The “implicit bias” wording is a relic.

Let's move on and study unconscious social categorization effects.

Olivier Corneille & Jérémy Béna

UCLouvain

This commentary has been accepted for publication at Psychological Inquiry: DOI: 10.1080/1047840X.2022.2106754

Corresponding author: Olivier Corneille, UCLouvain, PSP IPSY, 10 Place du Cardinal Mercier, 1348 Louvain-la-Neuve, Belgium. Email address: olivier.corneille@uclouvain.be

Acknowledgment: This work was supported by an FRS-FNRS grant [grant T.0061.18] awarded to Olivier Corneille.

Abstract:

In their target article, Gawronski et al. (this issue) propose to define *implicit bias* as the unconscious effect of social category cues on behavioral responses. Based on this definition, they reason that the study of *biases on implicit measures* may have little relevance to *implicit biases* arising in everyday life. We applaud Gawronski et al.'s commitment to a precise definition of the *implicit bias* construct and we agree with their clear-cut separation of the two constructs. However, we contest the (scientific, educational, societal) value of ideas disseminated around *bias on implicit measures* when relating these measures to "unconscious biases." We also contest the relevance of retaining the delusive *implicit* terminology in social cognition research and research inspired by it. We propose an alternative terminology for the construct of interest - *unconscious social categorization effect* - and point to current limitations and potential ways forward for studying it. More generally, our commentary points to risks associated with conceptual and methodological flaws in implicit attitude and social cognition research.

The "implicit bias" wording is a relic.

Let's move on and study unconscious social categorization effects.

In their article, Gawronski, Ledgerwood, and Eastwick (2022; hereafter, GLE) explain why "implicit bias" (defined as the unconscious effect of social category cues on behavioral responses) should not be confused with "bias on implicit measures." We see much value in their clarification and agree with their bleak assessment of research on implicit tasks when they are said to measure "implicit bias" (hereafter "implicit measures of bias"), the most prominent of which is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998).

The article opens with a puzzling statement, though. GLE celebrate the educational value of "implicit measures of bias": "implicit measures of bias deserve enormous credit for providing a tool for the widespread dissemination of the idea that people can be biased without being aware of it" (p. 3). However, while reading their article, it becomes quickly clear (1) that "implicit measures of bias" have little conceptual consistency, and (2) that critical assumptions underlying their use and interpretation are unsubstantiated (e.g., the assumption that these tasks tap into unconscious mental contents or hold a special relation to associative learning). GLE also note that social cognition research has barely started to study the unconsciousness of category-driven biases beyond responses entered on computer keyboards.

It is an open secret that we do not clearly know how to interpret outcomes from "implicit measures of bias" (see, e.g., Fiedler, Messner & Bluemke, 2006). The managers of Project Implicit, the largest educational and research-oriented platform conventionally said to study "implicit biases" feature an honest disclaimer on the website of the platform: the

designers of the task, their promoters, and their associated institutions "make no claim for the validity" of their suggested interpretations of IAT scores (https://implicit.harvard.edu/implicit/takeatest.html). If we want to be honest about it, we do not know much which and when social behaviors are driven by an unconscious influence of social categories either.

If social cognition research relied on tasks and study settings that are detached from "implicit biases" (as GLE define them), then this begs the question of how accurate and profitable the education around this notion has been. As a case in point, introductory psychology textbooks generally fail to accurately portray the most prominent "implicit measure of bias" (Bartels & Schoenrade, 2021). We suspect that extra-academic education does not fare better. In the present commentary, we speculate on how we got here, we discuss how bad it can get when scientists conflate science with mere opinions, and we propose ways forward. We argue that strong research on "implicit bias" can finally see the light if drastic changes are implemented in social cognition research, starting with radical terminological changes.

How did we get here?

One may think of many reasons why we reached this unenviable situation. Our inclination is to favor cognitive explanations. In their early cognitive account of stereotyping, Tajfel and Wilkes (1963) suggested that assigning dual category labels to a continuous reality elicits an accentuation of the perceived differences between the two categories and an accentuation of the perceived similarities within each category. Likewise, by drawing heavily on dual-process models of the mind, the explicit/implicit opposition in attitude research may have accentuated the perceived differences between, and similarities within, mental contents, mental processes, and tasks assigned to these categories.

The *between*-category accentuation is apparent in social cognition research when it draws a sharp distinction between automatic versus non-automatic processes. It is also apparent when it denies properties of "implicit" tasks to their "explicit" counterparts. A clear illustration of this principle discussed by GLE is the assumption that "implicit" measures necessarily capture unconscious processes whereas "explicit" measures do not. It is difficult to understand why so many researchers have claimed and keep claiming the special relevance of "implicit" tasks to the study of (un)awareness. Even Greenwald and Banaji (2017) have clarified that "implicit" should be understood as "indirect" and should not be understood as "unconscious".

More recently, Greenwald and Lai (2020) opened the abstract of their recent *Annual Review* article by stating: "In the last 20 years, research on implicit social cognition has established that social judgments and behaviors are guided by attitudes and stereotypes of which the actor may lack awareness". This is a resounding "may": what implicit social cognition research has established is a mere conjecture. As far as the IAT is concerned, GLE argue that current evidence rejects its relevance for investigating these questions (see also Hahn, Judd, Hirsh, & Blair, 2014). They further explain that the reaction of surprise communicated by IAT takers when they receive feedback on their performance may be due to a "mismatch between the naïve metric used by participants to describe the extremity of their biases and the metric used by researchers to convert numeric IAT scores into verbal feedback" (pp. 15-16). A less charitable interpretation is that the feedback they receive is simply invalid (e.g., Fiedler et al., 2006). Other indirect measures offer no better alternatives for probing people's lack of awareness of their attitude.

The *within*-category accentuation is also apparent in this research. Various features of automaticity (e.g., unintentionality and unawareness) are conflated with each other (e.g.,

Fiedler & Hütter, 2014; Melnikoff & Bargh, 2018; Moors & De Houwer, 2006, 2007) within the "implicit" category. This accentuation bias may also have misled researchers into thinking that "measures of implicit bias" capture a distinct mental entity and share privileged relations with each other. There is no reason, however, why tasks such as a race IAT and a race Affect Misattribution Procedure should bear special relations with each other if their defining feature is to be "indirect", the latter of which has no process-related meaning; alternatively, should "implicit" be defined as "automatic," no boundary can be set to "implicit bias measures" because any task involves some automaticity component, including slow-paced evaluative self-reports (see Corneille & Hütter, 2020).

The road to hell is paved with good intentions

Let us be clear about this. Like GLE, we *believe* that people may be unaware of the effect of social category cues on their behavior, that millions of people repeatedly suffer from these biases, and that these biases should be fought. However, the confusion between science and social activism can be damaging. The first author of the present commentary was recently accused of being "Trumpian" by an anonymous reviewer because he was making a point also made by GLE: the IAT does not provide a valid measure of unconscious racial bias. Calling for stronger research should not be considered incompatible with a motivation for social justice, though. *On the very contrary*: improving methods, measures, and theory may help design more effective social interventions. Until now, studying category-driven biases on tasks said to be implicit has not been very helpful in this regard. GLE note that this research "should not be used to draw inferences about how we can reduce (implicit bias)" (p. 27).

Even more worrisome, flawed interpretations disseminated about these tasks have the potential to harm, in part by exonerating unconscious perpetrators of these biases. Daumeyer Onyeador, Brown, & Richeson (2019) recently showed that "implicit bias" narratives reduce

the perceived accountability of perpetrators and may damage efforts to implement institutional changes: “the communication of scientific studies detailing the effects of implicit bias on behavior may unwittingly increase tolerance for both discriminators and discrimination itself, even when the harm is quite severe.” (p. 5; see also Cameron, Payne, & Knobe, 2010). Likewise, De Houwer (2019) argued that defining “implicit bias” as a latent construct is “(…) likely to hamper attempts to reduce implicit bias in society, (…) because the metaphor of a hidden mental structure encourages the idea that implicit bias is a stable entity that is difficult to change and control (e.g., Sukhera et al., 2018).” (p. 836). Recently, Corneille and Mertens (2021) illustrated how flawed interpretations of performance on tasks said to be implicit may contaminate interventions in the social, consumer, and health and clinical psychology domains.

Another important concern is that research around "implicit measures of bias” may have distracted research away from blatant forms of discrimination by suggesting that they have now become a thing of the past. Yet, blatant prejudice and discrimination are still alive. In a recent article soberly titled “explicit bias,” Clarke (2018) argued that the strong focus on “implicit bias” “may be leaving the homeland of equality law norms against overt discrimination undefended.” (p. 509). In a similar vein, Banks and Ford (2008) noted: “Just as it misdescribes the IAT by eclipsing the ambiguity of its findings, the unconscious bias approach prompts people to acknowledge the persistence of the race problem by misdescribing it. The unconscious bias approach not only discounts the persistence of knowing discrimination, but it also elides the substantive inequalities that fuel conscious and unconscious bias alike.” (p. 1058).

Let’s move on!

GLE state that research on "implicit measures of bias" should receive "enormous credit" for having disseminated the idea that people are unconsciously biased. We believe this enthusiasm should be largely tempered. First, as these authors note, these tasks were not suited to the study of consciousness. Second, as GLE also note, we have barely started studying the unconsciousness of biases on social behaviors that matter. Third, as just discussed, it remains unclear whether disseminating this idea has proved socially beneficial or, alternatively, if it has created more problems than it has solved. Fourth, and this should not be underestimated either, the credibility of psychological research may ultimately suffer from having confused the public between science and mere opinions.

In our view, implicit social cognition research has much to offer when it comes to studying the phenomenon that GLE refer to as "implicit bias". Moving on, however, will require significant efforts and the courage to leave bad habits behind us. In the remainder of this commentary, we discuss three ways forward.

1. Leave "implicit bias" behind.

The flawed equation discussed by GLE originates in the confusing construct of "implicit." In their recent conceptual review, Corneille and Hütter (2020) explained that "implicit" can be understood as indirect (procedural definition), automatic (processual definition), or associative (mental representation definition). All three definitions are problematic: "indirect" is process-agnostic and so is of little use for the study of cognition; "automatic" conflates various automaticity features and does not allow to set any clear boundary between "implicit" and "explicit" measures; as to "associative," research has now ruled out the notion that "implicit" tasks are particularly relevant to the study of associative attitude learning (assuming the latter form of attitude learning exists, which is itself debated; Corneille & Stahl, 2019; Corneille & Mertens, 2020; De Houwer, 2018; Mitchell, De

Houwer, & Lovibond, 2009; Shanks, 2010). The co-existence of these definitions further adds to current misunderstandings and overgeneralizations.

We see no justification for retaining the highly confusing "implicit" wording in the social cognition vocabulary anymore, be it attached to "task," "measure," "attitude," or "bias" (for a thorough analysis, see Corneille & Hütter, 2020). As we sampled among recent definitions of "implicit bias," we could easily identify several divergent *effect-based* explanations for this notion. GLE define "implicit bias" as an unconscious effect of social category cues on behavioral responses. However, Fazio, Granados Samayoa, Boggs, and Ladanyi (2020) define "implicit bias" as the unconscious influence of attitudes on perceptions and judgments. To make things even more complicated, the latter definition overlaps with Wittenbrink, Corell, and Ma's (2019) definition of another construct: "implicit prejudice," which these authors define as "a negative evaluative predisposition toward a social category that impacts judgments and behaviors without awareness and/or intent" (p. 167).

All three definitions involve a mix of effects, operating conditions, and latent constructs, and different ones. The operating condition is unawareness in the first two definitions, and unawareness and/or unintentionality in the third. Latent constructs are "attitude" and "evaluative predisposition" in the second and third definitions, and the perceived social category in the first one. Finally, close definitions relate to two different notions in Wittenbrink et al. ("implicit prejudice") and Fazio et al. ("implicit bias"), and the same notion ("implicit bias") relates to two different definitions in GLE and Fazio et al.

The confusion is further deepened by the recent proposal for a fourth, *multi-faceted and effect-hybrid* definition of "implicit bias" as "the prejudicial judgments, decisions, and behaviors a person enacts, but there is something about the bias about which the person is unaware." (Park et al., 2020, p. 5). According to these authors, the "something about which"

may be here the unawareness of the attitude driving the bias (coined “implicit attitude”), the unawareness of the impact of the attitude on judgments and behaviors (coined “implicit impact”), or the unawareness of the origin of the influential attitude (coined “implicit basis”).

Even though GLE’s definition may be the best among its peers, it seems unlikely that collective agreement will be ever reached on whether “implicit bias” should refer to attitudes or social category cues, should be defined as an effect, must be unitary or can be multi-faceted, or can accommodate intentionality in place of awareness.

We understand that it will take a lot of efforts to leave out the “implicit” tag from the scientific discourse when a whole field of research and prominent platforms and labs have been named after it, and when the word has also now become massively popularized (as we check for this, “*implicit bias”* delivers 3.130.000 entries in a Google search). Yet, the choice exists. To illustrate, after establishing the lack of convergence of the two most prominent said measures of implicit self-esteem (ISE), Jusepeitis and Rothermund (2022) recently called: “for a shift in perspective and terminology that is long overdue in most of the research involving ISE. (…) To strengthen the intellectual rigor in the field, our recommendation is to refrain from using the term ‘implicit self-esteem’ without actual justification (…)” (p. 16).

Cultural evolution tends to ascribe more precision to words, not less; using the vague “implicit bias” terminology would be like using “line” for “teaspoon” when we want to name the latter (Morin et al., 2022). So, what is the terminological alternative for naming the “unconscious influence of social category cues on behavioral responses”, the latter of which is characterized by high precision but low practicability? In principle, “unconscious bias” could work. Unfortunately, this notion is characterized by high semantic vagueness, too. Therefore, we do not recommend substituting “unconscious bias” for “implicit bias.”

We would like to propose “unconscious social categorization effect.” We see many advantages with it. First, this is an effect-based terminology. Second, the wording goes away from the confusing “implicit.” Third, it also goes away from the semantically ambiguous and normatively loaded “bias.” Fourth, it specifies the operating condition of interest: unconsciousness. Fifth, the terminology can be easily adapted to look at other features of automaticity (e.g., “uncontrolled” , “unintentional”, or “efficient” social categorization effect). Sixth, the terminology signals that the effects of interest are situated in a broader literature concerned with “categorization effects.” In doing so, it builds bridges between social cognition and other disciplines, such as cognitive psychology and psychophysics.

A relevant illustration of an “unconscious social categorization effect” may be found in attractors fields theory, which predicts a race by emotion interaction in the processing of human faces (Corneille, Hugenberg, & Potter, 2007). Although this should be tested, it is unlikely that participants are aware of this categorical effect on their perceptual sensitivity. In addition, attractor fields do not apply only to social categories. They also apply to the perceptual treatment of e.g., birds and cars (e.g., Tanaka & Corneille, 2007). This way, disciplines and study materials are unified under a common theoretical umbrella.

2. Start studying unconscious social categorization effects.

GLE tell an inconvenient truth: we would better start studying how social categories may unconsciously influence significant social behaviors rather than pretending we are doing so by studying performance on heterogeneous computerized tasks that are irrelevant to the study of awareness. Admittedly, this is more easily said than done, though. GLE define “social category” at the perceiver-level. On a methodological level, this implies to establish that the influence was (1) unconscious and (2) that it was driven by the said *percept*. Demonstrating that the influencing percept is the one the experimenters assume, in addition to

ascertaining that its influence is unconscious, poses tremendous methodological challenges. GLE offer little methodological guideline for implementing this research. This comment also applies to other researchers who have recently called for more out-of-the-lab research on unconscious influences on social behaviors (e.g., Bargh & Hassin, 2022). We agree with these authors that this research is dearly needed. Yet, we should be humble and explain that we currently do not really know how to achieve it. Accordingly, we should perhaps refrain from stating as a fact that these *unconscious* influences are pervasive.

We would like to note here that researchers interested in "implicit biases" generally endorse two strong assumptions. The first assumption - the unawareness assumption - is that social category cues influence meaningful social behaviors unconsciously. The current authors believe that this assumption is plausible. However, in the absence of rigorous research on these questions (which GLE are calling for and we have just mentioned is challenging), this view should be cautiously held as a working assumption. More specifically, thorough unawareness tests should be applied whenever a presumably unconscious social categorization effect is investigated. Admittedly, we may as well posit that these effects are by default unconscious and that the burden of proof should be on the shoulders of those who contend otherwise (Bargh & Hassin, 2022). In either case, we need to inform these questions by relying on valid procedures, and those await further development (see next point).

The second assumption - the remedial assumption - is that becoming aware of one's biases is conducive to positive changes at the individual, collective, and institutional levels. This assumption feels good. However, it awaits evidence, too. We currently do not know to what extent making people acknowledge their category-driven biases is helpful (Hahn & Gawronski, 2019) and some research suggests that it can potentially make things even worse (e.g., Vorauer, 2012).

3. Rely on adequate and more precise analytic tools.

We would like to clarify here that we do not recommend moving away from research making use of tasks conventionally said to be implicit. We see value in these tasks for basic research, and potentially for applied and translational research, too. However, whereas these tasks can be useful for studying effects under less controllable, intentional or resource-dependent conditions, it remains unclear whether they allow examining the role of consciousness. This implies that these tasks should be adapted or that other tools should be used or developed for the study of implicit bias. Below, we consider each of these options.

We generally recommend reliance on procedures and analytic strategies allowing for a more nuanced interpretation of task performance. It has been known for a long time that performance on both self-reports and indirect measures are influenced by a variety of mental processes. Process dissociation procedures and several multinomial processing tree models have been developed to gain precision about how we think about task performance in social cognition (e.g., Calanchini, Rivers, Klauer, & Sherman, 2018; Hütter & Klauer, 2016) and have helped solve apparent inconsistencies in measurement outcomes (Calanchini, 2020). We encourage using these models and note that, to our knowledge, they have hardly started to consider a parameter capturing consciousness when examining social categorization effects.

An alternative to this approach is to ensure structural fit when making task comparisons. Hundreds of studies have compared effects on self-reports vs. indirect measures. Because these tasks simultaneously differ on a variety of features, however, it is impossible to clearly interpret a divergence in performance when one is found. One solution to this problem is to design and compare performance on two versions of the same task that vary only on the factor of interest. For instance, Payne, Burkley, and Stokes (2008) designed an intentional version of the Affect Misattribution Procedure, in which participants are requested to directly

evaluate the prime, the latter of which must be disregarded for judging the target in the unintentional version of the task. Hence, by achieving structural fit, the difference in performance can be attributed to intentionality. It is puzzling that only a very few studies looked at dissociation effects (e.g., in learning or behavioral predictions) while achieving structural fit. We encourage researchers to revisit past dissociation hypotheses and to test new ones by relying on structurally fitted procedures (for a recent example, see Béna, Melnikoff, Mierop, & Corneille, 2022). Developing structurally fitted tasks to investigate unconscious social categorization effects represents another opportunity for future research.

There is no reason, however, why the study of awareness should be limited to the use of existing procedures and analytic tools such as multinomial processing trees. How awareness should be studied depends on how it is defined and on the specific effect that is under investigation (e.g., Dienes & Seth, 2022). Both broadening and deepening our understanding of existing tools and contributing to the development of new tasks and procedures will advance research on unconscious social categorization effects.

Conclusion

We very much welcome GLE's recommendation to distinguish between "implicit bias" and "bias on implicit measures." However, consistent with Corneille and Hütter (2020), we strongly advise abandoning the "implicit" construct in this domain of research. Instead, we proposed an alternative terminology, unconscious social categorization effect, the many advantages of which we discussed. More generally, we share the bleak assessment that GLE make of research around "measures of implicit bias". We also fully endorse their call for more research on the unconscious influence that social category cues may have on behaviors that matter. This research effort will be challenging but is necessary if we seek to advance psychological science and help design effective intervention programs aimed to identify and

fight unconscious social categorization effects. In the meantime, it would make sense to refrain from stating as a known fact the pervasiveness of “implicit biases” (as, e.g., in Kang & Banaji, 2006). It would be wise to acknowledge the corrosive persistence of explicit biases. And it would be beneficial to critically assess the educational value of the information that research around “measures of implicit bias” has disseminated within and beyond academia for the last 25 years.

References

Banks, R. R., & Ford, R. T. (2008). (How) Does Unconscious Bias Matter: law, Politics, and Racial Inequality. *Emory Law Journal*, 58, 1053.

Bargh, J. A., & Hassin, R. R. (in press) Human unconscious processes in situ: The kind of awareness that really matters. In A. Reber & R. Allen (Eds.), *The cognitive unconscious*. New York: Oxford University Press.

Bartels, J. M., & Schoenrade, P. (2021). The Implicit Association Test in Introductory Psychology Textbooks: Blind Spot for Controversy. *Psychology Learning & Teaching*, 14757257211055200. https://doi.org/10.1177/14757257211055200

Béna, J., Melnikoff, D. E., Mierop, A., & Corneille, O. (2022). Revisiting dissociation hypotheses with a structural fit approach: The case of the prepared reflex framework. *Journal of Experimental Social Psychology*, *100*, 104297. https://doi.org/10.1016/j.jesp.2022.104297

Calanchini, J. (2020). How multinomial processing trees have advanced, and can continue to advance, research using implicit measures. *Social Cognition*, *38*(Supplement), s165-s186. https://doi.org/10.1521/soco.2020.38.supp.s165

Calanchini, J., Rivers, A. M., Klauer, K. C., & Sherman, J. W. (2018). Multinomial processing trees as theoretical bridges between cognitive and social psychology. *Psychology of Learning and Motivation*, *69*, 39–65. https://doi.org/10.1016/bs.plm.2018.09.002

Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, *23*, 272–289. https://doi.org/10.1007/s11211-010-0118-z

Clarke, J. A. (2018). Explicit bias. *Nw. UL Rev.*, *113*, 505.

Corneille, O., Hugenberg, K., & Potter, T. (2007). Applying the attractor field model to social cognition: Perceptual discrimination is facilitated, but memory is impaired for faces displaying evaluatively congruent expressions. *Journal of Personality and Social Psychology*, *93*(3), 335. https://doi.org/10.1037/0022-3514.93.3.335

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*(3), 212-232. https://doi.org/10.1177/1088868320911325

Corneille, O., & Mertens, G. (2020). Behavioral and Physiological Evidence Challenges the Automatic Acquisition of Evaluations. *Current Directions in Psychological Science*, *29*(6), 569–574. https://doi.org/10.1177/0963721420964111

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*, *23*, 161-189. https://doi.org/10.1177/1088868318763261.

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*, 103812. https://doi.org/10.1016/j.jesp.2019.04.010

De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin*, *13*(3), 1-21. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science*, *14*(5), 835-840. https://doi.org/10.1177/1745691619855638

Dienes, Z. & Seth, A. K. (2022). Conscious and unconscious mental states. In O. Braddick (Ed.), *Oxford Research Encyclopedia of Psychology*, Oxford University Press. https://oxfordre.com/psychology

Fazio, R. H., Granados Samayoa, J. A., Boggs, S. T., & Ladanyi, J. (in press). Implicit bias? What is it? In J. A., Krosnick, T. H. Stark, & A.L., Scott (Eds.). *The Cambridge Handbook of Implicit Bias and Racism*. Cambridge, England: Cambridge University Press.

Fiedler, K., & Hütter, M. (2014). The limits of automaticity. In J. Sherman, B. Gawronski, & Y. Trope (Eds.). *Dual process theories of the social mind* (pp. 497-513). New York: Guilford Press.

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European review of social psychology*, *17*(1), 74-147. https://doi.org/10.1080/10463280600681248

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, *117*, 522-559. https://doi.org/10.1037/pspa0000160

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861. https://doi.org/10.1037/amp0000238

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*, 419-445. https://doi.org/10.1146/annurev-psych-010419-050837

Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, J. F., Friese, M., ... & Wiers, R. W. (2021). Best research practices for using the Implicit Association Test. *Behavior research methods*, 1-20. https://doi.org/10.3758/s13428-021-01624-3

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (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

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143,* 1369. 1369–1392. https://doi.org/10.1037/a0035028

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, 27(1), 116-159. https://doi.org/10.1080/10463283.2016.1212966

Jusepeitis, A., & Rothermund, K. (2022). No elephant in the room: The incremental validity of implicit self-esteem measures. *Journal of Personality*. https://doi.org/10.1111/jopy.12705

Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of affirmative action. *Calif. L. Rev.*, *94*, 1063.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in cognitive sciences*, *22*(4), 280-293. https://doi.org/10.1016/j.tics.2018.02.001

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183-198. https://doi.org/10.1017/S0140525X09000855

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297- 326. https://doi.org/10.1037/0033-2909.132.2.297

Moors, A., & De Houwer, J. (2007). What is automaticity? An analysis of its component features and their interrelations. In J. A. Bargh (Ed.), *Social psychology and the*

unconscious: The automaticity of higher mental processes (pp 11-50). New York: Psychology Press.

Morin, O., Müller, T. F., Morisseau, T., Winters, J. (2022). Cultural Evolution of Precise and Agreed-Upon Semantic Conventions in a Multiplayer Gaming App. *Cognitive Science*, 46, e13113. https://doi.org/10.1111/cogs.13113

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*(1), 16–31. https://doi.org/10.1037/0022-3514.94.1.16

Park, B., Petty, R., E., Correll, J., Feldman Barrett, L., Hutchings,V., Langer, G., Otten,, S., Parker, C., & von Hippel, W. (2020). Report from the NSF conference. To appear in J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.) (2021). *The Cambridge Handbook of Implicit Bias and Racism*. Cambridge, England: Cambridge University Press.Shanks, D. R. (2010). Learning: From Association to Cognition. *Annual Review of Psychology, 61*(1), 273–301. https://doi.org/10.1146/annurev.psych.093008.100519

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*(2), 101-114. https://doi.org/10.1111/j.2044-8295.1963.tb00865.x

Tanaka, J. W., & Corneille, O. (2007). Typicality effects in face and object perception: Further evidence for the attractor field model. *Perception & Psychophysics*, *69*(4), 619–627. https://doi.org/10.3758/bf03193919

Vorauer, J. D. (2012). Completing the Implicit Association Test reduces positive intergroup interaction behavior. *Psychological Science*, *23*(10), 1168-1175. https://doi.org/10.1177/0956797612440457

Wittenbrink, B., Corell, J., & Ma, D. S. (2019). Implicit prejudice. In K. Sassenberg& M. L. W. Vliek (Eds). *Social psychology in action: evidence-based interventions for theory and practice* (pp 163-177). Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-13788-5