> Note that in the interest of clarity we referred to this self-reported measure of awareness as "General Influence" in our pre-registration document and as an offline measure of awareness in the analytic section of that same document.



awareness (“Think back to the task you just completed. On how many trials was a valenced picture presented before the Chinese character? It is important that you are honest here.”) using a scale ranging from 1 (None) to 7 (All). The second indexed *intentional target stimulus responding* (“How often did you intentionally base your rating of the Chinese symbol on the image that immediately appeared before it?”). The third centered on *unintentional target stimulus responding* (“How often do you think that your rating of the Chinese symbol was unintentionally influenced by the pictures that appeared immediately before those symbols?”). Both questions were responded to using a scale ranging from 1 (Never) to 7 (Almost always). Finally, we asked participants if we should exclude their data for any reason (e.g., computer malfunction or a distraction during the study). These items were exploratory in nature were not part of our preregistered analyses.

## Results

### Data Preparation

For analyses at the trial-by-trial level we computed IA-AMP effects using the standard method outlined in Experiment 1. Given our interest in the *magnitude* of IA-AMP effects, regardless of their directionality, all analyses on trial-by-trial level effects examined absolute values (i.e., the difference in evaluations between the prime types, agnostic to the direction of the effect). We also calculated influence rates for each participant in the IA-AMP by dividing the number of trials where they reported being influenced by the prime (i.e., by pressing the spacebar) by the total number of trials in the IA-AMP.<sup>8</sup>

### Analytic Strategy

A logistic mixed-effects model was used to investigate our first question (is there evidence for an AMP effect) and second question (at the trial-by-trial level, is prime-consistent evaluation moderated by influence awareness). We opted for mixed-effects models given that they offer superior statistical power compared to more commonly used fixed-effects alternatives. To address our third question (are AMP effects at the group level moderated by those participants who are more highly aware of the prime’s influence) we scored IA-AMP effects for each participant (*see below*) and entered these into linear regression models. To compare online (IA-AMP) and offline (self-report) measures of influence awareness, correlational and regression analyses were used.

### Hypothesis Testing

<sup>8</sup> In our preregistration we incorrectly specified this as the ‘proportion of influenced to non-influenced trials’ (i.e., a ratio rather than a proportion). We opted to treat influenced trials as a proportion of the total number of trials.

**Was There Evidence For An AMP Effect?** A logistic mixed-effects model was carried out, with Valence Ratings (pleasant or unpleasant) of the target stimulus on each trial as the dependent variable, Prime Valence (pleasant or unpleasant) as the independent variable, and Participant as a random effect. This model acknowledges the non-independence of the multiple data points provided by each participant, trial type, and prime identity (i.e., the hierarchical nature of the data). Specifically, we modelled trial type as a random slope nested within a random intercept for participant, as well as modelling prime identity as a random intercept. Overall, an AMP effect emerged, such that participants were more likely to rate the target stimulus as positive when the prime valence was positive compared to when the prime valence was negative (and vice versa),  $OR = 3.69$ , 95% CI [2.81, 4.85],  $p < .001$ . The mean and standard deviations for standard and IA-AMP effects can be found in Tables 1 and 2 respectively.

**Are AMP Effects Moderated By Influence Awareness At The Trial Level?** We extended the above model by adding influence awareness on each trial (influence aware vs. non-influence aware) as a fixed effect. This allowed us to determine if the relationship between Valence Rating and Prime Valence was moderated by that subset of influence aware trials. Results revealed an interaction between Prime Valence and influence awareness, such that AMP effects were far stronger on influence aware trials,  $OR = 31.59$ , 95% CI [24.79, 40.28],  $p < .001$ .

**Are AMP Effects Moderated By Those Participants Who Are More Influence Aware?** We then sought to determine if AMP effects were moderated by those participants who were more frequently aware of the prime’s influence on their evaluations (i.e., whether awareness rates varied between individuals and whether this variation was associated with the magnitude of the AMP effect). An ‘awareness rate’ score was calculated for each participant by dividing the number of ‘aware’ trials by the total number of trials completed (i.e., 120). A linear regression analysis with AMP effect size as the dependent variable and influence awareness rate as a predictor variable was then conducted. Results indicated that influence awareness rate was a significant predictor of AMP effect size,  $B = 0.42$ , 95% CI [0.31, 0.53],  $\beta = 0.53$ , 95% CI [0.39, 0.57],  $p < .001$ .

**Do Online (IA-AMP) And Post-Hoc (Self-Report) Measures Of Awareness Correlate With One Another?** Simple correlations revealed that the IA-AMP and post-hoc awareness measures strongly associated with one another,  $B = 0.14$ , 95% CI [0.13, 0.16],  $\beta = 0.83$ , 95% CI [0.73, 0.92],  $p < .001$ .

### **Does Influence Awareness On The IA-AMP Predict AMP Effects Better Than Post-Hoc Self-Report Measures?**

Regression analyses were once again conducted with the two awareness measures added into the model. This allowed us to determine their relative contribution in predicting AMP effects. Results indicated that only awareness assessed during the IA-AMP task predicted AMP effect sizes,  $B = 0.38$ , 95% CI [0.15, 0.61],  $\beta = 0.42$ , 95% CI [0.17, 0.67],  $p = .001$ ; whereas awareness assessed after the task (post-hoc self-report) did not,  $B = 0.02$ , 95% CI [-0.02, 0.06],  $\beta = 0.12$ , 95% CI [-0.13, 0.37],  $p = .341$ ). Comparison of the beta estimates' confidence intervals indicated that assessing for awareness during the task was a significantly better predictor of effects than doing so afterwards.

### **Discussion**

The results from Experiment 2 are in-line with our preregistered hypotheses. AMP effects emerged and were moderated *at the trial-by-trial level* by performance on a subset of trials – namely – those where a participant was aware of the prime's influence on their evaluations. *At the group level*, AMP effects were moderated by participants who were highly influence aware and were significantly larger when calculated on the basis of the influence aware trials compared to when calculated on the basis of the non-influence aware trials. The online measure of awareness was a superior predictor of AMP effects than the offline measure.

### **Experiment 3: Awareness Assessed During an IA-AMP Predicts The Magnitude of Effects on a Previously-Completed Standard AMP (in the Same Attitude Domain)**

One question that comes to mind is how performance on the IA-AMP relates to performance on a standard AMP. It may be that asking about influence on every trial serves to artificially raise awareness of the prime as well as its influence on evaluations. This in turn may lead to a stronger relationship between awareness and AMP effects than would normally occur in the standard version of the task. Therefore, in Experiment 3 we sought to not only replicate our previous findings but also address this new question. Participants were asked to complete a standard AMP with generic valenced primes followed by an IA-AMP with similar stimuli (i.e., both task variants indexed attitudes from the same domain). This approach provided us with a baseline AMP effect for each participant that was unperturbed by awareness probes as well as a separate influence awareness measure from that same person. If influence awareness rates on an IA-AMP completed at Time 2 correlate with standard AMP effects obtained at Time 1 then this would suggest that

influence awareness may also be central to the standard AMP as well.

In short, Experiment 3 had both confirmatory and exploratory goals (all of which were preregistered). On the one hand we sought to confirm our earlier findings (i.e., that AMP effects would emerge on the IA-AMP; that these effects would be moderated by performance on the influence aware trials within a given individual, while at the group level, be moderated by highly aware participants). On the other hand, we also set out to explore the aforementioned question - would effects in the standard AMP be predicted by influence awareness rates in the IA-AMP? Based on our previous findings we predicted they would be.

We then explored a third and final question: would there be a difference in the magnitude of standard AMP effects relative to IA-AMP effects that are exclusively comprised of 'non-influence aware' trials? Recall that Payne et al. (2013; Experiment 3) asked a similar question when they compared scores on a standard AMP to those on a 'skip' AMP which was argued to provide a non-influence aware measure of evaluations. Unlike those authors (who found no difference between the two measures) we predicted that standard AMP effects would be significantly larger than those obtained from an IA-AMP comprised of exclusively "non-influence aware" trials, thus providing further support for the idea that influence awareness plays a role in standard AMP effects as well.

### **Method**

#### **Sample Selection Strategy**

Based on power analyses using identical criteria as Experiment 2, our *a priori* required sample size after exclusions was 150 participants.<sup>9</sup>

#### **Participants and Design**

206 participants took part, and of those, 176 (102 women) ranging in age from 18 to 64 years ( $M = 33.60$ ,  $SD = 11.45$ ) provided complete and analyzable data. A 2 (*Task Type*; standard vs. IA-AMP)  $\times$  2 (*Prime Type*; positive vs. negative) design was employed with both factors manipulated within participants. Two dependent variables were assessed: target stimulus evaluations and influence awareness responses.

#### **Materials**

AMP stimuli were similar to those used in Experiments 1 and 2 (i.e., 12 positively and 12 negatively valenced IAPS images [primes], and 72 Chinese pictographs [targets]).

#### **Procedure**

<sup>9</sup> We should note that more participants were sampled than originally specified in our preregistration due to an error in how exclusions were originally implemented in our data processing R script. Data collection was stopped when we believed we had 150 participants, as per the preregistration. A code review revealed that some participants had been erroneously excluded. The final analytic sample therefore includes these participants.

Participants first provided informed consent and demographic information (age and gender), and then completed a standard AMP, IA-AMP, the post-hoc self-reported awareness measure, and exploratory questions.

**AMPs.** Two version of the task were employed in Experiment 3: a standard AMP (similar to that used in Experiment 1) and an IA-AMP (similar to that used in Experiment 2). Both consisted of 10 practice and 72 test trials with the same trial parameters as reported in Experiment 2. Participants first encountered the standard AMP and then encountered the IA-AMP. This was the same general procedure as used in Experiment 1.

**Self-Report Measures.** The same set of self-reported awareness and exploratory questions were asked as in Experiment 2.

## Results

### Analytic Strategy

A linear regression model was used to examine our confirmatory questions (i.e., if AMP effects would emerge on the IA-AMP; if these effects would be moderated by performance on the influence awareness trials within a given individual, and be moderated by those participants who are more influence aware at the group level). A similar model was used to examine our first exploratory question (i.e., if influence awareness rates on the IA-AMP predict effect sizes in a previously completed standard AMP). A paired-samples *t*-test was used to investigate our second exploratory question (i.e., for differences between standard AMP effect sizes and non-influence aware only AMP effect sizes).

### Data Preparation

Three AMP scores were calculated for each participant: an overall effect for the standard task, an overall effect for the IA-AMP, and a 'non-influence aware' effect based on those trials from the IA-AMP where participants did not press the spacebar (i.e., did not indicate awareness of the prime and its influence on their evaluations). This score notionally reflects an AMP effect based exclusively on non-influenced trials.

### Hypothesis Testing

Was There Evidence For An AMP Effect And Was This Effect Moderated By Influence Awareness Within Individuals And At The Group Level? A significant effect emerged in both the standard AMP (OR = 3.81, 95% CI [2.86, 5.08],  $p < .001$ ) and IA-AMP (OR = 5.94, 95% CI [4.41, 7.99],  $p < .001$ ). At the trial-by-trial level, IA-AMP effects were moderated by influence aware trials, OR = 33.70, 95% CI [25.67, 44.23],  $p < .001$ . At the group level, IA-AMP effects were predicted by the influence awareness rates of participants,  $B = 0.44$ , 95% CI [0.34, 0.54],  $\beta = 0.56$ , 95% CI [0.44, 0.68],  $p < .001$ .

### Does Influence Awareness On An IA-AMP Completed At Time 2 Predict Standard AMP Effects At Time 1?

A regression analysis was conducted with standard AMP effect sizes as a dependent variable and influence awareness rate in the IA-AMP as a predictor variable. Results indicated that influence awareness rates in the IA-AMP predicted the magnitude of effects in the standard AMP,  $B = 0.34$ , 95% CI [0.24, 0.45],  $\beta = 0.44$ , 95% CI [0.30, 0.57],  $p < .001$ .

### Is There A Difference In The Magnitude Of Standard AMP Effects And Those Based Exclusively On Non-Influenced Trials?

Results from a partially overlapping *t*-test (see Experiment 1) indicated that AMP effects based exclusively on non-influenced trials were significantly smaller than effects in the standard AMP,  $t(163.85) = 5.09$ ,  $p < .001$ ,  $M_{diff} = 0.14$ . A between-subjects Cohen's  $d$  was also calculated, although this should be interpreted with caution as it does not acknowledge the partial dependence among the data,  $d = 0.41$ , 95% CI [0.19, 0.63].

## Discussion

Our findings replicated: AMP effects emerged and were moderated at the trial-by-trial level by performance on a subset of (influenced) trials, and at the group level by highly influence aware participants. We also extended these findings by showing that influence awareness during an IA-AMP at Time 2 predicted the size of standard AMP effects completed at Time 1. This suggests that asking about influence awareness of a trial-by-trial basis does not artificially raise awareness of the prime and its influence on evaluations. If it did then we would have expected no relationship between influence awareness rates and standard AMP effects to emerge, especially given that the IA-AMP was completed *after* the standard AMP. Yet influence awareness rates were strongly predictive of standard AMP effects, suggesting that people may in fact be aware of the prime, and use that stimulus when forming an evaluation of the target.

Finally, and in line with Experiment 1, we obtained further evidence that conflicts with the idea that standard and non-influenced AMP effects do not differ from one another. Results indicated that AMP effects exclusively generated from non-influenced trials were significantly smaller than standard AMP effects. These findings are consistent with what we observed in our direct replication attempt in Experiment 1 and converge on the same conclusion: AMP effects are stronger when participants are aware of the prime and its influence on evaluations.

#### **Experiment 4: Awareness Assessed During an IA-AMP Predicts the Magnitude of Effects on a Previously-Completed Standard AMP (in a Different Attitude Domain)**

Experiment 4 represents an even stronger test of our claims. Imagine if participants first complete a standard AMP in one domain (political attitudes) and then complete an IA-AMP in a completely different domain (attitudes towards generic valenced stimuli). Now imagine if the same pattern of findings once again emerges. This would mean that a given participant's influence awareness rates at Time 2 in one domain would be predicting the magnitude of their AMP effects at Time 1 in an entirely different domain. If so, then this would provide even stronger evidence that influence awareness is stable within individuals and moderates their task performance across AMPs in different attitude domains.

With this in mind, we employed a similar design to Experiment 3 but with one change: we varied the attitude domains being assessed by the standard AMP (political attitudes towards Donald Trump vs. Barack Obama) and the IA-AMP (attitudes towards generic valenced stimuli as in Experiments 1-3). If influence awareness rates reflect a stable (within-participant) pattern of responding regardless of content domain (politics vs. generic valenced primes), then influence awareness rates in a positive/negative IA-AMP at Time 2 should still predict effect sizes within a standard political AMP completed at Time 1.

#### **Method**

##### **Sample Selection Strategy**

Power analyses criteria were identical to Experiments 2 and 3 with an *a priori* required sample size after exclusions of 150 participants.

##### **Participants and Design**

Given that we were interested in assessing political attitudes we recruited a sample of residents from the USA who politically identified as Democrats. 175 participants took part in the study with data from 142 (74 women) ranging in age from 18 to 62 years ( $M = 31.90$ ,  $SD = 10.41$ ) eligible for analysis. The design, independent, and dependent variables were similar to those in Experiment 3.

##### **Materials**

The standard and IA-AMPs used in this study were similar to those used in Experiment 3 (i.e., participants encountered 10 practice and 72 test trials with the same trial parameters, instructions, and target stimuli as before). Critically, a new set of prime stimuli were used consisting of six images of Barack Obama and six images of Donald Trump that were taken from the Presidents-IAT materials of the Project Implicit website (see [osf.io/f38ag](https://osf.io/f38ag)).

#### **Procedure**

The procedure was similar to Experiment 3 with two exceptions: (a) the standard AMP now assessed political attitudes while the IA-AMP assessed generic attitudes towards valenced stimuli, and (b) two new exploratory questions were added.

**Exploratory Questions.** In addition to the same exploratory questions asked in Experiments 1-3 we also assessed for demand compliance and political alignment. *Demand compliance* was assessed using the following question: "Think back to the task with the Chinese characters. During the task, we asked you after each trial to indicate whether your response to the Chinese character was influenced by the image that appeared before it. Please choose the following option that is the most true for you". Response options were as follows: (1 = My responses were based on what I thought the researcher wanted me to say; 2 = My responses were genuinely based on whether I was influenced or not; 3 = I'm not sure or don't know why I responses the way I did). *Political alignment* was assessed using the following question: "In terms of the two major political parties in the US, do you consider yourself more Democratic or Republican?". Response options ranged from 1 (Very Republican) to 7 (Very Democratic).

#### **Results**

##### **Analytic Strategy**

For this and all subsequent experiments, we divide results into two sections. 'Replication hypotheses' refer to hypotheses that were first made in one of our previous experiments and which were retested in the current experiment. These will be briefly reported given that a detailed treatment is provided in a previous experiment. 'Critical hypotheses' refer to new hypotheses being made within a given experiment.

##### **Data Preparation**

Data preparation was similar to that of Experiment 3.

##### **Hypothesis Testing**

**Replicated Hypotheses: Was There Evidence For An AMP Effect And Was This Effect Moderated By Influence Awareness Within Individuals And At The Group Level?** A significant effect emerged on the IA-AMP,  $OR = 3.44$ , 95% CI [2.69, 4.39],  $p < .001$ , and standard AMP,  $OR = 6.31$ , 95% CI [4.08, 9.77],  $p < .001$ . At the trial-by-trial level, effects were moderated by the subset of trials where a participant was influence aware,  $OR = 67.46$ , 95% CI [49.70, 91.56],  $p < .001$ ; while at the group level, effect sizes were moderated by inter-individual differences in influence awareness, both within the IA-AMP ( $B = 0.57$ , 95% CI [0.40, 0.74],  $\beta = 0.49$ , 95% CI [.34, 0.63],  $p < .001$ ) and for a previously completed AMP ( $B =$

0.54, 95% CI [0.44, 0.65],  $\beta = 0.65$ , 95% CI [0.53, 0.78],  $p < .001$ ).

**Critical Hypotheses:** Does Influence Awareness On A Generic Valence IA-AMP Completed At Time 2 Predict Political AMP Effects At Time 1? A regression analysis with awareness rate in the IA-AMP as a predictor and standard AMP effect size as a dependent variable revealed that the former significantly predicted the latter,  $B = 0.54$ , 95% CI [0.44, 0.65],  $\beta = 0.65$ , 95% CI [0.53, 0.78],  $p < .001$ .

## Discussion

Our prior findings once again replicated: AMP effects emerged and were moderated at the trial-by-trial level by performance on influence aware trials, while at the group level, they were moderated by participants scoring high in influence awareness. Not only did influence awareness assessed by an IA-AMP retrospectively predict standard AMP effects, but it did so even when these tasks were assessing attitudes in entirely different domains. Taken together, Experiments 3 and 4 suggest that influence awareness is stable within individuals and moderates the magnitude of the AMP effect, even between attitude domains. It may be the case that, at least at the group level, the majority of the variance in AMP effects does not represent a measure of implicit evaluations *in general*, but rather the evaluations of a subset of individuals who are highly influence aware. We will return to this idea later on.

### Experiment 5: The AMP's Predictive Utility & Influence Awareness Are Strongly Related

Experiments 2-4 indicate that a subset of influence aware trials strongly moderate AMP effects at the trial-by-trial level, and inter-individual differences in influence aware rates moderate the magnitude of the AMP effect. Influence awareness rates on one AMP predict how one will respond on another, and this is true when both tasks assess the same or different attitude domains.

In Experiment 5 we set out to further replicate our findings while addressing three new questions. On the one hand, we wanted to know if influence awareness also played a key role in the AMP's predictive utility (i.e., its ability to discriminate between two known-groups). We therefore recruited two groups of participants (Democrats and Republicans) and asked them to complete a political IA-AMP followed by a positive/negative IA-AMP. Our preregistered hypothesis was that the AMP's ability to predict whether a person was a Democrat or Republican would be higher when effects were solely derived from influence aware trials and lower when they were derived from non-influence aware trials.

At the same time, we wanted to know if there was intra-individual stability in influence awareness from

one AMP to another. Experiments 2-3 offered indirect evidence such that influence awareness rates in the IA-AMP predicted the same person's scores in the standard AMP. However, a more direct demonstration requires that we have a measure of awareness on both tasks. We therefore examined if a participant's influence rate on one IA-AMP was correlated with their influence rate on a second IA-AMP.

Finally, a unidirectional relationship between influence awareness and AMP effect sizes emerged in Experiments 2-4. However, given that we were now administering two IA-AMPs, we could also now test for a bidirectional relationship between these variables (i.e., if influence rates in AMP #1 predict effects in AMP #2, and if influence rates from AMP #2 predict effects in AMP #1). Demonstrating such a relationship would suggest that *influence rates in general* are predictive of *AMP effects in general*, yet further evidence supporting the idea that AMP effects are highly related to awareness of the primes and its influence on one's evaluations.

## Method

### Sample Selection Strategy

Power analyses for interactions in mixed-effects models are difficult to determine due to the large increase in the number of parameters involved, therefore no power analysis was conducted for our first analysis. For our second analysis, we used the pwr package in R to compute the number of participants required to detect a medium  $f^2$  effect size (i.e., 0.15) in a regression analysis with a single IV, at the conventional alpha level (.05) and at 95% power. Given these criteria, 89 participants would be required. The aforementioned power analysis is also applicable for our third analysis. With 89 participants, at a standard alpha level and a power of .90, we would be able to detect a correlation of  $r = .33$ . We chose to collect data from at least 200 participants (100 Democrats and 100 Republicans) based on the availability of resources.

### Participants and Design

A total of 334 participants took part, and of these, 207 (105 Democrats, 102 Republicans; 106 women) ranging in age from 18 to 65 years ( $M = 34.03$ ,  $SD = 11.15$ ) provided complete and analyzable data. A 2 (*Content Type*; political vs. generic valence)  $\times$  2 (*Prime Type*; positive vs. negative)  $\times$  2 (*Political Orientation*; Democrat vs. Republican) design was employed with the first two factors manipulated within and the third manipulated between participants. Influence awareness rates and evaluations were the two dependent variables.

### Materials

Two IA-AMPs were employed. The first was an IA-AMP that used the same political (Obama and Trump) primes as used in the standard AMP in Experiment 4 while the second was a generic valence

IA-AMP identical to that used in our previous studies. All instructions, trial parameters, and other aspects of the tasks were similar to those described in Experiment 4.

### Procedure

Participants provided informed consent and demographic information. They then completed a politics IA-AMP, a generic valence IA-AMP, the post-hoc self-reported awareness measure, and exploratory questions. The latter (self-report) measures were identical to those reported in Experiment 4.

### Results

#### Analytic Strategy

Our replicated hypotheses were assessed using a similar analytic strategy as before. To examine the relationship between influence awareness and the AMP's predictive validity, two AMP scores were calculated, one based solely on the influence aware trials and another based solely on the non-influence aware trials. We then used two between-groups *t*-tests to examine their relative ability to discriminate between Democrats and Republicans. To examine the consistency of influence awareness rates within participants across different AMPs, a simple correlation test was used. Finally, two linear regression models were used to assess the bidirectional relationship between influence rates and AMP scores. In the first regression, influence awareness rate was taken from the politics IA-AMP, and the effect size from the positive-negative IA-AMP. In the second regression, influence awareness rate was taken from the positive-negative IA-AMP and effect sizes from the politics IA-AMP.<sup>10</sup>

#### Hypothesis Testing

Replicated Hypotheses: Was There Evidence For An AMP Effect And Was This Effect Moderated By Influence Awareness Within Individuals And At The Group Level? A significant effect emerged on the positive-negative IA-AMP,  $OR = 4.11$ , 95% CI [3.17, 5.33],  $p < .001$ . Effects also emerged on the political IA-AMP and in the opposite direction for Republicans and Democrats. Specifically, a significant interaction effect was obtained for Prime Type (Trump vs. Obama) and Political Orientation (Republicans vs. Democrats),  $OR = 0.04$ , 95% CI [0.02, 0.09],  $p < .001$ . At the trial-by-trial level, effects on both IA-AMPs were moderated by the subset of trials where a participant was influence aware (valence:  $OR = 58.58$ , 95% CI [43.07, 79.68],  $p < .001$ ; politics:  $OR = 272.73$ , 95% CI [155.85, 477.26],  $p < .001$ ) while at the group level, effect sizes on a given IA-AMP were moderated by inter-individual

differences in influence awareness on that task (valence:  $B = 0.49$ , 95% CI [0.40, 0.58],  $\beta = 0.61$ , 95% CI [0.50, 0.72],  $p < .001$ ; politics:  $B = 0.63$ , 95% CI [0.53, 0.74],  $\beta = 0.64$ , 95% CI [0.54, 0.75],  $p < .001$ ).

#### Critical Hypotheses: Does Influence Awareness Moderate the AMP's Predictive Validity?

Results indicated that the overall IA-AMP effect (i.e., an AMP including both types of trials) was  $d = 1.25$ , 95% CI [0.95, 1.55]. Effects based solely on influence aware trials were superior in discriminating between Democrats and Republicans ( $d = 2.08$ , 95% CI [1.62, 2.55]) than effects based solely on the non-influence aware trials ( $d = 0.62$ , 95% CI [0.33, 0.91]),  $Q(df = 1) = 27.51$ ,  $p < .001$  (see Figure 2).<sup>11</sup> Note that the confidence intervals of these three effects do not overlap, with influence-only AMP effects better at discriminating between individuals than an overall effect, which in turn was superior to a non-influence aware effect.

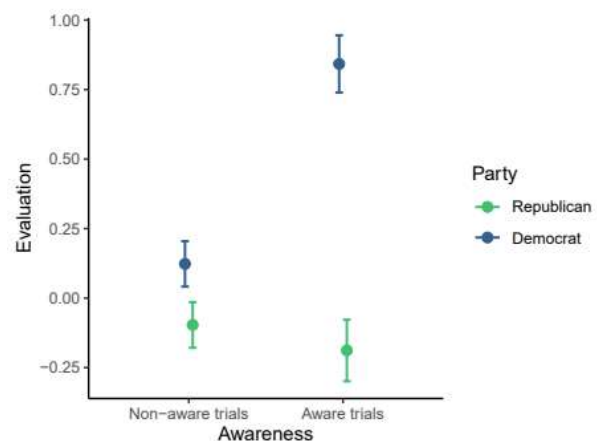


Figure 2. The political IA-AMP's ability to discriminate between Democrats and Republicans on the basis of influence aware and non-influence aware trials. A negative score indicates a preference for Trump over Obama whereas a positive score indicates a preference for Obama over Trump. Error bars represent 95% confidence intervals.

#### Are Influence Awareness Rates Consistent Within Individuals And Across Different AMPs?

Results revealed a strong correlation between influence awareness rates in the two task variants,  $r = 0.82$ , 95% CI [0.77, 0.86],  $p < .001$ .

<sup>10</sup> Note that in Experiments 2-4 we focused on the absolute magnitude of AMP effect sizes. Here in Experiment 5 we took the directionality of AMP effects into account when comparing the AMP scores of Republicans to those of Democrats. In all other cases, absolute AMP scores were assessed when testing hypotheses at the participant level.

<sup>11</sup> Our preregistration stated that we would compare differences between these conditions via the confidence intervals on the two Cohen's *d* estimates. We subsequently discovered a method to produce a *p* value for this comparison via the metafor package's heterogeneity test. Both are reported and results are congruent across these analytic choices.

### **Does Influence Awareness In One IA-AMP Predict Performance In Another IA-AMP, And Is This Relationship Bidirectional?**

Results indicated that influence awareness rates in the politics IA-AMP predicted scores in the positive-negative IA-AMP,  $B = 0.46$ , 95% CI [0.36, 0.55],  $\beta = 0.54$ , 95% CI [0.42, 0.66],  $p < .001$ , and that influence awareness rates in the positive-negative IA-AMP predicted scores in the politics IA-AMP,  $B = 0.49$ , 95% CI [0.38, 0.60],  $\beta = 0.52$ , 95% CI [0.40, 0.63],  $p < .001$ . It is also useful to consider the implications of these results in terms of temporal order rather than domain. Although it was not part of our original research plan, these results also suggest that the temporal order of the tasks, and therefore the order of assessment of the AMP effect versus the influence rate, does not matter. Participants always completed the politics IA-AMP first and the valence IA-AMP second. The influence rate in the politics IA-AMP (completed first) predicted the absolute magnitude of the valence IA-AMP (completed second),  $B = 0.46$ , 95% CI [0.36, 0.55]. Equally, the influence rate of the valence IA-AMP (completed second) predicted (or more accurately 'postdicted') the absolute magnitude of the politics IA-AMP (completed first),  $B = 0.49$ , 95% CI [0.38, 0.50]. The very similar estimates and strongly overlapping confidence intervals provide no evidence that order of presentation moderated the effect.

### **Discussion**

Experiment 5 offers three new insights into the relationship between influence awareness and AMP effects. First, the predictive validity of AMP effects based solely on influence aware trials is superior to that of effects based solely on non-influence aware trials. Second, a given person's influence awareness rate on one AMP is strongly correlated with their influence awareness on another AMP, even when those tasks are targeting entirely different domains. Third, the relationship between influence awareness and AMP scores is *bidirectional*, insofar as AMP effect sizes predict how influence aware one will later report being, and how influence aware one is at an earlier point in time will predict the later magnitude of their AMP effects.

In short, these findings provide yet further support for the idea that (a) AMP effects are moderated by a subset of influence aware trials and participants who are highly influence aware, and (b) that the influence aware participants who are mainly responsible for AMP effects in one domain are the same participants who are responsible for effects in another domain. They also imply that the AMP's predictive validity is strongly related to influence awareness. Although non-influence aware trials retain some degree of predictive validity,

this pales in comparison to that of influence aware trials.

### **Experiment 6: Performance on the Mann et al. AMP is Also Moderated by Influence**

#### **Awareness**

We are not the first to argue that AMP effects are strongly moderated by a well-defined subset of participants. In a recent review of the AMP literature, Mann et al. (2019) noted that data from AMP studies exhibit a strong bimodal distribution, with a subset of participants showing a very strong AMP effect, and the others producing scores that follow a normal distribution (for related arguments also see Bar-Anan & Nosek, 2012).

Mann et al. argued that this cluster of extreme scoring participants (i.e., those who responsible for the bimodality) represent a small group of *intentional* responders, whereas the remaining participants reflect *unintentional* responders. They sought to eliminate the contaminating influence of these intentional responders (and thus reduce this bimodality) by creating a new and improved variant of the AMP. This task employed visually stimulating paintings as target stimuli, rather than the less visually stimulating Chinese pictographs, in order to increase the chances that participants would pay attention to the target rather than prime. They also included additional instructions imploring participants to avoid intentionally responding to the prime while reassuring them that it was acceptable if they sometimes did so. Mann et al. concluded that their modifications to the AMP decreased bimodality compared to a standard AMP (and thus reflected a less intentional measure of evaluations).

In Experiment 6 we examined if awareness of influence of the prime is also reduced in the Mann et al. AMP (referred to hereafter as the 'Mann AMP'). We did so by replicating Experiment 3 using this new version of the task. That is, participants were first asked to complete a standard Mann AMP and then complete a version of that task where they could also indicate if they were aware of the prime and its influence on their evaluations (referred to as the 'Mann IA-AMP'). If the Mann AMP successfully limits or excludes influence aware trials and participants, then we should not expect to replicate our prior findings. However, if we do replicate those findings, then it would suggest that even this purportedly 'improved' version of the task is also heavily dependent on influence awareness.

Based on our findings to date, we preregistered two hypotheses. On the one hand, we argued that, at both the group- and trial-level, Mann IA-AMP effects would be strongly moderated by influence awareness. On the other hand, we hypothesized that influence awareness rates of a given participant in the Mann IA-AMP at

Time 2 would predict the size of that same person's Mann AMP effects completed at Time 1.

## **Method**

### **Sample Selection Strategy**

Power analyses began with an examination of the association between the IA-AMP influence awareness rates and absolute AMP effects observed in Experiment 2. Results from that study indicated that this association was in the range  $\beta = 0.56$ , 95% CI [0.44, 0.68]. However, we were unsure whether the Mann et al. modification to the AMP would impact the magnitude of this association compared to our previous studies. We therefore opted to power our analyses to detect an even smaller effect size (i.e.,  $\beta = .20$ ). To power a regression analysis to detect a  $\beta = .20$  at a 0.05 alpha level (two-tailed) with 95% power requires 320 participants. This was defined as our *a priori* sample size after exclusions and participants were sampled in a similar fashion to our previous experiments.

### **Participants**

410 participants took part, and of those, 330 (171 women) ranging in age from 18 to 65 ( $M = 33.40$ ,  $SD = 11.05$ ) provided complete and analyzable data.

### **Procedure**

A similar procedure to Experiment 2 was used with one exception: the standard AMP was replaced with Mann et al.'s AMP, and the IA-AMP was replaced with a Mann et al. variant of our IA-AMP.

**AMPs.** The standard Mann AMP consisted of 60 test trials, the same 12 positive and 12 negative valenced prime images as before, as well as 60 abstract paintings which served as target stimuli. The trial parameters were similar to those reported in our previous studies (i.e., a fixation cross for 500ms, a valenced prime stimulus for 100ms, followed by a blank screen for 100ms, a target painting for 100ms, after which, a white noise image appeared and replaced that painting).

This version of the AMP also provided participants with the following instructions: "It is crucial that you try to evaluate only the paintings, rather than the pictures that come before them. Even if you are trying to rate only the paintings, you may notice occasionally that your ratings of the paintings are consistent with how you would have rated the previous pictures. That is totally okay. You do not have to worry about whether your ratings of the paintings are consistent, or not, with how you feel about the pictures. Your only job is to evaluate the paintings, and we really need your help in doing this. Do not even worry about the pictures, or how your paintings judgements occasionally match the pictures. Your help to this research is really important. Our data depend on participants like you trying as hard as possible to follow instructions. We really need and appreciate your help."

The Mann IA-AMP consisted of similar parameters as those used in the standard Mann AMP. The same instructions were provided along with the following: "Remember: if you do feel like your response to a painting was influenced by the picture that came before it, make sure to press the spacebar when prompted to do so."

**Self-reported measures.** The same self-reported awareness measure and exploratory questions were asked as in Experiment 2.

## **Results**

### **Analytic Strategy**

Our analytic strategy was identical to that of Experiment 2.

### **Data Preparation**

Our data preparation was identical to that of Experiment 2.

### **Hypothesis Testing**

#### **Replication Hypotheses: Do We Find**

**Evidence For Mann IA-AMP Effects?** A significant effect emerged on both the Mann AMP,  $OR = 4.39$ , 95% CI [3.52, 5.48],  $p < .001$ , and the Mann IA-AMP,  $OR = 5.12$ , 95% CI [4.14, 6.34],  $p < .001$ .

**Critical Hypotheses: Does Influence Awareness Predict Mann IA-AMP Effects At The Trial Level And Trial-By-Trial Level?** Results revealed an interaction between influence awareness and Prime Type in the Mann IA-AMP,  $OR = 24.48$ , 95% CI [19.70, 30.43],  $p < .001$ , such that IA-AMP effects were moderated by influence aware trials. Results also indicated that influence awareness rates significantly predicted the magnitude of Mann IA-AMP effects,  $B = 0.54$ , 95% CI [0.47, 0.62],  $\beta = 0.61$ , 95% CI [0.53, 0.70],  $p < .001$ .

**Does Influence Awareness On A Mann IA-AMP Completed At Time 2 Predict The Magnitude Of Mann AMP Effects Completed At Time 1?** Results indicated that influence awareness rates in the Mann IA-AMP predicted scores in the previously completed Mann AMP,  $B = .38$ , 95% CI [0.30, 0.47],  $\beta = .42$ , 95% CI [0.32, 0.52],  $p < .001$ .

#### **Non Preregistered Analyses: Does The Predictive Utility Of Influence Awareness Vary Between The Standard And Mann AMPs?**

Following data collection, we noted that effect sizes in the influence rates predicting Mann AMP effects was relatively similar to that reported in Experiment 2. We therefore examined if effect sizes for this analysis in Experiment 2 (where a standard AMP was used) differed significantly from those in Experiment 6 (where a Mann AMP was used). Data from Experiments 2 and 6 were pooled and a similar



regression model was constructed as used in those experiments (i.e., Influence Rate as IV, [Mann] AMP effect as DV), also adding AMP type (i.e., Experiment) as a fixed effect in the model. If Influence Rate significantly differed in how well it predicted AMP effects between the standard and Mann AMPs, then an interaction between Influence Rate and AMP type (i.e., Experiment) should emerge. However, no such interaction was observed,  $B = 0.04$ , 95% CI [-0.09, 0.18],  $\beta = 0.05$ , 95% CI [-0.11, 0.21],  $p = .534$ . In order to quantify evidence for the absence of this interaction, a Bayes Factor for the interaction effect was computed using the BayesFactor R package (Morey & Rouder, 2019) by comparing models within and without this interaction effect. This Bayesian analysis using the default prior (Cauchy distribution placed on the effect size with scaling factor  $r = 0.5$ ) revealed moderate evidence in support of the null hypothesis,  $BF_{10} = 0.12$ .

## Discussion

The Mann et al. AMP was recently introduced with the aim of eliminating a similar phenomenon as in our previous experiments: namely, that only a subset of participants contribute to the AMP effect. However, we found the same pattern in that version of the task as we did in the standard task: a subset of influence aware trials (within participants), and highly influence awareness participants (between participants) strongly moderated AMP effects. Influence awareness rates in the Mann IA-AMP also predicted effects sizes in a previously completed Mann AMP. Furthermore, the extent to which influence awareness rates predicted the size of AMP effect sizes did not differ from, and was credibly equivalent to, what was observed in Experiment 2 with the standard AMP. Put simply, we obtained the same pattern of outcomes as reported in Experiments 2-5 with a variant of the AMP specifically designed to eliminate subset effects seen in other AMP research.

## Experiment 7: Prospective Influence

### Awareness Measures Also Predict AMP Effects

Experiments 2-6 show that influence awareness is a strong moderator of the magnitude of AMP effects at the individual and group levels, across different versions of the task (standard, Mann et al., IA-AMP), and within the same or between different content domains. Critically, however, we always assessed influence awareness in a *retrospective* fashion such that people were first asked to emit an evaluative response and only then reflect on it. Although this reflection occurs mere milliseconds after the evaluative response itself, it is still in some sense post-hoc. We therefore wanted to know if our findings would replicate when a *prospective* measure was used, one where awareness is assessed before the evaluative response is emitted. To the best of our knowledge no such (prospective)

measure of awareness has ever been used to examine awareness in the AMP to date.

Experiment 7 represents an exact replication of Experiment 3 wherein a standard AMP is completed and then followed by an IA-AMP with primes from the same attitude domain. This IA-AMP was modified into a prospective awareness measure: participants first encountered a prime followed by a target. They were then asked to indicate if they were influence aware *before* they provided their evaluative response to the target (instead of after as in Experiments 1-6 and all prior studies in this area). If our findings were to replicate with this new measure it would lend further evidence to the idea that people are aware of the influence of the prime on evaluative responses, both in retrospective and prospective ways (see Figure 1).

## Method

### Sample Selection Strategy

Power analyses were identical to that of Experiment 3 and the sampling strategy was identical to previous experiments.

### Participants

184 participants took part, and of those, 153 (94 women) ranging in age from 18 to 63 ( $M = 32.58$ ,  $SD = 10.86$ ) provided complete and analyzable data.

### Procedure

The procedure was similar to that of Experiment 3 (i.e., informed consent and demographic information, a standard AMP, IA-AMP, post-hoc self-reported awareness measure, and exploratory questions), with one key exception: the IA-AMP was changed from a retrospective to a prospective measure of influence awareness.

**Prospective IA-AMP.** Prior to this task participants were told: “The second part of the study will now begin. This part of the study is similar to what you just did. However there is one important difference. This time, before you rate a Chinese character as being pleasant or unpleasant, we will ask you a question: do you think the response you are going to give to the Chinese character will be influenced by the image that first appeared onscreen? If so, and you think that how you are going to respond to the Chinese character will be influenced by the first image then press the spacebar. If not, then simply wait, and when given the opportunity, evaluate the Chinese character as you normally would.”

The IA-AMP consisted of the same parameters as in previous studies (i.e., 12 positively and 12 negatively valenced prime stimuli, 72 Chinese pictographs, 10 practice trials, 72 test trials) with one notable change: after the presentation of the target stimulus, but before emitting the evaluative response, participants were given the opportunity to press the spacebar to indicate if they believed their response to the target *will be*

influenced by the prime. This was achieved through the presentation of the cue to “Press spacebar if the picture will influence your response to the Chinese symbol” for a fixed 2000ms interval. The above sentence was removed from the screen following a response, although the response window was fixed regardless of whether a response was emitted or not.

## Results

### Analytic Strategy

Our analytic strategy was similar to that of Experiment 3.

### Data Preparation

Data preparation was identical to that of Experiment 3.

### Hypothesis Testing

**Replication Hypotheses.** A significant effect emerged on both the standard AMP,  $OR = 2.67$ , 95% CI [1.85, 3.87],  $p < .001$ , and the prospective IA-AMP,  $OR = 3.18$ , 95% CI [2.30, 4.40],  $p < .001$ .

**Critical Hypotheses: Are Prospective IA-AMP Effects Moderated By Influence Awareness?** Influence awareness moderated evaluations at the trial-by-trial level,  $OR = 10.38$ , 95% CI [7.86, 13.71],  $p < .001$ , and inter-individual differences in influence awareness moderated the magnitude of IA-AMP effects at the group level,  $B = 0.54$ , 95% CI [0.43, 0.64],  $\beta = 0.63$ , 95% CI [0.50, 0.75],  $p < .001$ .

**Does Prospective Influence Awareness In The IA-AMP Predict Standard AMP Effects?** A person’s influence awareness rate in the prospective IA-AMP completed at Time 2 predicted the magnitude of their effect in a standard AMP completed at Time 1,  $B = 0.41$ , 95% CI [0.27, 0.54],  $\beta = 0.45$ , 95% CI [0.30, 0.59],  $p < .001$ .

## Discussion

A prospective measure of influence awareness yielded similar findings to the retrospective measures used in Experiments 1-6. Participants are able to identify that the prime is going to influence their response to the target even before they evaluate the target itself.

### Experiment 8: Prospective Influence

#### Awareness Measures (Prior to Prime

#### Presentation) Also Predict AMP Effects

In our final study we wanted to replicate and extend our findings with the prospective measure even further. Experiment 7 assessed for influence awareness before an *overt* evaluative response was emitted. However, it is possible that people may still have formed a *covert* evaluation of the target after having seen the prime. We therefore had participants register their influence awareness response *before* the target

stimulus was shown (see Figure 1). In this way, it is highly unlikely that they formed a covert evaluation of the target stimulus given that it had yet to appear. Likewise, it is highly unlikely that their performance on the influence awareness measure was associated with a covert evaluation of the target stimulus for the very same reason. In such a situation it seems most plausible that participants are aware of the prime and the impact it is going to have on their evaluations of a target that will be encountered at a later point in time.

## Method

### Sample Selection Strategy

Power analyses were similar to Experiment 3 and the sampling strategy was similar to prior studies.

### Participants

188 participants took part, and of those, 154 (89 women) ranging in age from 18 to 64 ( $M = 29.81$ ,  $SD = 10.98$ ) provided complete and analyzable data.

### Materials

Materials were similar to Experiment 3 with the exception that the influence awareness response emitted on each trial was recorded prior to the presentation of the target.

### Procedure

The procedure was similar to Experiment 7 (i.e., informed consent, demographics, standard AMP, prospective IA-AMP, self-reported awareness, exploratory questions), with one exception: influence awareness was now assessed before an *overt* evaluative response was emitted.

**Prospective IA-AMP.** A similar prospective IA-AMP was used as in Experiment 7. However, this time, after the presentation of the prime stimulus, but before presentation of the target stimulus, and thus before a target evaluative response was emitted, participants were given the opportunity to press the spacebar to indicate if they believe their evaluation of the target *will be influenced* by the prime. This was achieved through the presentation of a cue to “Press spacebar if the picture will influence your response to the next image” for a fixed 2000ms interval. The above sentence was removed from the screen following a response (although the response window was fixed regardless of whether a response was emitted or not).

## Results

### Hypothesis Testing

**Replication Hypotheses.** A significant effect emerged on both the standard AMP,  $OR = 2.49$ , 95% CI [1.75, 3.56],  $p < .001$ , and the prospective IA-AMP,  $OR = 2.98$ , 95% CI [2.22, 4.00],  $p < .001$ .

**Critical Hypotheses: Are Prospective IA-AMP Effects Moderated By Influence Awareness?** At the trial level, prospective influence

awareness moderated evaluative responses,  $OR = 4.78$ , 95% CI [3.78, 6.05],  $p < .001$ . This was also the case at the participant level: participants' prospective influence awareness rates strongly moderated the magnitude of their IA-AMP effect,  $B = 0.49$ , 95% CI [0.37, 0.61],  $\beta = 0.55$ , 95% CI [0.41, 0.68],  $p < .001$ .

### **Does Prospective Influence Awareness In An IA-AMP Completed At Time 2 Predict Effects In A Standard AMP Completed At Time 1?**

Influence awareness rates in the prospective IA-AMP significantly predicted the magnitude of their effect in a previously completed standard AMP,  $B = 0.34$ , 95% CI [0.20, 0.49],  $\beta = 0.35$ , 95% CI [0.20, 0.50],  $p < .001$ .

### **Discussion**

Results indicate that an even more conservative prospective measure of influence awareness resulted in the same pattern of results as in Experiments 1-7 (see Figures 3 and 4). Participants are able to report that a prime is going to influence their response to a target even before they have seen that target for themselves.

### **Meta-Analyses**

Meta-analyses were conducted in order to estimate the effect with greater precision across studies, and to estimate heterogeneity in the effect between experiments and across methodological variations. Meta-analyses were conducted using the lme4 R package (Bates, Mächler, Bolker, & Walker, 2015). Meta-analyses were not preregistered, although the hypotheses assessed within them and their model specifications were identical to those preregistered in the original experiments, with the addition of a random intercept for experiment. Data from all novel experiments (2-8) were included (total  $N = 1309$ ,  $k = 7$ ).

### **The AMP Effect Is Strongly Moderated By Awareness**

#### ***Inter-Individual Differences In Awareness Moderate the IA-AMP Effect***

As in the preregistered analyses for the individual studies, we assessed whether the absolute magnitude of the AMP effect on the IA-AMP was associated with the influence awareness rate on that IA-AMP. Results demonstrated that a large proportion of the variance in AMP effects was attributable to the influence awareness rate between participants,  $B = 0.52$ , 95% CI [0.48, 0.55],  $\beta = 0.60$ , 95% CI [0.56, 0.64],  $p < .001$ .

Recall that the AMP effect is the difference in evaluations on trials involving positive versus negative primes, and can range from 0 (evaluations unrelated to prime valence) and 1 (all evaluations congruent with primes). The model intercept was  $B = 0.14$ , 95% CI [0.12, 0.16],  $\beta = -0.01$ , 95% CI [-0.07, 0.05],  $p < .001$ . At the two extremes, in participants who report being aware of the influence of the prime on their evaluations on 0% of trials, the estimated marginal mean AMP

effect on the IA-AMP was therefore 0.14. In contrast, in participants who report being aware of the influence of the prime on their evaluations on 100% of trials, the estimated marginal mean AMP effect on the IA-AMP was 0.66. The AMP effect was therefore estimated to be three times larger in fully influence aware participants than fully non-influence aware participants. Very little of the variance was attributable to differences between experiments ( $R^2 = .07$ ) compared to influence awareness ( $R^2 = .36$ ). Given the methodological differences between the experiments, this implied that the strong moderation of the AMP effect on the IA-AMP by inter-individual differences influence awareness had good generalizability.

#### ***Inter-Individual Differences In Awareness On The IA-AMP Moderate The AMP Effect On A Previously Completed AMP***

An identical set of analyses was conducted using the AMP effect on the standard AMP rather than the IA-AMP as the dependent variable. Results demonstrated that a large proportion of the variance in AMP effects was attributable to the influence awareness rate between participants,  $B = 0.39$ , 95% CI [0.34, 0.44],  $\beta = 0.42$ , 95% CI [0.37, 0.48],  $p < .001$ .

The model intercept was  $B = 0.18$ , 95% CI [0.15, 0.21],  $\beta = 0.01$ , 95% CI [-0.08, 0.09],  $p < .001$ . At the two extremes, in participants who report being aware of the influence of the prime on their evaluations on 0% of trials, the estimated marginal mean AMP effect on the IA-AMP was 0.18. In participants who report being aware of the influence of the prime on their evaluations on 100% of trials, the estimated marginal mean AMP effect on the IA-AMP was 0.57. The AMP effect was therefore estimated to be three times larger in fully aware participants than fully non-aware participants. Very little of the variance was attributable to differences between experiments ( $R^2 < .01$ ) compared to influence awareness ( $R^2 = .18$ ). Given the methodological differences between the experiments, this again implied that the strong moderation of the AMP effect on a standard AMP by inter-individual differences influence awareness had good generalizability.

In summary, knowing an individual's influence awareness rate is sufficient to predict the magnitude of their AMP effect on a standard AMP that was completed prior to capturing the influence awareness rate (i.e., the AMP effect could not have been perturbed as awareness was only asked about later). This effect was found across 7 studies with very little evidence of heterogeneity, suggesting high replicability and generalizability across methodological variations (e.g., when within each IA-AMP trial influence awareness was assessed, and the domain being assessed; see Figure 3). As noted in Experiment 4, this effect holds

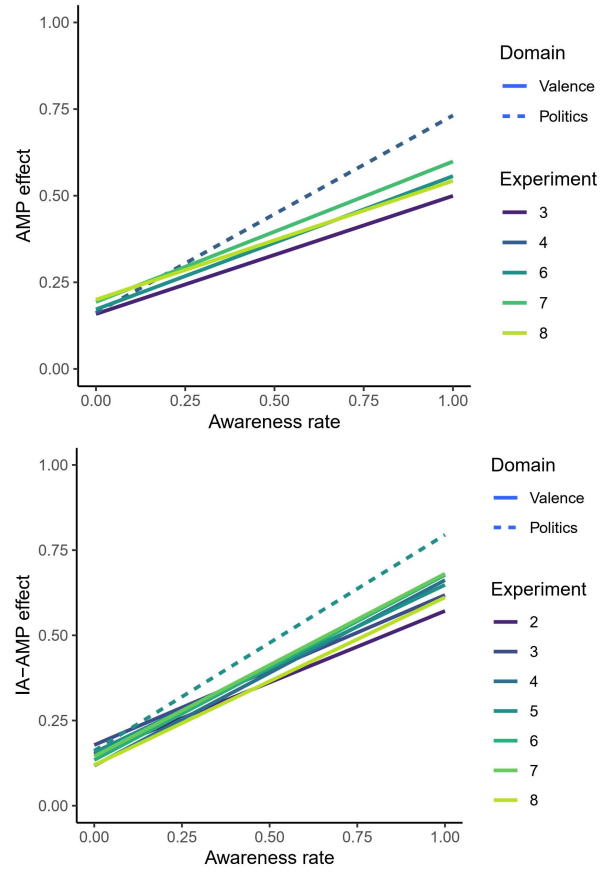
even when the IA-AMP (used to capture the influence awareness rate) and the standard AMP are assessing different domains (valence vs. politics).

### ***Trial-By-Trial Awareness Moderates the AMP Effect***

IA-AMP effects were found to be moderated not only by inter-individual differences in awareness (as in the above analysis), but also intra-individual at the trial level,  $OR = 22.82$ , 95% CI [20.72, 25.14],  $p < .001$  (see Figure 4).

### **What Is The Distribution Of Influence Awareness Across Participants?**

Data were pooled in order to understand the distribution of influence awareness rates between participants. This analysis was exploratory and not included in the preregistrations for the individual experiments. Hartigan's dip test demonstrated non-unimodality ( $D = 0.03$ ,  $p < .001$ ). As a robustness test, we also analyzed the subset of AMPs that employed the Mann et al. (2019) modifications, which also demonstrated non-unimodality ( $D = 0.03$ ,  $p = .030$ ). Visual inspection of distribution kernel-density plots demonstrated clear bimodality (see Figures 5-6). Gaussian kernel density estimation was used to estimate the two modes: influence awareness rates were found to cluster around participants being either fully non-aware (Mode = .01) or fully awareness (Mode = .97; range = 0 to 1). This provided convergent evidence that it is the subset of highly influence aware participants and their subset of influence aware trials that represent the majority of variance in observed AMP effects.



*Figure 3.* Influence awareness rates on the IA-AMP and the absolute magnitude of AMP effects on the IA-AMP (upper panel) and a previously completed standard AMP (lower panel), across Experiments 2-8.

Figure 4. Trial-level influence awareness moderates the magnitude of IA-AMP effects. Point estimates represent marginal means from the meta-analytic model of Experiments 2-8 and their 95% Confidence Intervals.

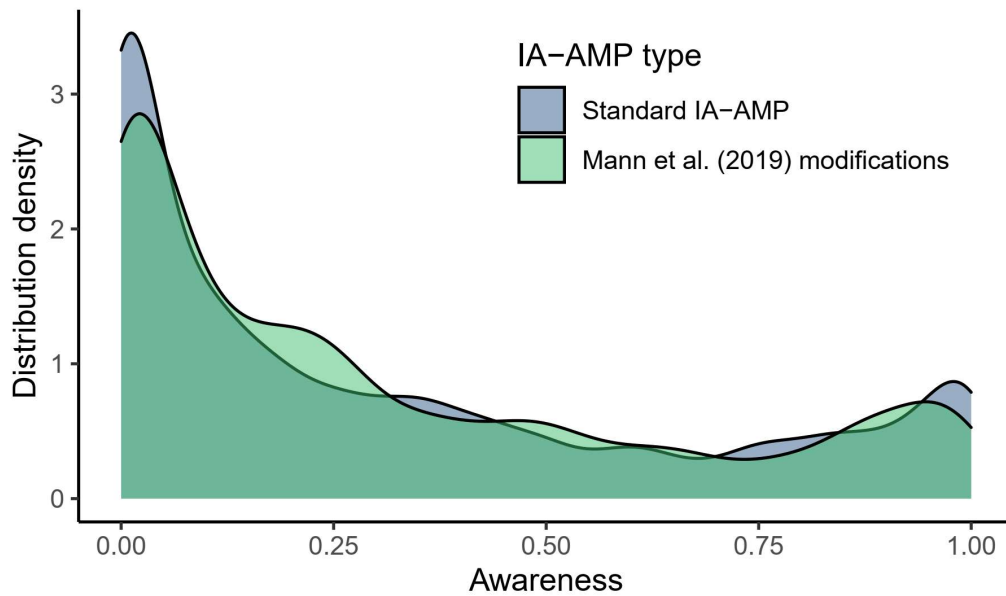
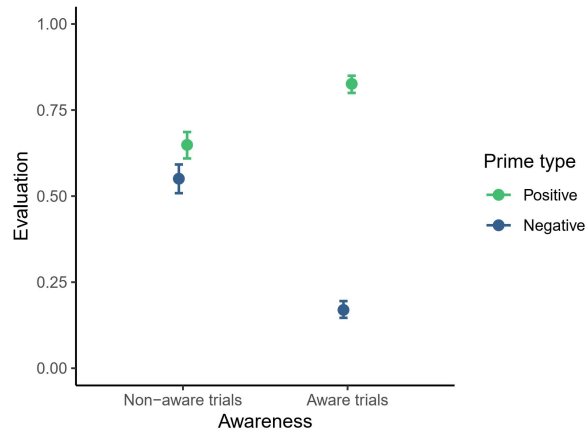


Figure 5. Bimodality in the distribution of participants' influence awareness rates in the IA-AMPs pooled across Experiments 2-8.

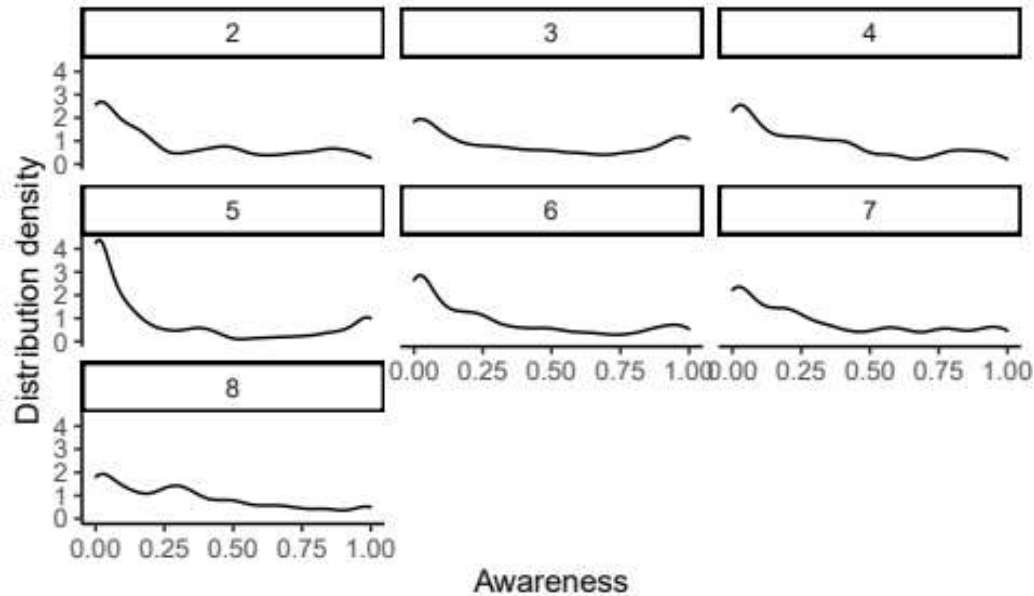


Figure 6. The distribution of influence-awareness rates in the valence IA-AMP in Experiments 2-8.

### General Discussion

Over the past 15 years people have used the AMP under the assumption that its effects represent an implicit measure of attitudes, stereotypes, and other biases. Effects on the task are said to be ‘implicit’ insofar as prime stimuli influence how target stimuli are evaluated without a person’s intention for this to happen, or awareness that such an occurrence has taken place. In this paper we extend on past work that examined the role of awareness in AMP effects. Across eight preregistered, high-powered studies, including a direct replication and meta-analyses, we re-examined if participants are indeed aware of the prime’s influence on their evaluations. In what follows we summarize our findings and then discuss their implications for research and theorizing on the AMP.

### Overview of Findings

Experiment 1 began with an attempt to replicate the finding that standard and ‘skip’ AMP effects do not differ. The finding that ‘skip’ AMP effects (i.e., those where ‘influenced’ responses have been detected and removed) are no different to standard AMP effects is viewed as strong support for the implicit account. We examined if these findings would replicate using an improved and more highly-powered design. Results indicated that this finding did not replicate, such that scores on the standard AMP were significantly different from those on the skip AMP. Exploratory analyses also revealed that a given person’s awareness of the prime’s influence on their evaluations (during the skip AMP at Time 2) was strongly related to the magnitude of their effects in the standard AMP at Time 1.

A limitation of the ‘skip’ AMP is that it forces people to either skip or evaluate the target and thus only provides partial data. To overcome this we developed an influence aware IA-AMP in Experiment 2 wherein participants rated the target (thus providing evaluative information) and then indicated if that evaluation had been influenced by the prime (thus providing influence information). Results indicated that AMP effects emerged and that these were moderated at the trial-by-trial level by a subset of (influence aware) trials, and AMP effects were produced predominantly by highly influence aware participants at the group level.

In Experiments 3-4 we controlled for the possibility that by probing for influence awareness on each trial of the IA-AMP we artificially altered the relationship between awareness and AMP effects. Participants now completed a standard AMP at Time 1 and an IA-AMP at Time 2, either from the same (Experiment 3) or different attitude domains (Experiment 4). Because the standard AMP was always completed prior to the IA-AMP, effects on the former were always unperturbed by modifications to the latter. Yet in both studies influence awareness during an IA-AMP at Time 2 predicted the magnitude of standard AMP effects at Time 1, indicating that influence awareness is a stable (within-participant) pattern of responding that holds within and between content domains.

Experiment 5 extended our analyses to three additional questions: is the predictive validity of the AMP effect also dependent on influence awareness; is there intra-individual stability in influence awareness from one AMP to another; and is the relationship

between influence awareness and AMP effects bidirectional, such that the presence of one predicts the presence of the other. Two groups of participants (Democrats and Republicans) first completed a political IA-AMP and then an IA-AMP with generic valenced primes. We found that the AMP's ability to correctly classify a person as a Democrat or Republican was superior when effects were based solely on influence aware trials and inferior when based solely on non-influence aware trials. A given person's influence awareness rate on one AMP was also strongly correlated with their influence awareness on another AMP, even when those tasks target entirely different attitude domains. Finally, the predictive relationship between influence awareness and AMP scores was bidirectional; influence awareness from AMP #1 predicted the AMP effect of AMP #2, and vice versa.

Experiment 6 took a newly developed version of the AMP that purportedly reduces subset effects within the AMP (the Mann et al. [2019] AMP) and examined if influence awareness also plays a role here too. Participants first completed a standard Mann AMP and then a Mann IA-AMP. Once again, the same pattern of findings emerged: a subset of influence aware trials (within participants), and highly influence awareness participants (between participants) moderated AMP effects. Influence awareness rates in the Mann IA-AMP also predicted effects sizes in a previously completed Mann AMP. Put simply, the same pattern of outcomes emerged even within a variant of the task designed to optimize the implicitness of the AMP.

In our final two studies we modified the IA-AMP so that influence awareness was measured *prospectively*, either before the target was evaluated (Experiment 7) or before the target stimulus was even presented (Experiment 8). In this way influence awareness was measured before an overt evaluation took place or a covert evaluation could even be formed. In both studies the same pattern of findings emerged as before.

In short, our findings demonstrate that (a) the AMP effect and its predictive validity appear to be moderated by influence aware responding, (b) influence awareness rates vary widely between individuals but are highly consistent within individuals, within and between attitude domains, (c) participants who are more highly influence aware contribute most to variance in group-level AMP effects, and that (d) recent modifications to the AMP that purportedly control for such subsample effects do not reduce or resolve this issue. Although non-influence aware trials retain some degree of predictive validity and contribute to some extent to the magnitude of effects, their contributions pale in comparison to that of influence aware trials.

## Implications

### *Implicit vs. Explicit Accounts*

Our findings consistently show that the majority of variance in group-level AMP effects is explained by those trials wherein participants are aware of how the prime will (prospective measures) or has (retrospective measures) influence(d) their evaluations. This claim holds across eight preregistered studies, different attitude domains, multiple versions of the AMP (standard, Mann et al., IA-AMP), and different influence awareness measures (prospective and retrospective measures taken on each trial vs. post hoc self-report questions). Thus it seems reasonable to suggest that AMP effects may not be 'implicit' in the sense of unaware. Rather our findings suggest that people are aware of the prime's influence on how they are responding to the target.<sup>12</sup>

### *Theoretical Implications*

Do AMP Effects Reflect A Misattribution Process? So far we have focused on the role of awareness in AMP effects. However, our findings are also relevant to another issue, namely the idea that AMP effects are mediated at the mental level by misattribution of prime valence to the target stimulus. Misattribution is traditionally conceived of as occurring in the absence of awareness (Schwarz & Clore, 1983; Payne et al., 2005). Indeed, as a reviewer of this paper noted, misattribution by definition cannot occur with awareness. If AMP effects rely heavily on those participants who show awareness of the prime and its influence on their evaluations (as our results indicate), then this suggests two possibilities.

On the one hand, AMP effects may reflect misattribution, as is often claimed, yet people are fully aware that misattribution is taking place. Our findings with prospective measures in Experiments 7-8 would also require people to not only be aware of misattribution but be able to predict that it is going to occur even before a target is evaluated or a target stimulus is even presented. Although this idea runs contrary to how misattribution is traditionally defined (Schwarz & Clore, 1983) it is possible. For instance, a reviewer of this paper argued that misattributions are - by definition - mistakes. As such, they cannot be entirely understood by the person making the misattribution otherwise they would presumably not make that error. When it comes to the AMP participants may be aware that their target evaluations are influenced by the primes (especially in cases where they feel strongly about the primes). But they may also hold the belief that the target really does have a particular valence. In other words, just because

<sup>12</sup> Note that we are agnostic to the AMP's implicitness in other senses of the word (e.g., intentional): our goal here was to reevaluate the AMP's specific claim to implicitness in the sense of awareness.

participants are aware of the priming effect does not necessarily mean that they are aware of how it occurred. And despite this awareness they may also still genuinely consider the target to be positive or negative.

On the other hand, if one contends that misattribution is a mistake made by the individual, and one that occurs without either prospective or retrospective awareness of the prime and its eventual influence on their target responses, then it may be that misattribution is not the best mechanism to explain our AMP effects. This possibility would have implications for those theories and methods that rest on this idea. For instance, it would challenge a misattribution account of AMP effects as well as theoretical perspectives on misattribution that rely on the AMP for support. This includes theoretical models relating to the process of misattribution itself (e.g., the process model of misattribution: Payne, Hall, Cameron, & Bishara's, 2010), as well as claims that evaluative conditioning is based on a misattribution process (Jones et al., 2009), and that psychological properties beyond evaluations can also be misattributed (Blaison, Imhoff, Hühnel, Hess, & Banse, 2012). It would also call into question a number of second-generational tasks that attempt to exploit the misattribution of meaning (the Semantic Misattribution Procedure: Sava et al., 2012) and truth (the Truth Misattribution Procedure: Cummins & De Houwer, 2019). It seems likely that the very same issues associated with influence awareness in the standard AMP are likely to play similar roles in these other procedures. Future work could employ a similar IA-AMP style manipulation to these variants to investigate this issue in more detail.

**Alternative Theoretical Possibilities.** During the review process, several reviewers offered alternative theoretical accounts of our findings. One advanced an *inferential perspective*, according to which participants generate a series of inferences about the stimuli they encounter during the task which then guide how they respond. For instance, when retrospectively asked about awareness in Experiments 1-6 (i.e., before or after they emitted an evaluative response on each trial in Experiments 2-8, or in lieu of an evaluative response in Experiment 1), participants may have inferred that their evaluative response had been influenced by the preceding prime stimulus. This could also have been the case in Experiment 7 (prospective awareness assessed before a response was emitted), if one were to assume that participants knew what their evaluation of the target was going to be when they were given the chance to emit it at the end of the trial. One could even postulate an inferential explanation for Experiment 8 (prospective awareness before the target was even presented), such that participants may have inferred that priming is likely to occur based on the

extremity of their reaction to the prime; perhaps because they knew that, in general, they do not have much of an opinion about the target stimulus. This inferential perspective also makes assumptions about some of the more unexpected findings in our paper. Take the fact that standard AMP effects at Time 1 were strongly related to influence awareness rates in the IA AMP at Time 2. It is possible that participants who showed priming effects in the standard AMP may be more likely to infer that priming sometimes occurs, based on their observation of the congruency between their evaluation of the primes and their evaluation of the targets. Then, in the IA-AMP, when they see primes that elicit strong reactions, they are more likely to predict that priming is about to happen. Such an account shares commonalities with predictive processing theories elsewhere in psychology (Corlett, Mohanty, & MacDonald, 2020) as well as inferential accounts in attitude research in particular (De Houwer, 2018). It's important to note that if one accepts this account it necessarily also applies to the original study that is cited as providing evidence that AMP effects are unaware (i.e., Payne et al. 2013, Experiment 3). That is, this account does not argue that AMP effects are in fact unaware, but rather than the evidence to date cannot answer this question. As such, the key implication of this account is congruent with our own account: that this is little evidence supporting the claim that the AMP is unaware. Regardless, it is important to acknowledge that this account cannot easily account for the results of Experiment 8, where participants were asked to report whether the prime stimulus is likely to influence their later evaluation of the target stimulus, given that this rating was provided before the target stimulus was presented or an evaluation of it was given.

Another reviewer advanced an *attitude strength* perspective on our findings. The idea here is that if the AMP is equally valid for all participants, and judgments of influence are caused by attitude strength (i.e., the strength of affective reactions to the primes), then participants with stronger attitudes will consequently report more influence of the primes. When participants are separated based on their influence awareness ratings, the ones who report little influence will be those with weak attitudes and little variability on AMP scores. Those who report high influence will have strong attitudes and more extreme priming effects. Thus, influence awareness would be a by-product or corollary of attitude strength rather than the main factor driving AMP effects. This implications of this account for what we should conclude about whether AMP effects are unaware could be one of two options. The first possibility is that this issue also applies to the evidence presented by Payne et al. (2013, Experiment 3). For example, perhaps the decision to skip highly influenced trials would also be a corollary of the



evaluation. In this case, this would again remove the evidence both for and against the claim that the AMP is unaware, leaving this core claim unsupported. The second possibility is that this issue does not apply to the Skip-AMP paradigm employed by Payne et al. (2013, Experiment 2), perhaps because this paradigm does not ask participants to report any evaluation they may have if they were influenced by the prime. However, in this case, then this issue also does not apply to our failed replication of Payne et al.'s study in Experiment 1. As such, regardless of the theoretical account one adopts, there is reason to doubt that there is firm evidence that AMP effects are unaware. Regardless of which theoretical account one advocates for, our findings inspire new theoretical perspectives on the AMP effects and highlight the need for more elaborated consideration of the role and nature of awareness within the task.

## **Future Research Directions**

### ***Creating a Better Implicit Measure***

One option is to modify the AMP effect in ways that exclude influence aware trials or refine the task itself in some way that diminishes the role of influence awareness on those effects. These changes would allow the AMP to maintain its status as an implicit measure. Our results could be seen as supporting this approach given that (a) even those participants with influence awareness rates of zero demonstrated (very small) IA-AMP effects, and (b) IA-AMP effects calculated from non-influence aware trials still possessed some predictive validity for discriminating between known groups. As such, researchers may be tempted to set the standard AMP to the side, employ an IA-AMP, exclude all influence aware trials, and calculate an effect. This is certainly one way forward. Yet it also comes with its own issues.

First, one should not conflate ‘non-influence aware’ with ‘influence *unaware*’ responding. The IA-AMPs used here asked participants to press the spacebar if their evaluation was influenced by the prime. The presence of such a response provides a measure of influence awareness. Yet the absence of such a response is far more ambiguous. It may be that such trials are free from influence awareness (i.e., are ‘influence-unaware’), or they could equally reflect uncertainty about influence, momentary distraction, or other sources of control over responding. Put simply, caution should be exercised when assigning a specific meaning to non-influence aware trials in the IA-AMP. To better investigate influence-unaware trials, one would need to develop and test a hypothetical ‘Influence-Unawareness AMP’ (IU-AMP): for example, by asking people to respond when their evaluation was *not* influenced by the prime.

Yet even an IU-AMP would not be without issue. Imagine that an applied researcher in a specific domain wishes to examine differences between two known-groups using the IU-AMP and obtains results similar to what we report in Experiment 5 ( $d = 0.62$ , using IA-AMP non-influence aware trials for comparison). To appropriately power her study using the IU-AMP to detect group differences would require at least 138 participants.<sup>13</sup> For the applied researcher, collecting such sample sizes is often either unfeasible or a poor use of limited resources. In contrast, if predictive utility was more important to her than ‘process purity’, then an IA-AMP capturing influence aware responses could detect group differences with as few as 16 participants (i.e., IA-AMP effects calculated from influence aware trials:  $d = 2.08$ ).

Now imagine the flipside. For basic researchers, the need to collect larger sample sizes may be both feasible and desirable if this allows them to study implicit processes in a relatively ‘pure’ manner. The problem here is that an IU-AMP will likely also lead to a significant number of people being discarded due to zero, or near-zero, levels of unaware task responding. The implication here is that although such an IU-AMP might provide a better implicit measure by implementing changes to the task, the effects obtained from such a task would still not reflect behaviors (or mental processes) in people *in general*. Yet this is exactly what the AMP is primarily used for. Therefore, just as other fields acknowledge the variety of issues associated with making inferences or generalizations about people in general from non-representative samples (e.g., WEIRD individuals: Henrich, Heine, & Norenzayan, 2010; neuroscience tending to only study the brains of right-handed people: Willems, der Haegen, Fisher, & Francks, 2014; animal models of pathology that are based on male biology but not female: Mogil, 2016), we need to do the same. Both applied and basic researchers using the AMP (or AMP-like tasks, including the IU-AMP) need to carefully attend to the dangers of making inferences and generalizations about people *in general* from an effect that reflects a special subset of people.

### ***Exploring The (Interactive) Role of Automaticity Conditions Within the AMP***

Our work focused on one particular automaticity condition (awareness), and one particular sub-type of awareness (influence awareness). It would be interesting to explore if other types of awareness also impact AMP effects in the way that influence awareness does. For instance, are AMP effects moderated by a participant’s awareness of the content under investigation (i.e., that they have an automatic

<sup>13</sup> Using G\*Power (Faul, Erdfelder, Buchner, & Lang, 2009): Independent *t*-test, alpha = 0.05 (two-sided), power = 0.95.

evaluative preference for one group over another). Or their awareness of how their automatic content is being assessed by the task (e.g., via speeded categorization of one social group with positive words and a second group with negative words as in the IAT). Or the presence of particular stimuli and response requirements in the procedure (e.g., that a black person is being presented as a prime stimulus). Or their awareness of the origin of their AMP effects (e.g., stemming from society or their personal learning history).<sup>14</sup>

Similarly, it would be interesting to explore if and how awareness interacts with other automaticity conditions, and how their interaction contributes to the AMP effect. Take, for instance, intention. Is it that awareness of the prime's influence on target evaluations is a precondition for intentional responding (i.e., one can only intend to respond to the target in-line with the prime once they are aware of both stimuli and the latter's relationship to the former)? Is it a precondition for controlling how one responds (demand) or for combating the unwanted influence of certain stimuli over automatic behavior (e.g., if one wants to exert control over their automatic responses do they first need to be aware of the relationship between stimuli in the environment and how they influence those responses)? Is it possible to be influence aware and still respond in an unintentional manner? Future work could explore these and related issues.

### **Revision of Existing Findings**

Research on the AMP is typically founded on two assumptions: (a) that AMP effects are reflective of implicit attitudes, and (b) that AMP effects represent an equally valid measure of such attitudes across all individuals (e.g., Fox et al., 2018; Kalmoe & Piston, 2013; Mann et al., 2019; Payne et al., 2005; Rinck & Becker, 2007; Spring & Bulik, 2014). To illustrate, consider a study by Franklin, Puzia, Lee, and Prinstein (2014), which concluded that “young adults with a history of non-suicidal self-injury (NSSI) display a significantly stronger *implicit* identification with [images of skin] cutting” compared to their counterparts without such a history of NSSI. Our findings suggest that such a result should instead be interpreted as “in those young adults who are highly influence aware on the AMP, those with a history of NSSI also self-identify more with NSSI compared to those who had no such

history. However, little can be said about those with low influence awareness rates.” This is just one example; similar revisions need to be applied to the core claims of research using the AMP (e.g., via systematic review), which may alter the conclusions derived from that body of work.

### **What Form Does Influence Awareness Take in a Standard AMP?**

Our findings highlight a deep connection between performance on the IA-AMP and standard AMP. The fact that performance on the IA-AMP predicts the magnitude of standard AMPs that either come before or after it, both within and between different attitude domains, and when influence awareness is measured in different prospective and retrospective manners indicates that what holds in the former likely plays out in the latter. But this raises the question – how does influence awareness play out in standard AMPs? Is it that participants pause after each and every prime-target presentation, reflect on what they just saw, and then use this information to inform every individual target evaluation that they make? This seems unlikely to us given that trial latencies (while not critical to the effect or the task) are typically short. One alternative is that people don't stop and reflect on every standard AMP trial but do stop and reflect on multiple trials, and this awareness of the prime's influence grows as one completes the task. Yet another alternative is that stopping and reflecting on a single trial early on in the task may lead people to generate inferences that influence how they subsequently perform (e.g., “Certain primes keep influencing how I respond...” or “Certain stimuli [primes] keep flashing up on screen... maybe I'm supposed to respond to them in a particular way...”). In other words, as one Reviewer suggested, responding to the targets than follow primes may make people self-conscious about their performance, and attune them to possible explanations for that performance, that subsequently influence it. In short, future research could examine how influence awareness actually plays out within the standard AMP. This would not only provide useful information about the task itself, but also about influence awareness as a process.

### **What Makes A Person Influence Aware?**

Our results also raise the question of what characterizes and differentiates highly influence aware participants who moderate AMP effects from the rest of the population. Experiment 6, which employed the modifications to the AMP suggested by Mann et al. (2019), suggests that both influence awareness rate and moderation of the magnitude of AMP effects by influence awareness is not reduced through simple alterations to the task itself. Experiment 5 suggests that an individual's influence awareness rate is consistent across IA-AMPs assessing different domains, and that

<sup>14</sup> The different ways that awareness can be potentially defined and studied makes it difficult to (a) arrive at general conclusions about the role of awareness from past work, and (b) requires precision about what type of awareness one is measuring, and how they have done so (for a more detailed treatment of this issue see Hahn & Goedderz, 2020; Corneille & Hütter, 2020). It is for this reason that we continually made reference to *influence* awareness throughout this paper, and sought to measure it using a range of measures (e.g., trial-by-trial vs. post-hoc self-reports), attitude domains, and both prospectively and retrospectively.

the influence awareness rate demonstrated in one domain predicts AMP effects in another. As such, it seems that influence awareness rates are an individual difference variable rather than merely random noise or properties of the task itself. While beyond the scope of the current research, future work could examine whether influence awareness is a state- or trait-like property (e.g., whether it is consistent across time and context), whether it is related to other individual differences (e.g., Need for Cognition: Cacioppo & Petty, 1982), or indeed whether influence awareness on the AMP is related to performance on other kinds of implicit measures (e.g., the Implicit Association Test).<sup>15</sup>

### Diversity and Inclusiveness

Experiments 1-8 were carried out online on a platform that recruits participants from the general population (Prolific Academic). An analysis of the demographic data we requested in our studies (age, gender, and political orientation [in Experiments 4 and 5]) revealed that our samples were broadly representative in terms of age (ranging from 18 to 65 years), and balanced in terms of gender (847 women, 741 men, 15 other). As such, in our efforts to reexamine different properties of the AMP effect, we recruited samples that were, at worst, no less representative as the original AMP studies that we build and extend upon. At best, our samples are likely more representative than the original studies given that we did not recruit solely from undergraduate students (i.e., more balanced in terms of age and levels of education). That said, we did not request information on other demographic variables (e.g., sexuality, ethnicity). Although we had no theoretical reason to assume that such variables would moderate performance on a generic valence AMP, it is perhaps plausible they are associated with performance on the political IA-AMPs in particular. Likewise, given that Experiments 4-5 were sampled exclusively from US residents, and that the majority of Prolific Academic participants reside in the UK, our samples are primarily made up of (and potentially over-represent) individuals from these nations. Future work may therefore wish to capture more detailed demographics information or could replicate our findings across different nationalities to further expand its remit.

<sup>15</sup> One could also explore other factors that may impact how much one is likely to be influence aware during an AMP, from one's previous experience with the task (e.g., extensive vs. limited prior AMP exposure), as well as their motivation or opportunity to process information during the task in an effortful manner. Likewise, one could examine what properties of the AMP alter one's likelihood of being influence aware. Is it that certain types of primes that are more likely to elicit influence awareness than others (e.g., because of their valence, extremity, familiarity). Is it that certain types of instructions direct attention towards vs. away from the prime, or the relationship between the prime and target, and this also alters influence awareness? Both topics seems worthy of future investigation.

### Conclusion

AMP effects do not appear to be implicit in at least one sense (i.e., unaware). Our results show that both the magnitude of AMP effects and their predictive validity are strongly moderated by awareness. As such, insofar as (1) the AMP has been claimed to be implicit in the sense of being (influence) unaware and (2) utility can be defined in terms of large AMP effects and/or AMP effects that possess predictive validity (i.e., the criteria employed in this previous literature to date; e.g., Payne et al., 2013), what is useful about the effect is not particularly implicit, and what is implicit about the effect is not particularly useful. This finding raises a host of conceptual, theoretical, and applied questions for past and future research using the task as well as its ability to make inferences about psychological phenomena in people *in general* (i.e., rather than merely in a subset of highly aware participants).

### References

- Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality & Social Psychology Bulletin*, 38, 1194–1208.  
<https://doi.org/10.1177/0146167212446835>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48.  
<https://doi.org/10.18637/jss.v067.i01>
- Blaison, C., Imhoff, R., Hühnel, I., Hess, U., & Banse, R. (2012). The Affect Misattribution Procedure: hot or not? *Emotion*, 12, 403–412.  
<https://doi.org/10.1037/a0026907>
- Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science*, 10, e1501.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116–131.  
<https://doi.org/10.1037/0022-3514.42.1.116>
- Chapman, M. V., Hall, W. J., Lee, K., Colby, R., Coyne-Beasley, T., Day, S., ... Payne, B. K. (2018). Making a difference in medical trainees' attitudes toward Latino patients: A pilot study of an intervention to modify implicit and explicit attitudes. *Social Science & Medicine*, 199, 202–208.  
<https://doi.org/10.1016/j.socscimed.2017.05.013>
- Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, 24, 212–232.

- Cummins, J., & De Houwer, J. (2019). An inkblot for beliefs: The truth misattribution procedure. *PloS one*, 14, e0218661.
- De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The Handbook of Implicit Cognition and Addiction* (pp. 11–28). Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781412976237.n2>
- Derrick, B., Toher, D., & White, P. (2017). How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017). *The Quantitative Methods for Psychology*, 13, 120–126. <https://doi.org/10.20982/tqmp.13.2.p120>
- Ditonto, T. M., Lau, R. R., & Sears, D. O. (2013). AMPing Racial Attitudes: Comparing the Power of Explicit and Implicit Racism Measures in 2008. *Political Psychology*, 34, 487–510. <https://doi.org/10.1111/pops.12013>
- Dunham, Y., & Emory, J. (2014). Of affect and ambiguity: The emergence of preference for arbitrary ingroups. *Journal of Social Issues*, 70, 81–98. <https://doi.org/10.1111/josi.12048>
- 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. <https://doi.org/10.3758/BF03193146>
- Fox, K. R., Ribeiro, J. D., Kleiman, E. M., Hooley, J. M., Nock, M. K., & Franklin, J. C. (2018). Affect toward the self and self-injury stimuli as potential risk factors for nonsuicidal self-injury. *Psychiatry Research*, 260, 279–285. <https://doi.org/10.1016/j.psychres.2017.11.083>
- Franklin, J. C., Puzia, M. E., Lee, K. M., & Prinstein, M. J. (2014). Low Implicit and Explicit Aversion Toward Self-Cutting Stimuli Longitudinally Predict Nonsuicidal Self-Injury. *Journal of Abnormal Psychology*, 123, 463–469.
- Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd, *Handbook of Research Methods in Social and Personality Psychology* (pp. 283–310). Cambridge University Press.
- Gawronski, B., & Payne, B. K. (2010). *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. Guilford Press.
- Gawronski, B., & Ye, Y. (2014). What Drives Priming Effects in the Affect Misattribution Procedure? *Personality and Social Psychology Bulletin*, 40, 3–15. <https://doi.org/10.1177/0146167213502548>
- Görge, S. M., Joormann, J., Hiller, W., & Witthöft, M. (2015). The Role of Mental Imagery in Depression: Negative Mental Imagery Induces Strong Implicit and Explicit Affect in Depression. *Frontiers in Psychiatry*, 6, 94. <https://doi.org/10.3389/fpsy.2015.00094>
- Greenwald, A., McGhee, D., & Schwartz, J. L. K. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Hahn, A., & Gawronski, B. (2019). Facing one's implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*, 116(5), 769–794. <https://doi.org/10.1037/pspi0000155>
- Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion*, 8(6), 513–533. <https://doi.org/10.1080/02699939408408957>
- 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 and Emotion*, 25(4), 676–690. <https://doi.org/10.1080/02699931.2010.508260>
- Jasper, F., & Witthöft, M. (2013). Automatic Evaluative Processes in Health Anxiety and Their Relations to Emotion Regulation. *Cognitive Therapy and Research*, 37(3), 521–533. <http://dx.doi.org/10.1007/s10608-012-9484-1>
- Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, 96(5), 933–948. <https://doi.org/10.1037/a0014747>
- Kalmoe, N. P., & Piston, S. (2013). Is Implicit Prejudice against Blacks Politically Consequential? Evidence from the AMP. *Public Opinion Quarterly*, 77(1), 305–322. <https://doi.org/10.1093/poq/nfs051>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259–269. <https://doi.org/10.1177/2515245918770963>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*. NIMH Center for the Study of Emotion and Attention.
- Mann, T. C., Cone, J., Heggeseth, B., & Ferguson, M. J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology*, 116(3), 349–374. <https://doi.org/10.1037/pspa0000146>
- Mann, T. C., & Ferguson, M. J. (2017). Reversing Implicit First Impressions through Reinterpretation

- after a Two-Day Delay. *Journal of Experimental Social Psychology*, 68, 122–127.  
<https://doi.org/10.1016/j.jesp.2016.06.004>
- McCarthy, R. J., Skowronski, J. J., Crouch, J. L., & Milner, J. S. (2017). Parents' spontaneous evaluations of children and symbolic harmful behaviors toward their child. *Child Abuse & Neglect*, 67, 419–428.  
<https://doi.org/10.1016/j.chiabu.2017.02.005>
- Mogil, J. S. (2016). Perspective: Equality need not be painful. *Nature*, 535, S7
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326.  
<https://doi.org/10.1037/0033-2909.132.2.297>
- Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, D., & Lundberg, K. (2013). Intention Invention and the Affect Misattribution Procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin*, 39(3), 375–386.  
<https://doi.org/10.1177/0146167212475225>
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89(3), 277–293. <https://doi.org/10.1037/0022-3514.89.3.277>
- Payne, B. K., & Lundberg, K. (2014). The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms. *Social and Personality Psychology Compass*, 8(12), 672–686. <https://doi.org/10.1111/spc3.12148>
- Quertemont, E. (2011). How to Statistically Show the Absence of an Effect. *Psychologica Belgica*, 51(2), 109–127. DOI: <https://doi.org/10.5334/pb-51-2-109>
- Rinck, M., & Becker, E. (2007). Approach and avoidance in fear of spider. *Journal of Behavior Therapy and Experimental Psychiatry*, 38(2), 105–120. <https://doi.org/10.1016/j.jbtep.2006.10.001>
- Sava, F. A., MaricutToiu, L. P., Rusu, S., Macsinga, I., Vîrgă, 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(6), 613–628. <https://doi.org/10.1002/per.1861>
- Schreiber, F., Witthöft, M., Neng, J. M. B., & Weck, F. (2016). Changes in negative implicit evaluations in patients of hypochondriasis after treatment with cognitive therapy or exposure therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 50, 139–146.  
<https://doi.org/10.1016/j.jbtep.2015.07.005>
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgements of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3), 513–523.
- Smith, A. R., Forrest, L. N., Velkoff, E. A., Ribeiro, J. D., & Franklin, J. (2018). Implicit attitudes toward eating stimuli differentiate eating disorder and non-eating disorder groups and predict eating disorder behaviors. *The International Journal of Eating Disorders*, 51(4), 343–351.  
<https://doi.org/10.1002/eat.22843>
- Spring, V. L., & Bulik, C. M. (2014). Implicit and explicit affect toward food and weight stimuli in anorexia nervosa. *Eating Behaviors*, 15(1), 91–94.  
<https://doi.org/10.1016/j.eatbeh.2013.10.017>
- Teige-Mocigemba, S., Becker, M., Sherman, J., Reichardt, R., & Klauer, K. C. (2017). The Affect Misattribution Procedure: In Search of Prejudice Effects. *Experimental Psychology*, 64(3), 215–230.  
<https://doi.org/10.1027/1618-3169/a000364>
- Van Dessel, P., Mertens, G., Smith, C. T., & De Houwer, J. (2017). The Mere Exposure Instruction Effect. *Experimental Psychology*, 64(5), 299–314.  
<https://doi.org/10.1027/1618-3169/a000376>
- Ye, Y., & Gawronski, B. (2018). Validating the Semantic Misattribution Procedure as an Implicit Measure of Gender Stereotyping. *European Journal of Social Psychology*, 48(3), 348–364.  
<https://doi.org/10.1002/ejsp.2337>
- Zerhouni, O., Bègue, L., Comiran, F., & Wiers, R. W. (2018). Controlled and implicit processes in evaluative conditioning on implicit and explicit attitudes toward alcohol and intentions to drink. *Addictive Behaviors*, 76, 335–342.  
<https://doi.org/10.1016/j.addbeh.2017.08.026>

## Appendix

*Table 1.* Means and standard deviations of absolute magnitude of standard AMP effects

Experiment	Domain	Mean	SD
1	Valence	0.17	0.33
3	Valence	0.30	0.29
6	Valence	0.29	0.29
7	Valence	0.32	0.29
8	Valence	0.32	0.29

*Table 2.* Means and standard deviations of absolute magnitude of IA-AMP effects

Experiment	Domain	Effect	Mean	SD
2	Valence	IA-AMP Effect	0.28	0.25
	Valence	IA-AMP Influenced Effect	0.68	0.31
	Valence	IA-AMP Non-Influenced Effect	0.20	0.24
3	Valence	IA-AMP Effect	0.36	0.29
	Valence	IA-AMP Influenced Effect	0.67	0.31
	Valence	IA-AMP Non-Influenced Effect	0.24	0.25
4	Valence	IA-AMP Effect	0.28	0.24
	Valence	IA-AMP Influenced Effect	0.79	0.23
	Valence	IA-AMP Non-Influenced Effect	0.20	0.23
5	Valence	IA-AMP Effect	0.30	0.29
	Valence	IA-AMP Influenced Effect	0.73	0.31
	Valence	IA-AMP Non-Influenced Effect	0.21	0.22
6	Valence	IA-AMP Effect	0.30	0.28
	Valence	IA-AMP Influenced Effect	0.69	0.28
	Valence	IA-AMP Non-Influenced Effect	0.22	0.27
7	Valence	IA-AMP Effect	0.32	0.27
	Valence	IA-AMP Influenced Effect	0.62	0.31
	Valence	IA-AMP Non-Influenced Effect	0.22	0.25
8	Valence	IA-AMP Effect	0.29	0.27
	Valence	IA-AMP Influenced Effect	0.49	0.32
	Valence	IA-AMP Non-Influenced Effect	0.20	0.22

*Table 3.* Co-occurrence of consistent and inconsistent prime-evaluation responses with trials registered as influence-aware and non-influence-aware.

Experiment	Prime Consistent Response	Reported Influence Awareness	
		No	Yes
2	No	5672	1048
	Yes	6670	4250
3	No	3389	923
	Yes	4050	4310
4	No	3403	442
	Yes	3772	2607
5	No	4859	667
	Yes	5839	3539
6	No	6081	1115
	Yes	7675	4929
7	No	3400	922
	Yes	4097	2597
8	No	3336	1037
	Yes	3919	2796