**Supplementary Materials: Table 1 with annotations**

Each theoretical premise in Table 1 is based on previously published literature. In most cases there are many examples of each claim – our goal here is not to present an exhaustive list of all work that has made these claims, but to provide notable examples.

Table 1b. *Premises of past and present research on the AMP and their logical conclusions, with annotated references*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Research question/issue | | | |
|  | Implicitness | | Mechanism | Validity |
|  | Awareness | Intentionality |  |  |
| 1. Common theoretical premises | AMP effects are implicit in the sense of unaware 1, 2 | AMP effects are implicit in the sense of unintentional 3 | AMP effects are mediated by misattribution 4 | The AMP effect is a valid measure of evaluations 5 |
|  | - | - | Misattribution only occurs in the absence of awareness 6 | - |
| 2. Premises based on previous work | Previous literature on the AMP effect’s unawareness contains many statistical, methodological, and conceptual issues that weaken its conclusions 7 | Previous literature on the AMP effect’s unintentionality contains many  statistical, methodological, and conceptual issues that weaken its conclusions 7 | - | - |
| 3. Premises derived from the current work | Our results suggest that AMP effects are driven by awareness | - | Our results suggest that AMP effects are driven by awareness | AMP effects demonstrate structural invalidity in the majority of participants (i.e., when influence-awareness is not high) |
| Logical conclusions | The AMP effect is not implicit in the sense of unaware | There is no clear evidence to suggest that the AMP is implicit in the sense of unintentional | The AMP effect is not mediated by misattribution | The AMP effect is not a structurally-valid measure of evaluations for the majority of individuals |
| *Notes:* Premises and conclusions are arranged vertically in columns, so that conclusions follow from the premises above them. | | | | |

1 Bar-Anan & Nosek (2012): “the effect of attitudes in the AMP *depends* on people’s unawareness of that effect” (p. 1195)

2 Payne et al. (2005): “We suspect that if participants recognized that their judgment on any given trial was being influenced by the prime, they would be able to correct by simply giving the opposite response . . . the effect was difficult to control because participants did not believe they were experiencing it.” (p. 291)

3 Payne et al. (2005): “Those evaluations that bias performance nonetheless are taken to reflect automatic (i.e., unintentional) influences of attitudes.” (p.278)

4 Payne & Lundberg (2014): “The evidence to date suggests that misattribution is the best supported mechanism by which primes influence responses in the AMP.” (p.667)

5 This is tacitly endorsed by any paper which uses the AMP to study a domain on the assumption that it is a valid measure of attitudes, stereotypes, identities, etc. (agnostic to its implicitness), for example with 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), political beliefs (Payne et al., 2005; Kalmoe & Piston, 2013), or many clinically relevant behaviours such as eating disorders, non-suicidal self-injury, alcoholism, anxiety, depressive symptoms, or physical abuse of children (Fox et al., 2018; Görgen, Joormann, Hiller, & Witthöft, 2015; Jasper & Witthöft, 2013; McCarthy, Skowronski, Crouch, & Milner, 2017; Smith, Forrest, Velkoff, Ribeiro, & Franklin, 2018; Zerhouni, Bègue, Comiran, & Wiers, 2018; for a recent review see Payne & Lundberg, 2014).

6 Payne et al. (2005): “Only if the participant is unable to separate his or her reaction to the [prime] from his or her reaction to the symbol will any misattribution occur” (p.278; see also footnote 2).

7 See the introduction section of our article.