PULMONOLOGY

Beyond Relative Assessment of Implicit Bias: Toward Absolute Measurement in Clinical Decision-Making and Healthcare Equity Research

Dr. Sofia Almeida[1]*, Dr. Ricardo Mendes[1]

[1]Department of Psychiatry and Behavioral Sciences, Hospital Clínico San Carlos, Madrid, Spain

Abstract

A relative assessment of implicit biases is limited because it produces a combined summary evaluation of two attitudinal beliefs, while concealing the biases driving this evaluation. Similar limitations occur for relative explicit measures. Here, we will discuss the benefits and weaknesses of using relative versus absolute (individual/separate) assessments of implicit and explicit attitudes. The Implicit Association Test (IAT) will be the focal implicit measure discussed, and we will present a new perspective challenging the evidence that the IAT can only be utilized to measure relative, not absolute, implicit attitudes. Modeling techniques (i.e., Quad models) that can determine the separate biases behind the relative summary evaluation will also be considered. Accurately utilizing absolute implicit bias scores will enable academia and industry to answer more complex research questions. For implicit social cognition to maintain and expand its usefulness, we encourage researchers to further test and refine the measurement of absolute implicit biases.

Keywords: Attitudes, Implicit Association Test, Quad Modelling, ReAL model, Reaction Times

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 794913.

Introduction

For attitude researchers, the mid to late 90s ushered in fresh perspectives for describing automatic or implicit processes (Greenwald & Banaji, 1995). Crucially, two new tools to measure these complex processes were developed - Evaluative Priming (Fazio, Jackson, Dunton, & Williams, 1995) and the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) – and these tools accelerated research into implicit cognition, with the area being described a decade ago as "one of the liveliest and most active research areas in social psychology" (Payne & Gawronski, 2010, p.9). The IAT has undoubtedly become the most popular measure (Nosek, Hawkins, & Frazier, 2011), having four times more citations than Evaluative Priming, according to Google Scholar. For example, in June 2019, the initial IAT publication (year 1998) had 12,664 citations, while the Evaluative Priming publication (year 1995) had 3,152 citations. The academic popularity of the IAT, combined with prominent media exposures (Bartlett, 2017; Singal, 2017), and public attention through Project Implicit, which is the world's largest online virtual laboratory aimed at educating the public about implicit bias (implicit.harvard.edu), have resulted in the IAT rightly receiving intense scrutiny (Mitchell & Tetlock, 2017; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). However, strong criticisms have been met with persuasive rebuttals and fruitful future pursuits (see Gawronski, 2019; Jost, 2019).

Despite contributing important advancements in academia (e.g., clinical psychology, political science, & law, see Gawronski, Galdi, & Arcuri, 2015; Kang & Lane, 2010; Teachman, Clerkin, Cunningham, Dreyer-Oren, & Werntz, 2019), and industry (e.g., diversity training & marketing, see Bezrukova, Spell, Perry, & Jehn, 2016; Dimofte, 2010), we argue that the IAT, and implicit measures in general, will become less relevant and useful due to the need to tackle more complex research questions. This stagnation will be the result of over-dependence on and the ease

of using relative assessments of implicit biases, with their (current) superior psychometric properties (i.e., reliability and validity; Bar-Anan & Nosek, 2014).

Implicit processes, implicit attitudes, and implicit biases are used synonymously throughout this article. We define these terms as automatic evaluations that do not require an individual to deliberate or introspect on their feelings. We have refrained from using the term 'unconscious' when referring to implicit attitudes, due to the validity challenges of implicit measures detecting this construct (Hahn & Gawronski, 2019; Hahn, Judd, Hirsh, & Blair, 2014). Moreover, De Houwer (2019) proposed that implicit biases would benefit from being viewed as something that an individual does (implicitly biased behavior) rather than something that an individual possesses (a proxy for uncontrollable hidden/unconscious mental biases). In general, we are supportive of this reframing (see also De Houwer, Gawronski, & Barnes-Holmes, 2013). We define explicit attitudes as reflective evaluations that can involve deliberation or introspection.

Overall, this paper will focus on the advantages, as well as the various challenges/pitfalls, that the measurement of absolute attitudes, particularly absolute implicit attitudes, can have for researchers. Crucially, when we use the term "absolute" attitudes, we are not referring to something that is 'true' or is invariant across contexts or measures (see Gschwendner, Hofmann, & Schmitt, 2008; Schwarz, 2007; Stewart, Brown, & Chater, 2005), but rather we are simply referring to an evaluation towards only one attitude object (e.g., Black People, a person called Bob, Flowers). A "relative" attitude refers to an evaluation that involves contrasting one attitude object (i.e., Black People) with another (i.e., White People).

First, to orient the reader, we provide evidence showing the advantages relative explicit measures have over absolute explicit measures. In the second section, we emphasize the crucial reasons why researchers would find value in absolute implicit measurement. Third, we describe a

new method of decomposing the relative IAT scores. Next, we describe current modeling techniques used to determine absolute biases using IAT data, and following this section, we provide evidence showing how extraneous factors can impact absolute implicit measurement. Finally, we use the ideas from the previous sections to emphasize where absolute implicit measurement would be particularly suited (i.e., between subject assessments).

Explicit attitudes: Relative measures outperform absolute measures

A 7 point Likert scale (Likert, 1932) is often used to measure relative explicit attitudes (e.g., 1 = I strongly prefer White People to Black People, 4 = I like Black People and White People equally, 7 = I strongly prefer Black People to White People). Relative scales essentially give each respondent an anchor on which to base their judgments through the use of "to" in the previous examples (Goffin, Jelley, Powell, & Johnston, 2009; Wagner & Goffin, 1997). No anchors or comparisons are present when an absolute explicit measure is used (Taylor & Parker, 1964), such as when a feeling thermometer is used to gauge participants' attitudes towards an individual attitude object (e.g., Please rate how warm or cold you feel towards Black People: 0=coldest feelings, 5=neutral, 10=warmest feelings). A similar feeling thermometer could also be used to estimate absolute biases towards White People.

Although the Black and White People feeling thermometers are meant to be used to measure separate biases, placing them next to each other on the questionnaire likely results in respondents using both absolute explicit measures when making their judgment, essentially resulting in a relative judgment (Schwarz, 1999). Furthermore, a relative scale can be created using these absolute scales by calculating the difference between the two. To clarify, two White individuals may explicitly have no bias or preference in favor of White People over Black People.

On the feeling thermometer, White Person A could choose 5 for Black People and 5 for White People, while White Person B could choose 10 for Black People and 10 for White People. Relatively both these participants are expressing similar biases (neutral), but absolutely these responses are, of course, very different (A = neutral, B = positive). This example illustrates that compared to relative explicit scales, absolute explicit scales have increased measurement error between participants, which can result in reduced validity of absolute explicit scales (for a full discussion of this point see Olson, Goffin, & Haynes, 2007; Seligman, Swedish, Rose, & Baker, 2018).

To empirically illustrate the reduced variance that absolute explicit measures can explain, we analyzed explicit racial attitudes using the 2018 Race Project Implicit data (Xu, Nosek, & Greenwald, 2014). We correlated the relative Likert scale and the two feeling thermometers described above with a six-item Bayesian Racism Scale (BRS; Uhlmann, Brescoll, & Machery, 2010), and the strength of the correlation acted as the criterion variable. The BRS measures beliefs related to the appropriateness of discriminating against individuals based on stereotypes about their racial group, with higher scores indicating more Bayesian racist views. We hypothesized that those with more negative biases towards Black People will show higher Bayesian racism and that higher correlation coefficients will be shown for the relative rather than either of the two absolute explicit measures.

White participants who responded to all the questions were included in the analysis. A Fisher's r-to-z transformation was used and showed that the Black (r = -.22) and White (r = .03) People feeling thermometers showed significantly weaker associations with the BRS compared to the relative explicit Likert scale question (r = .33, zs > 7.82, df = 8422, ps < .001). Therefore, the higher correlation coefficient implies that the relative explicit measure has higher validity than the

absolute explicit scales. Similar findings were shown when explicit scales related to Bayesian racism were used (e.g., Modern Racism Scale). All the data are available here: https://osf.io/2gxn6/. Importantly, the absolute scores are still offering value by showing that feelings towards Blacks rather than Whites are more strongly related to the Bayesian racism scores.

There are various reasons why absolute explicit measures will show reduced reliability and validity compared to relative explicit measures. One reason is due to classical test theory, where a measure that incorporates two referents (i.e., Black People & White People) will be more reliable than a measure that only uses one referent (i.e., Black People; Crocker & Algina, 1986; see also Payne, Bettman, & Schkade, 1999). Another reason is due to the possibility that each respondent will use a different reference point, contrast category, or anchor on which to base their absolute response (see Olson, Goffin, & Haynes, 2007; Seligman, Swedish, Rose, & Baker, 2018). A final reason is because most judgments in life are relative (Parducci, 1968; Stewart, Chater, & Brown, 2006; Suls, Martin, & Wheeler, 2002). According to Social Comparison Theory (Festinger, 1954), relative/comparative judgments of social stimuli across various dimensions is a natural and continuous process engaged in by humans (e.g., Kruglanski & Mayseless, 1990). Furthermore, people are generally good at discriminating between stimuli, but less capable of identifying or estimating the magnitude of a stimulus (see Stewart, Brown, & Chater, 2005). Social cognitive and evolutionary perspectives have been proposed to explain why humans spontaneously engage in judgments using comparative rather than absolute assessments (see Goffin & Olson, 2011).

From a socio-cognitive perspective, there is value in having accurate assessments of one's attributes, and continuously comparing the self to other similar and dissimilar individuals can function as a method of self-enhancement and self-improvement (e.g., Buunk & Gibbons, 2007; Roese & Olson, 2007). Likewise, from an evolutionary perspective, prioritizing comparative over

absolute judgments is crucial for navigating complex social problems such as mate choice (choosing the "best" partner), social status (figuring out your place in a social hierarchy), self-protection (fighting or retreating from an aggressive competitor), and relationship maintenance (sticking with your current partner or finding someone better; Goffin & Olson, 2011; Kenrick et al., 2002).

Despite humans' spontaneous, effortless, and unintentional engagement in relative comparisons (Gilbert, Giesler, & Morris, 1995), absolute explicit items/measures are far more common than relative measurements in research (Goffin & Olson, 2011). This observation is likely due to the enhanced understanding that absolute items offer the researcher and the respondents (i.e., bias estimates towards the attitude object is not influenced by the anchor, and differential knowledge or exposure to the contrasting category will impact responses). However, in accordance with the correlational findings above, there is strong evidence across various disciplines in psychology (i.e., organizational, social, and personality) that relative explicit measures outperform absolute explicit measures regarding criterion-related validity (see Goffin et al., 2009; Goffin & Olson, 2011; Olson et al., 2007). Therefore, on the explicit level, it appears that relative measures are superior to absolute measures when predicting behaviors or outcomes, but regardless, absolute explicit measures are more common due to the enhanced specificity they offer researchers (i.e., is an attitude negative, neutral or positive towards an attitude object? – this absolute evaluation cannot be determined with a relative explicit measure).

To clarify, on the 7 point relative Likert scale (recoded to -3 to +3) and a relative score created using the race feeling thermometers, a mean score of 0.29 (SD = 0.78, neutral = 0) and 0.22 (SD = 1.61, neutral = 0) were shown respectively. Both scores indicate a significant pro-White/anti-Black bias, but with a small effect size, $ts(>348{,}627) > 80.78$, $ps < .001$, $d = 0.14$ –

0.39. We cannot decompose/unpack the relative Likert scale, but we can get absolute estimates using the feeling thermometers. These absolute scores indicate that the White participants have a significantly positive bias towards both Black People (M = 7.06, SD = 1.93, *neutral = 5.5*, $t(351{,}908) = 477.71$, $p < .001$, $d = 0.80$) and White People ($M = 7.28$, $SD = 1.93$, $t(351{,}908) = 545.38$, $p < .001$, $d = 0.93$), but their pro-White preferences were stronger. Likewise, Black participants show positive absolute biases towards both groups (Black People: $M = 8.50$, $SD = 1.82$, $t(58{,}446) = 400.07$, $p < .001$, $d = 1.65$; White People: $M = 6.44$, $SD = 2.31$, $t(58{,}362) = 94.44$, $p < .001$, $d = 0.40$), but they expressed more exaggerated preferences in favor of their own group and less positivity towards their out-group than White respondents, which others have previously documented (Howell, Gaither, & Ratliff, 2015; Nosek et al., 2007). As this example illustrates and where we will further elaborate below, absolute estimates clearly offer great value to researchers who appreciate specificity.

Implicit attitudes: **Relative versus absolute assessment**

Various absolute and relative implicit measures have been developed to assess implicit attitudes (for a review, see Gawronski & De Houwer, 2014). In contrast to the predominant use of absolute measures in assessing explicit attitudes (Goffin & Olson, 2011), relative measurement techniques are far more popular in assessing implicit attitudes (see Nosek et al., 2011). This popularity is primarily driven by the wide use of the IAT and the superior validity and reliability it has over other relative and absolute implicit measures (Bar-Anan & Nosek, 2014). The IAT is normally classified as a relative implicit measure mainly due to its structural design, as it assesses the associations between two categories and two valenced attributes in each critical block.

Take the Race IAT as an example. It consists of two critical blocks: in one block, the labels 'White People' and 'Good' share the same response key, and 'Black People' and 'Bad' share another response key (the congruent block for White participants). A trial involves a stimulus appearing at the center of the screen, which corresponds to one of the four labels, and the correct response must be made before moving onto the next trial. In the other critical block, the instruction is reversed, and the labels "White People" and "Bad" share the same response key, and "Black People" and "Good" share the same response key (incongruent block for White participants). Standard IAT scores are calculated by subtracting the average reaction times in the "congruent' block from the average reaction times in the "incongruent" block with higher scores indicating a pro-White/anti-Black implicit bias. In this calculation, the IAT score is inherently relative because the average reaction times to both the White People and the Black People are included in each critical block (see Greenwald, Nosek, & Banaji, 2003, for further details about the IAT *D*-score).

Although the benefits of relative measurement of implicit attitudes are likely similar to those of relative measurement of explicit attitudes, there are four primary reasons why we would want to break relative implicit (and explicit) attitudes into their component processes (i.e., evaluation of each attitude object separately or absolute attitudes):

(1) From a conceptual perspective, intergroup relations (i.e., Black People versus White People) are impacted by the dynamics between the separate evaluations of each group. For example, strongly positive evaluations towards White People and neutral evaluations towards Black People would manifest as pro-White/anti-Black biases on a relative measure, as would neutral evaluations towards White People and strongly negative evaluations towards Black People. These two forms of intergroup bias are clearly conceptually different, but relative scoring methods cannot distinguish between them. Therefore, separating attitudes into their component evaluations

could help confirm theories related to intergroup bias, which posit the primacy of positive ingroup evaluations (i.e., ingroup favoritism) over negative outgroup evaluations (i.e., outgroup derogation) (e.g., Greenwald & Pettigrew, 2014, cf., Riek, Mania, & Gaertner, 2006).

(2) Another crucial reason to decompose relative implicit scores is because they cannot be used to determine why an intervention that aims to increase or reduce implicit biases have their effect (see Lai et al., 2016). For example, if an intervention/diversity training was effective at reducing White participants' pro-White/anti-Black IAT *D*-scores, the observed change could reflect a reduction of ingroup favoritism (pro-White), a reduction of outgroup derogation (anti-Black), or a combinatorial effect. Only using absolute assessments can we uncover why an intervention was effective in the first place. This enhanced understanding will enable researchers to refine their interventions, by detecting biases that are most malleable or by targeting the specific bias that is primarily maintaining a negative evaluation (cf. Gladwin et al., 2015)

(3) Research on the shifting standard of stereotypes model (Biernat, 2003; Biernat, Manis, & Nelson, 1991) indicates that when subjective judgments are used, contrast (or null) effects to the stereotype are more probable (i.e., participants predict that White and Black applicants will perform similarly on an academic test when the response options ranged from “very poorly” to “very well”). However, when more objective judgments are used, assimilation to the stereotype is more probable (i.e., predicting higher scores for White than Black applicants when numerical scores are used; see Biernat, Collins, Katzarska-Miller, & Thompson, 2009; Biernat & Kobrynowicz, 1997).

“The shifting standard model assumes that, initially, there is an automatic, perceptual-level *assimilative* effect in terms of the mental representation of the targets; that is, consistent with the stereotype” (Biernat & Manis, 2007, p. 77). We would argue that since implicit measures are

measuring objective behavior (i.e., reaction time differences), rather than subjective judgments (i.e., preferences towards White versus Black People), the IAT is more likely detecting assimilative effects, while contrast (or null) effects will be more likely on explicit scales (see Axt, Ebersole, & Nosek, 2016). Moreover, contrast effects may be particularly pronounced for subjective absolute explicit estimates, while we believe that absolute implicit estimates will better align with stereotyping biases due to assimilative effects. See Mussweiler and Strack (2000) for related discussions on how selective accessibility, when using objective measures, can lead to assimilation effects, while a reference point/contrast category is crucial for contrast effects to occur when using subjective measures. Consequently, further research related to these points could be fruitful.

(4) Overcoming socially desirable responding has been a key reason for the popularity of implicit measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Although both relative and absolute explicit measures are susceptible to impression management, we suspect that absolute explicit measures will be even more impacted by this phenomenon. This susceptibility is because most participants can rationalize their relative preference as being primarily driven by an ingroup preference, while respondents are likely conscious of the negative ramifications of openly expressing strong negative biases towards groups. As the race evaluations in the previous section indicate, participants generally respond with a pro-ingroup/anti-outgroup relative explicit preference. But interestingly, negative absolute biases towards outgroups were not detected. Of course, these results are entirely possible and could be accurate attitudinal estimates, but through using absolute implicit measures, we can test if these scores diverge or converge with absolute explicit scores.

Still to be published results using the decomposed absolute IAT scoring technique, described in the next section, show that Black respondents' absolute explicit and implicit bias

estimates converge (both pro-White and pro-Black biases). In contrast, white participants scores diverge, such that explicitly they show pro-White and pro-Black preferences, but implicitly pro-White and anti-Black biases are shown (O'Shea, 2020a). Future research will need to determine why this divergence occurred. As already mentioned, social desirability concerns are an obvious contender, as well as the previous point related to contrast versus assimilation effects. Still, these findings may simply be a methodological artifact of the IAT, impacted by salience-asymmetries or recoding processes (see Meissner, Grigutsch, Koranyi, Müller, & Rothermund, 2019; Meissner & Rothermund, 2015b; Rothermund & Wentura, 2004). Yet regardless, these findings warrant further investigation, particularly due to the clear oppression Black People in the US and across the globe are still experiencing, leading to the rise of the Black Lives Matter movement (https://blacklivesmatter.com/about/).

Decomposing the relative IAT scores: Absolute IAT scores (*DD*-scores)

Unlike relative explicit scales, such as the Likert scale described above, the IAT can measure separate, absolute attitudes towards White and Black People. For example, a participant's absolute bias towards White People can be calculated by averaging the correct reaction time responses to 'White People' and 'Good', doing the same for 'White People' and 'Bad', and calculating the difference between these two scores. A similar calculation can also be performed for the Black People sorting task. Furthermore, another calculation can be performed on error trials across the congruent and incongruent blocks to test for convergent validity (i.e., calculating the relative difference between the number of errors on the "White People – Good" trials to "White People – Bad" trials). For full details describing and validating this method of decomposing the

relative IAT *D*-scores, see O'Shea, Glenn, Millner, Teachman, and Nock (in press). They call this method the decomposed *D*-scores or *DD*-scores for short.

As the previous calculations illustrate, absolute scores created using implicit measures involve a relative comparison ("White People – Good" versus "White People – Bad"). Indeed, even implicit measures specifically designed to measure absolute evaluations such as the Go/No-Go Association Test (GNAT; Nosek & Banaji, 2001) and Single Category/Target - IAT (Bluemke & Friese, 2008; Karpinski & Steinman, 2006), relatively compare reaction times or error rates, when an attitude object is associated with positively versus negatively valanced words. Similarly, absolute scores using Evaluative Priming (Fazio et al., 1995; i.e., an image appearing prior to classifying a word as positive or negative can facilitate or hinder accurate and quick responses) can be calculated by relatively comparing classifying responses to the target attitude object prime (e.g., Black or White People) with neutral primes (e.g., furniture). However, finding universally neutral primes is challenging, which is a similar challenge to finding an appropriate contrasting category (Penke, Eichstaedt, & Asendorpf, 2006). Regardless, we feel the benefits of accurately estimating absolute implicit biases will outweigh the costs of tackling these complicated issues. We would recommend researchers to always use a contrasting category, even with absolute implicit measures, where one is not required. This addition enables the assessment of relative attitudes if extraneous influences (e.g., faster responses to positive than to negative stimuli; Kuperman, Estes, Brysbaert, & Warriner, 2014; O'Shea, Watson, & Brown, 2016) impact the absolute implicit estimates.

Researchers have previously performed absolute bias calculations using IAT data (e.g., de Jong, Pasman, Kindt, & van den Hout, 2001; Gemar, Segal, Sagrati, & Kennedy, 2001), but such assessments fell out of use (cf. Gladwin et al., 2015). To understand why the pursuit of absolute

IAT scoring research ended, we first must delve into how implicit measures are typically validated. When validating explicit attitudinal measures, they are usually related to other explicit measures or behaviors (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018). To validate an implicit measure, directly testing the relationship it has with another implicit measure is less common, while correlating implicit measures with explicit measures is often used due to the cost-effectiveness (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). This method of validation is preferred, because implicit attitudes are less stable over time (e.g., Enkavi et al., 2019), and are more susceptible to extraneous influences such as distractions or executive control (Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010) than explicit measures.

Consequently, Nosek, Greenwald, and Banaji (2005) tested the effectiveness of the relative and the absolute IAT scores at predicting participants' corresponding relative and absolute explicit scores. Their criterion variable was the strength of the correlation. They proposed that the absolute scores are only valid if they show stronger correlations for the matched absolute pairs (e.g., the relationship of Black People implicit scores to Black People explicit scores) compared to the relative scores. Nosek et al. (2005) showed that the relative IAT scores always outperformed the absolute IAT scores regarding the criterion variable. However, as they indicated in a footnote, when the relative IAT scores were calculated with the same number of trials as the absolute IAT scores, no differences were observed between the criterion variables. Furthermore, the relative explicit question they used was conceptually more similar to the IAT then the absolute feeling thermometers, advantaging the relative scores due to the correspondence principle (Ajzen & Fishbein, 1977; see also Gawronski, 2019, Lesson 2; Irving & Smith, 2020).

In contrast to the observation that relative explicit scores normally outperform absolute explicit scores (see the explicit attitudes section above), it is quite remarkable that the absolute and

relative IAT scores are performing comparably (c.f. Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). We would argue this comparability is indicating that the absolute IAT scores are, in fact, detecting meaningful implicit biases towards a single attitude object. Indeed, Bar-Anan and Nosek (2014) showed that the IAT and the Brief IAT (Sriram & Greenwald, 2009) had good convergent validity when the absolute scores were used to detect known groups (e.g., Black respondents, Republican supporters); however, no signs of discriminant validity were shown. More recently, O'Shea and colleagues (in press) showed that the IAT *DD*-scores essentially performed comparably to the standard *D*-scores at correctly classifying suicide attempters from non-attempters, even though the *DD*-scores have half the number of trials.

In the race IAT, if we were to decompose the relative IAT scores and get a separate or absolute bias for Black People, it is normally assumed that the contrasting category (White People) will influence participants responding towards Black People. However, this assumption has never been tested using the IAT. Instead, evidence has shown that the contrasting category can change the relative IAT scores. Perhaps, it is conceivable that the absolute scores are stable, and it is only the alternating comparisons that are impacting the relative scores. For example, Robinson, Meier, Zetocha, and McCaul (2005) performed two IAT experiments, one where smoking was contrasted with non-smoking and another where smoking was contrasted with stealing. In the former condition, stronger pro-non-smoking/anti-smoking biases were shown, while in the latter condition, stronger pro-smoking/anti-stealing biases were observed. Due to the automaticity or fast responses required when using an implicit measure, it is possible that the contrasting category will have less of an impact on absolute implicit measurement than on explicit absolute measurement. Future research would benefit from re-analyzing Robinson et al.'s (2005) data, and that of other

studies like it to determine if the contrasting category significantly influences absolute metrics in the IAT.

Discouragingly, an implicit measure called the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010), which was developed to measure absolute implicit biases, indicates that the contrasting category does indeed impact the focal category absolute estimates (Hussey et al., 2016). However, similar to the IAT, when participants are completing the IRAP, they are required to remember two opposing associations (e.g., On this block, please respond as if Black People are Positive and White People are Negative). Therefore, implicit measures that have only one focal category during a block of trials may show more stable absolute estimates, regardless of the focal categories in the other blocks. For example, see the Multicategory IAT (Axt, Ebersole, & Nosek, 2014), which is a variant of the Brief IAT (Sriram & Greenwald, 2009). Further research should comprehensibly test the impact of the contrasting category in both relative and absolute implicit measures.

Researchers have challenged the idea that the relative IAT scores are capable of correctly and accurately detecting those with a neutral, weak, or strong implicit bias due to the arbitrary nature of the calculation (Blanton & Jaccard, 2006; Blanton, Jaccard, Strauts, Mitchell, & Tetlock, 2015). Similar critiques will also apply to absolute scores. However, when the IAT is used in domains that are not socially sensitive, such as political ideology, the relative IAT scores closely correspond to the slope gradient of explicit responses (Greenwald, Nosek, & Sriram, 2006). This evidence indicates that the relative IAT score of zero signifies a neutral or an ambivalent bias between the two categories (i.e., US republican or democratic candidate), while scores that deviate from this zero mark indicate a more positive attitude towards one candidate and/or a negative

attitude towards the other. It would be useful to determine if the absolute IAT scores in the political domain match the gradient of absolute explicit political attitudes.

Models used to estimate absolute implicit biases

Multinomial processing trees (MPTs: Batchelder & Riefer, 1999) are one prominent class of methods to separate attitudes into their component evaluations. MPTs have been applied to a wide variety of implicit measures (for reviews, see Erdfelder et al., 2009, and Hütter & Klauer, 2016) with the Quad Model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) and the ReAL model (Meissner & Rothermund, 2013) previously being applied to IAT data. MPTs can provide relatively more absolute estimates of evaluations of each of the attitude objects, typically included in relative implicit measures (i.e., MPTs reflect evaluations of one attitude object, with the contrasting category acting as a reference point). Consequently, the contrasting category is believed to impact the absolute score. Comparable to standard reaction time or error rate calculations, future research should test the impact of the contrasting category when using MPTs.

Quad modeling has shown that the effects identified by relative IAT scores mainly reflect differences in control-oriented processes (i.e., an individuals' ability to overcome a bias) rather than being solely due to attitude evaluations (e.g., Gonsalkorale, Sherman, & Klauer, 2009, 2014). A limitation of using the Quad model to estimate absolute evaluations is that it is typically configured to estimate the extent to which one attitude object is associated with positive attributes and, separately, the extent to which the other attitude object is associated with negative attributes. Other configurations are possible, such as modifying the Quad model to measure four distinct evaluations (e.g., White-Good; White-Bad; Black-Good; Black-Bad). However, to date, no published research has tested alternative Quad model configurations.

Nevertheless, the ReAL model is configured to estimate the extent to which each attitude object is associated with either positive, negative or neutral attributes. The ReAL model achieves this enhanced resolution of attitude object evaluations by using a short response time window in the IAT, similar to the GNAT, to ensure more errors are made. Unfortunately, the ReAL model cannot be applied to all the data that has been collected via Project Implicit as their IATs do not use any enforced response window.

Meissner and Rothermund (2013) provide an excellent illustration of the utility of absolute versus relative implicit attitude measurement. Initially, participants completed an IAT measuring their evaluations towards fruit versus chocolate. Then, in a seemingly unrelated task, they were offered either fruit or chocolate to snack on while watching a short movie. Relative IAT scores were unrelated to the amount of either food consumed by participants, but the ReAL model's absolute estimates towards chocolate predicted the amount of chocolate each participant ate. Given ongoing discussions about the predictive utility of the IAT (Kurdi et al., 2019; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013; Schimmack, 2019), Meissner and Rothermund's (2013) findings suggest that responses on implicit measures may indeed be related to behavior, but that the relative scores obscure these relationships.

Of note, Quad modeling and the ReAL model solely rely on error rates when calculating their estimates. Due to restrictive variance and ceiling effects, accuracy-based calculations are often less reliable (e.g. (Draheim, Mashburn, Martin, & Engle, 2019; cf. Gladwin et al., 2015). Recently, MPT models have been expanded to incorporate reaction time data (see Heck & Erdfelder, 2016; Klauer & Kellen, 2018) and therefore, may offer exciting new avenues for absolute implicit measurement. Future research would benefit from testing when, why, and for

whom do the Quad or ReAL model parameters converge or diverge with other absolute implicit attitude estimates such as the *DD*-scores.

Removing extraneous influences in implicit measures

Systematic response biases or other extraneous influences such as the category labels or the target stimuli used can impact relative implicit scores (e.g., Bluemke & Friese, 2006; Klauer & Mierke, 2005; Klauer et al., 2010; Meissner & Rothermund, 2015; Mierke & Klauer, 2003). These influences are expected to have an even more significant impact on absolute implicit scores (e.g., Positive Framing Bias in the IRAP, see O'Shea, Watson, & Brown, 2016). Consequently, greater efforts will need to be introduced to reduce these systematic biases or extraneous influences for absolute implicit assessment.

We will use a new implicit measure called the Simple Implicit Procedure (SIP: O'Shea et al., 2016; O'Shea, 2017) to give a concrete example of how systematic biases can be more problematic for absolute implicit scores than relative scores. If we measure implicit biases towards insects using the SIP, during one block of trials, participants must correctly use the response rule, 'Respond as if Insects are Positive'. When an insect (word/image) and a negative word appear on the screen, participants must press the 'No' keypress, and when an insect and a positive word appear, they must press 'Yes'. In the other block, 'Respond as if Insects are negative', the opposite keypresses must be used. Across these two blocks of trials, 'Yes' versus 'No' reaction time differences are calculated, one for the Insect positive word associations, and another calculation for the Insect negative word associations.

Humans are faster at responding with an affirming 'Yes' rather than a negating 'No' response option, especially when associating any attitude object with positively valenced words (

O'Shea, 2020b; see the density hypothesis by Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008, which can account for this affirming bias). Consequently, the affirming bias results in participants' absolute implicit biases being positively overestimated in the SIP than would occur if an affirming bias wasn't present (i.e., participants show a neutral implicit evaluation towards insects, rather than the expected negative bias, which would be predicted based on their explicit reports and findings from other implicit measures, see Meissner & Rothermund, 2013). However, if we also measure implicit biases towards flowers, similar affirming biases will be present for this attitude object, and when a difference/relative score is created, the affirming bias will get canceled out and hence, the relative SIP (and IRAP) scores will produce accurate relative estimates (O'Shea et al., 2016).

Researchers should endeavor to devise new implicit measures, as well as apply design alternatives, transformations, or modeling techniques to existing implicit measures, to remove any systematic biases detected. To clarify, a method to remove the systematic affirming bias occurring in the absolute measurement of the SIP would be to measure an individual's bias towards "neutral" made-up words (i.e., non-words) in place of flowers and insects, while using the same positive and negative attribute words that will appear in their subsequent Flower-Insect SIP. Similar to the neutral stimuli used in Evaluative Priming, these non-words may not be a universally neutral stimulus. Nevertheless, positive biases towards these non-words would be expected due to the affirming bias. Importantly, we can then use each individual's response biases towards the non-words to transform the scores on the Flower-Insect IAT, in order to remove the systematic affirming biases. Therefore, this transformation should result in more accurate estimates of individuals' absolute biases towards flowers and insects. Similar transformations could also be carried out if any systematic biases are detected with other implicit measures, especially absolute

implicit measures, such as the GNAT (Nosek & Banaji, 2001), and the Single Category/Target - IAT (Bluemke & Friese, 2008; Karpinski & Steinman, 2006).

Group versus individual estimates of absolute implicit biases

If one finds that the contrasting category or other extraneous influences significantly impact the absolute IAT scores, then we will have less confidence in the accurate interpretation of these scores (e.g., zero = neutral, above zero = more positive, below zero = more negative). However, great value can still be ascertained from decomposing the IAT scores or using models to determine absolute biases, especially when observing between-group differences. Firstly, the absolute scores will give researchers an enhanced understanding of the between-group mechanism, resulting in similar or different relative IAT scores. These observations are possible because the relative scores are essentially a composition of the two separate absolute scores, and any influences (i.e., contrasting category) should impact all groups equally.

Secondly, strong evidence has accumulated to highlight that the IAT is particularly suited to measuring context effects or situational level variance, rather than individual-level variance (Payne, Vuletich, & Lundberg, 2017; Vuletich & Payne, 2019). The strongest implicit and explicit correlations were observed when using larger areas/regions of analysis (e.g., US states) compared to smaller levels of analysis (e.g. US counties), with individual-level analysis performing the worst (Hehman, Calanchini, Flake, & Leitner, 2019). These findings are to be expected as reaction time difference scores were the unit of analysis (Draheim et al., 2019). To clarify, in the IAT, the within-subject variance is large between the congruent and incongruent blocks, while the overall individual-subject effect generally shows a much lower variance. As a result, the IAT has a stronger ability to distinguish between groups (intergroup) or regional differences, while

'paradoxically' it has a weaker ability to distinguish between individual-level (intragroup) differences (e.g., Hedge, Powell, & Sumner, 2018).

Using the IAT to predict individual-level variance has produced modest effects (Kurdi et al., 2019; Oswald et al., 2013; Schimmack, 2019), yet researchers have shown the IAT to be apt at predicting group (e.g., Glenn et al., 2017) and regional level variance (Hehman, Flake, & Calanchini, 2018; Nosek et al., 2009; O'Shea, Watson, Brown, & Fincher, 2020). All these prior studies have only used the relative IAT scores, but using absolute metrics will illuminate the biases behind certain observations, such as why some US states have lower implicit racial prejudice (i.e., reduced preference for Whites, increased preference for Blacks or both of these changes) or why suicide attempters differ from non-attempters on the Death IAT (i.e., a stronger "Me=Death", weaker "Me=Life" or both these biases, see O'Shea et al., in press).

Thirdly, just as individual respondents have expressed less explicit prejudice over time (see https://gssdataexplorer.norc.org/trends), implicit biases in the US from 2004 to 2016 have become more egalitarian on the Race, Skin-Tone and Sexuality IAT (Charlesworth & Banaji, 2019). Comparable to regional level differences, it would be highly beneficial for researchers to know precisely, using the absolute IAT estimates, why implicit biases in these domains are changing over time. Additionally, it is possible that certain groups' biases (e.g., political moderates) are changing faster than others (e.g., conservatives and liberals), and the absolute scores would enable researchers to determine why. This knowledge would be extremely valuable for understanding and predicting behavior, as well as developing large scale bias-reduction interventions to target specific groups or biases.

Conclusion

The study of implicit cognition has immensely benefited from tools such as Evaluative Priming and the IAT. Although various absolute implicit measures have been developed, generally they show weaker validity and reliability than relative implicit measures (Bar-Anan & Nosek, 2014). These weaker properties are due to various factors, such as human cognition being inherently relative when making evaluations (Goffin & Olson, 2011) and/or systematic response biases, which exert a greater impact on absolute implicit assessment (O'Shea et al., 2016). Research would benefit from detecting and developing methods to remove or mitigate these extraneous biases, to facilitate the detection of accurate estimates of absolute implicit biases.

Due to the lack of process purity in implicit measures (Conrey et al., 2005), Gawronski (2019, Lesson 6) recommended that findings from an implicit measure should be replicated using another implicit measure that is based on different underlying processes (e.g., IAT versus SIP). We propose to extend this recommendation to different analytic/measurement techniques using the same implicit measure. For example, showing that both the quad model and the IAT *DD*-scores show similar absolute metrics would be useful. However, if a divergence is detected, perhaps participants overcoming bias abilities could be used to explain the divergence. Due to clear limitation that the IAT has when measuring individual-level differences (Draheim et al., 2019; Hedge et al., 2018; Hehman et al., 2019; Schimmack, 2019), testing which assessment techniques (i.e., relative, absolute or modeling scores) are better related to behaviors of groups (e.g., suicide attempters versus non-attempters) or biases across regions (US states), are important future research pursuits.

To sum up, measuring absolute implicit biases will greatly expand the type and complexity of questions academics and industry can answer. Vast quantities of data have been gathered using

the IAT through the Project Implicit and Project Implicit Mental Health platforms. Therefore, applying the absolute assessment techniques discussed above to this data can tremendously expand our understanding of implicit processes, and how they impact society. Furthermore, detecting the reason for changes in implicit biases across regions and over time will be crucial for large scale policy interventions to ameliorate discrimination and intergroup tensions.

References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The Rules of Implicit Evaluation by Race, Religion, and Age. *Psychological Science*, *25*(9), 1804–1815. https://doi.org/10.1177/0956797614543801

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2016). An Unintentional, Robust, and Replicable Pro-Black Bias in Social Judgment. *Social Cognition*, *34*(1), 1–39. https://doi.org/10.1521/soco.2016.34.1.1

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*(3), 668–688. https://doi.org/10.3758/s13428-013-0410-6

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*(3), 527–542.

Bartlett, T. (2017, January 5). Can We Really Measure Implicit Bias? Maybe Not. *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Can-We-Really-Measure-Implicit/238807

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bezrukova, K., Spell, C., Perry, J., & Jehn, K. (2016). A Meta-Analytical Integration of Over 40 Years of Research on Diversity Training Evaluation. *Psychological Bulletin*, *142*(11), 1227–1274.

Biernat, M. (2003). Toward a Broader View of Social Stereotyping. *American Psychologist*, *58*(12), 1019–1027. (2003-10099-002). https://doi.org/10.1037/0003-066X.58.12.1019

Biernat, M., Collins, E. C., Katzarska-Miller, I., & Thompson, E. R. (2009). Race-Based Shifting Standards and Racial Discrimination. *Personality and Social Psychology Bulletin*, *35*(1), 16–28. https://doi.org/10.1177/0146167208325195

Biernat, M., & Kobrynowicz, D. (1997). Gender and Race-Based Standards of Competence: Lower Minimum Standards but Higher Ability Standards for Devalued Groups. *Journal of Personality and Social Psychology 72:544–57*.

Biernat, M., & Manis, M. (2007). Stereotypes and Shifting Standards: Assimilation and Contrast in Social Judgment. In *Assimilation and contrast in social psychology* (pp. 75–97). New York, NY, US: Psychology Press.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*(4), 485–499. https://doi.org/10.1037/0022-3514.60.4.485

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*(1), 27–41. https://doi.org/10.1037/0003-066X.61.1.27

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful metric of implicit prejudice. *Journal of Applied Psychology*, *100*(5), 1468–1481. https://doi.org/10.1037/a0038379

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*(2), 163–176. https://doi.org/10.1016/j.jesp.2005.03.004

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, *38*(6), 977–997. https://doi.org/10.1002/ejsp.487

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*. https://doi.org/10.3389/fpubh.2018.00149

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*(1), 3–21. https://doi.org/10.1016/j.obhdp.2006.09.007

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability From 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487. https://doi.org/10.1037/0022-3514.89.4.469

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887 ($44.

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*(1), 252–287.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347–368. https://doi.org/10.1037/a0014211

de Jong, P. J., Pasman, W., Kindt, M., & van den Hout, M. A. (2001). A reaction time paradigm to assess (implicit) complaint-specific dysfunctional beliefs. *Behaviour Research and Therapy*, *39*(1), 101–113. https://doi.org/10.1016/S0005-7967(99)00180-1

Dimofte, C. V. (2010). Implicit measures of consumer cognition: A review. *Psychology & Marketing*, *27*(10), 921–937. https://doi.org/10.1002/mar.20366

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508–535. (2019-15647-001). https://doi.org/10.1037/bul0000192

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472–5477. https://doi.org/10.1073/pnas.1818430116

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial Processing Tree Models: A Review of the Literature. *Zeitschrift Für*

Psychologie / Journal of Psychology, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. http://dx.doi.org/10.1037/0022-3514.69.6.1013

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Gawronski, Bertram. (2019). Six Lessons for a Cogent Science of Implicit Bias and Its Criticism. *Perspectives on Psychological Science*, *14*(4), 574–595. https://doi.org/10.1177/1745691619826015

Gawronski, Bertram, & 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* (Vol. 2, pp. 283–310). New York, NY: Cambridge University Press.

Gawronski, Bertram, Galdi, S., & Arcuri, L. (2015). What Can Political Psychology Learn from Implicit Measures? Empirical Evidence and New Directions. *Political Psychology*, *36*(1), 1–17. https://doi.org/10.1111/pops.12094

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology*, *110*(2), 282–289. https://doi.org/10.1037/0021-843X.110.2.282

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of Personality and Social Psychology*, *69*(2), 227–236. (1995-43669-001). https://doi.org/10.1037/0022-3514.69.2.227

Gladwin, T. E., Rinck, M., Eberl, C., Becker, E. S., Lindenmeyer, J., & Wiers, R. W. (2015). Mediation of cognitive bias modification for alcohol addiction via stimulus-specific alcohol avoidance association. *Alcoholism, Clinical and Experimental Research*, *39*(1), 101–107. https://doi.org/10.1111/acer.12602

Glenn, J. J., Werntz, A. J., Slama, S. J. K., Steinman, S. A., Teachman, B. A., & Nock, M. K. (2017). Suicide and self-injury-related implicit cognition: A large-scale examination and replication. *Journal of Abnormal Psychology*, *126*(2), 199–211. https://doi.org/10.1037/abn0000230

Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management*, *48*(2), 251–268. https://doi.org/10.1002/hrm.20278

Goffin, R. D., & Olson, J. M. (2011). Is It All Relative?: Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others. *Perspectives on Psychological Science*, *6*(1), 48–60. https://doi.org/10.1177/1745691610393521

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, *45*(2), 410–414. https://doi.org/10.1016/j.jesp.2008.11.004

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of Implicit Attitudes May Conceal Differences in Implicit Associations: The Case of Antiaging Bias. *Social*

Psychological and Personality Science, *5*(3), 271–278. https://doi.org/10.1177/1948550613499239

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

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.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Greenwald, Anthony G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test: Comment on Blanton and Jaccard (2006). *American Psychologist*, *61*(1), 56–61. https://doi.org/10.1037/0003-066X.61.1.56

Greenwald, Anthony G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, *69*(7), 669–684. https://doi.org/10.1037/a0036056

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Differential Stability: The Effects of Acute and Chronic Construct Accessibility on the Temporal Stability of the Implicit Association Test. *Journal of Individual Differences*, *29*(2), 70–79. https://doi.org/10.1027/1614-0001.29.2.70

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*(3), 1369–1392. https://doi.org/10.1037/a0035028

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, *23*(5), 1440–1465. https://doi.org/10.3758/s13423-016-1025-6

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General*, *148*(6), 1022–1040. https://doi.org/10.1037/xge0000623

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate Use of Lethal Force in Policing Is Associated With Regional Racial Biases of Residents. *Social Psychological and Personality Science*, *9*(4), 393–401. https://doi.org/10.1177/1948550617711229

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Howell, J. L., Gaither, S. E., & Ratliff, K. A. (2015). Caught in the Middle: Defensive Responses to IAT Feedback Among Whites, Blacks, and Biracial Black/Whites. *Social Psychological and Personality Science*, *6*(4), 373–381. https://doi.org/10.1177/1948550614561127

Hussey, I., Mhaoileoin, D. N., Barnes-Holmes, D., Ohtsuki, T., Kishita, N., Hughes, S., & Murphy, C. (2016). The IRAP Is nonrelative but not acontextual: Changes to the contrast category influence men's dehumanization of women. *The Psychological Record*, *66*(2), 291–299.

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

Irving, L. H., & Smith, C. T. (2020). Measure what you are trying to predict: Applying the correspondence principle to the Implicit Association Test. *Journal of Experimental Social Psychology*, *86*, 103898. https://doi.org/10.1016/j.jesp.2019.103898

Jost, J. T. (2019). The IAT Is Dead, Long Live the IAT: Context-Sensitive Measures of Implicit Attitudes Are Indispensable to Social and Political Psychology. *Current Directions in Psychological Science*, *28*(1), 10–19. https://doi.org/10.1177/0963721418797309

Kang, J., & Lane, K. (2010). Seeing Through Colorblindness: Implicit Bias and the Law. *UCLA Law Review*, *58*, 465–520.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16–32. https://doi.org/10.1037/0022-3514.91.1.16

Kenrick, D. T., Maner, J. K., Butner, J., Li, N. P., Becker, D. V., & Schaller, M. (2002). Dynamical Evolutionary Psychology: Mapping the Domains of the New Interactionist Paradigm. *Personality and Social Psychology Review*, *6*(4), 347–356. https://doi.org/10.1207/S15327957PSPR0604_09

Klauer, K. C., & Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, *82*, 111–130. https://doi.org/10.1016/j.jmp.2017.12.003

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: new evidence for a task-set switching account of the implicit association test effect. *Personality & Social Psychology Bulletin*, *31*(2), 208–217. https://doi.org/10.1177/0146167204271416

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology*, *63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Kruglanski, A. W., & Mayseless, O. (1990). Classic and Current Social Comparison Research: Expanding the Perspective. *Psychological Bulletin*, *108*(2), 195–208. https://doi.org/10.1037//0033-2909.108.2.195

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065–1081. https://doi.org/10.1037/a0035669

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., … Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A

meta-analysis. *American Psychologist*, *74*(5), 569–586. https://doi.org/10.1037/amp0000364

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016. https://doi.org/10.1037/xge0000179

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22 140*, 55–55.

Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., & Rothermund, K. (2019). Predicting Behavior With Implicit Measures: Disillusioning Findings, Reasonable Explanations, and Sophisticated Solutions. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02483

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45–69. https://doi.org/10.1037/a0030734

Meissner, F., & Rothermund, K. (2015a). A thousand words are worth more than a picture? The effects of stimulus modality on the Implicit Association Test. *Social Psychological and Personality Science*, *6*(7), 740–748. https://doi.org/10.1177/1948550615580381

Meissner, F., & Rothermund, K. (2015b). The Insect-Nonword IAT Revisited: Dissociating Between Evaluative Associations and Recoding. *Social Psychology*, *46*(1), 46–54. https://doi.org/10.1027/1864-9335/a000220

Mierke, J., & Klauer, K. C. (2003). Method-Specific Variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*(6), 1180–1192. https://doi.org/10.1037/0022-3514.85.6.1180

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a Poor Proxy for Utility: The case of implicit prejuidce. In S. O. Lilienfeld & I. D. Waldman, *Psychological Science Under Scrutiny* (Eds., pp. 164–195). New York, NY: John Wiley & Sons. https://doi.org/10.1002/9781119095910.ch10

Mussweiler, T., & Strack, F. (2000). The 'relative self': Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology*, *79*(1), 23–38. https://doi.org/10.1037/0022-3514.79.1.23

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*(6), 625–666. https://doi.org/10.1521/soco.19.6.625.20886

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*(2), 166–180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*(4), 152–159. https://doi.org/10.1016/j.tics.2011.01.005

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. https://doi.org/10.1080/10463280701489053

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

Olson, J. M., Goffin, R. D., & Haynes, G. A. (2007). Relative versus absolute measures of explicit attitudes: Implications for predicting diverse attitude-relevant criteria. *Journal of Personality and Social Psychology*, *93*(6), 907–926. https://doi.org/10.1037/0022-3514.93.6.907

O'Shea, B. A. (2017). *Advancing the applicability of absolute implicit measures : using the Simple Implicit Procedure (SIP) to measure responses to pathogen threats* (PhD, University of Warwick). University of Warwick, United Kingdom. Retrieved from http://webcat.warwick.ac.uk/record=b3140400~S15

O'Shea, B. A. (2020a). Convergence and divergence between absolute explicit biases and decomposed IAT scores. *Unpublished Manuscript*.

O'Shea, B. A. (2020b). *'Glorious', 'Garbage': Positive adjectives and negative nouns elicit faster responding in valence judgment tasks.* Unpublished Manuscript.

O'Shea, B. A., Glenn, J. J., Millner, A. J., Teachman, B. A., & Nock, M. K. (in press). Decomposing Implicit Associations about Life and Death Improves the Understanding and Prediction of Suicidal Behavior. *Suicide and Life-Threatening Behavior*.

O'Shea, B. A., Watson, D. G., Brown, G. D. A., & Fincher, C. L. (2020). Infectious Disease Prevalence, Not Race Exposure, Predicts Both Implicit and Explicit Racial Prejudice Across the United States. *Social Psychological and Personality Science*, *11*(3), 345–355. https://doi.org/10.1177/1948550619862319

O'Shea, B., Watson, D. G., & Brown, G. D. A. (2016). Measuring implicit attitudes: A positive framing bias flaw in the Implicit Relational Assessment Procedure (IRAP). *Psychological Assessment*, *28*(2), 158–170. https://doi.org/10.1037/pas0000172

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. https://doi.org/10.1037/a0032734

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, *108*(4), 562–571. https://doi.org/10.1037/pspa0000023

Parducci, A. (1968). The Relativism of Absolute Judgments. *Scientific American*, *219*(6), 84–93.

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going. In B. Gawronski & K. B. Payne, *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1–15). New York, NY: Guilford Press.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice. *Psychological Inquiry*, *28*(4), 233–248. https://doi.org/10.1080/1047840X.2017.1335568

Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring Constructed Preferences: Towards a Building Code. *Journal of Risk and Uncertainty*, *1–3*(19), 243–270. https://doi.org/10.1023/A:1007843931054

Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single-Attribute Implicit Association Tests (SA-IAT) for the Assessment of Unipolar Constructs. *Experimental Psychology*, *53*(4), 283–291. https://doi.org/10.1027/1618-3169.53.4.283

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, *10*(4), 336–353.

Robinson, M. D., Meier, B. P., Zetocha, K. J., & McCaul, K. D. (2005). Smoking and the Implicit Association Test: When the Contrast Category Determines the Theoretical Conclusions. *Basic and Applied Social Psychology*, *27*(3), 201–212. https://doi.org/10.1207/s15324834basp2703_2

Roese, N. J., & Olson, J. M. (2007). Better, Stronger, Faster: Self-Serving Judgment, Affect Regulation, and the Optimal Vigilance Hypothesis. *Perspectives on Psychological Science*, *2*(2), 124–141. https://doi.org/10.1111/j.1745-6916.2007.00033.x

Rothermund, K., & Wentura, D. (2004). Underlying Processes in the implicit association test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139–165. https://doi.org/10.1037/0096-3445.133.2.139

Schimmack, U. (2019). The Implicit Association Test: A Method in Search of a Construct. *Perspectives on Psychological Science*, 1745691619863798. https://doi.org/10.1177/1745691619863798

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105. https://doi.org/10.1037/0003-066X.54.2.93

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*(5), 638–656.

Seligman, L. D., Swedish, E. F., Rose, J. P., & Baker, J. M. (2018). An initial investigation of the use of comparative referents to assess social anxiety. *European Journal of Psychological Assessment*, *34*(6), 367–375. (2016-38442-001). https://doi.org/10.1027/1015-5759/a000349

Singal, J. (2017, January 11). Psychology's Favorite Tool for Measuring Racism Isn't Up to the Job. Retrieved 17 September 2019, from The Cut website: https://www.thecut.com/2017/01/psychologys-racism-measuring-tool-isnt-up-to-the-job.html

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology*, *56*(4), 283–294. https://doi.org/10.1027/1618-3169.56.4.283

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute Identification by Relative Judgment. *Psychological Review*, *112*(4), 881–911. https://doi.org/10.1037/0033-295X.112.4.881

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Suls, J., Martin, R., & Wheeler, L. (2002). Social Comparison: Why, With Whom, and With What Effect? *Current Directions in Psychological Science*, *11*(5), 159–163. https://doi.org/10.1111/1467-8721.00191

Taylor, J. B., & Parker, H. A. (1964). Graphic ratings and attitude measurement: A comparison of research tactics. *Journal of Applied Psychology*, *48*(1), 37–42. https://doi.org/10.1037/h0040880

Teachman, B. A., Clerkin, E. M., Cunningham, W. A., Dreyer-Oren, S., & Werntz, A. (2019). Implicit Cognition and Psychopathology: Looking Back and Looking Forward. *Annual*

Review of Clinical Psychology, *15*(1), null. https://doi.org/10.1146/annurev-clinpsy-050718-095718

Uhlmann, E. L., Brescoll, V. L., & Machery, E. (2010). The Motives Underlying Stereotype-Based Discrimination Against Members of Stigmatized Groups. *Social Justice Research*, *23*(1), 1–16. https://doi.org/10.1007/s11211-010-0110-7

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*(1), 36–49. https://doi.org/10.1037/0022-3514.95.1.36

Vuletich, H. A., & Payne, B. K. (2019). Stability and Change in Implicit Bias. *Psychological Science*, *30*(6), 854–862. https://doi.org/10.1177/0956797619844270

Wagner, S. H., & Goffin, R. D. (1997). Differences in Accuracy of Absolute and Comparative Performance Appraisal Methods. *Organizational Behavior and Human Decision Processes*, *70*(2), 95–103. https://doi.org/10.1006/obhd.1997.2698

Xu, K., Nosek, B., & Greenwald, A. (2014). Psychology data from the Race Implicit Association Test on the Project Implicit Demo website. *Journal of Open Psychology Data*, *2*(1). https://doi.org/10.5334/jopd.ac

Acknowledgments

We would like to express our gratitude to the Banaji lab, Kirsty Lee, Miao Quin, Anne O'Shea, Radhika Raheja, and Lianna Valdes, for feedback on this manuscript. The advice and guidance from Jeffery Sherman and two anonymous reviewers throughout the review process were also invaluable. This research was supported by an EU Horizon 2020 Marie Curie Global Fellowship (No.794913) awarded to B. O'Shea.