

# The Validity of the IAT and the AMP as Measures of Racial Prejudice\*

Katherine Clayton<sup>†</sup>      Jordan Horrillo<sup>‡</sup>      Paul M. Sniderman<sup>§</sup>

December 10, 2020

## Abstract

Implicit measures such as the Implicit Association Test (IAT) and the Affect Misattribution Procedure (AMP) have been hailed as promising methods for gauging subconscious or covert racial prejudice. However, their validity has been contested. Using data from the ANES 2008–2009 panel survey, we demonstrate that: (1) although they purport to measure the same thing, the relationship between the IAT and the AMP in a high-quality nationally representative sample of white respondents is substantively indistinguishable from zero; (2) both the IAT and the AMP classify an improbable proportion of white Americans as having more positive attitudes towards Black Americans than white Americans; and (3) substantial numbers of white Americans that the IAT and the AMP classify as free of prejudice openly endorse explicitly negative stereotypes of Black Americans.

**Word Count:** 3,423 (including footnotes and references)

**Keywords:** implicit bias, racial prejudice, implicit association test, affect misattribution procedure

---

\*We thank Ellen Chapin, Stephen Haber, Andy Hall, Shanto Iyengar, Rachel Lienesch, Greg Mitchell, Natalya Rahman, Cole Tanigawa-Lau, and Chenoa Yorgason for helpful comments.

<sup>†</sup>**Corresponding Author:** Ph.D. student, Department of Political Science, Stanford University, Email: kpc.14@stanford.edu, URL: <https://sites.google.com/view/kclayton>

<sup>‡</sup>Ph.D. student, Department of Political Science, Stanford University, Email: horrillo@stanford.edu

<sup>§</sup>Fairleigh S. Dickinson Jr. Professor in Public Policy and Professor in Political Science, Email: paulms@stanford.edu

# 1 Introduction

Prejudice can be explicit – palpable, blatant. But it also can be implicit – subtle, even unconscious (e.g., Devine 1989; Fazio et al. 1995; Fazio and Olson 2003; Greenwald and Banaji 1995).<sup>1</sup> In political science, two methods for assessing implicit prejudice have become standard – the Implicit Association Test (IAT) (Greenwald, McGhee, and Schwartz 1998) and the Affect Misattribution Procedure (AMP) (Payne et al. 2005).<sup>2</sup>

Research both upholding and critiquing the validity of the IAT and the AMP measures has been ongoing for more than two decades (Arkes and Tetlock 2004; Banaji, Nosek, and Greenwald 2004; Blaison et al. 2012; Blanton et al. 2009; Greenwald, Banaji, and Nosek 2015; Oswald et al. 2013; Payne and Lundberg 2014). This study takes a new approach. We assess the validity of the IAT and the AMP as measures of racial prejudice, specifically, by examining them together. We leverage the American National Election Study (ANES) 2008–2009 panel study. The ANES is the gold standard for nationally representative survey samples in political science. To our knowledge, the ANES panel study is the only study in which both the IAT and the AMP have been administered to the same respondents in a high-quality representative sample of the general population.

Our analysis addresses three questions that have not been asked before. The first is the relationship between the IAT and the AMP – how well does a score on one measure of implicit prejudice predict a score on the other measure of implicit prejudice? Consequently, how likely is it that an individual who is classified as prejudiced on one measure will also be classified as prejudiced on the other? The second question is to what extent, if any, do the IAT and the AMP overestimate the number of white respondents who have more positive attitudes toward Black Americans than white Americans? The third question is to what

---

<sup>1</sup>The concept of implicit attitudes, taken as a whole, has become a briar patch of distinctions. Greenwald and Banaji (2017), for example, equate “implicit” with “indirect” and not “unconscious.” But see Corneille and Hütter (2020) for a five-fold distinction: implicit-as-indirect, implicit-as-automatic, implicit-as-associative, implicit-as-indirect-and-automatic, and implicit-as-driven-by-affective-gut reactions.

<sup>2</sup>A third measure of implicit attitudes is Affective Priming (Fazio et al. 1986). The method employed – capturing affect based on response latency – is similar to the IAT. This measure has not been widely employed in studies of public opinion and electoral behavior.

extent, if any, are white respondents who openly express derogatory views of Black people classified as free of racial prejudice on the IAT or the AMP?

Our most important result is this. The IAT and the AMP both purport to measure racial prejudice. If they measure the same thing, scores on one should be highly correlated with scores on the other. In fact, the correlation between the IAT and AMP is substantively indistinguishable from zero. Moreover, both measures classify an improbably large number of white Americans as having a more favorable opinion of Black Americans than of white Americans. Finally, our results show that both the IAT and the AMP classify as free of prejudice large numbers of white Americans who rate Black people as less intelligent and lazier than white people.

The paper proceeds as follows. First, we review previous research on implicit prejudice and political behavior. Second, we summarize how the IAT and the AMP were administered in the 2008–2009 ANES panel study. Third, we present the main results. Finally, we call attention to limitations of our data and analyses and note some implications of our findings.

## 2 Implicit Prejudice and Political Behavior

The IAT and the AMP can be used to measure reactions to an array of objects. Our concern is only with their use to assess racial prejudice in studies of public opinion and electoral behavior.<sup>3</sup> For example, when Barack Obama won the 2008 presidential election, Greenwald et al. (2009) found that both the IAT and the AMP predicted vote choice independently of explicit racial prejudice measures and ideology (see also Messing, Jabon, and Plaut 2016; Pasek et al. 2009; Payne et al. 2010). Similarly, Kam (2007) and Pérez (2010, 2016) have demonstrated evidence of prejudice against Hispanics using these methods. Other studies have argued that implicit attitudes, as measured by the IAT and/or the AMP, are strong predictors of vote choice for undecided voters in particular (Lundberg and Payne 2014; Ryan

---

<sup>3</sup>The reliability and hence the validity of both the IAT and the AMP vary dramatically, depending on the specific object that they attempt to assess (Gawronski et al. 2017).

2017; but see Frieze et al. 2016).

For studies of prejudice and electoral behavior, a natural focus is comparisons of the relative explanatory power of implicit and explicit measures of prejudice. These studies find that explicit attitudes play a more powerful role than implicit attitudes in shaping vote choice. Ditonto, Lau, and Sears (2013) find that traditional explicit measures of racism are much stronger predictors of racial and ethnic attitudes than estimates obtained from the AMP. Finn and Glaser (2010) demonstrate that the AMP predicts vote choice, but that the relationship between the AMP and vote choice disappears when covariates including emotions toward the candidates are taken into account. Segura and Valenzuela (2010) compare explicit and implicit measures of prejudice in the 2008 ANES cross-sectional study and find that the AMP does not add explanatory power above and beyond standard models of vote choice. Likewise, using three nationally representative data sets and also employing the AMP, Kalmoe and Piston (2013) find little evidence of implicit prejudice effects. Finally, Kinder and Ryan (2017) examine the IAT implemented in the 2008–2009 ANES panel study and the AMP used in the 2008 ANES cross-sectional study and find that implicit measures do not improve over explicit measures in predicting outcomes such as presidential vote choice, presidential approval ratings, or racial policy attitudes.

The objective of this study is to take the next step. Rather than just comparing implicit and explicit measures of prejudice, we compare the implicit prejudice measures themselves.

### **3 Data and Methods**

The 2008–2009 ANES panel study presents a unique opportunity to investigate how responses to the IAT and the AMP are related to one another in a representative sample of white respondents in the general population.

The ANES panel used a brief version of the IAT (BIAT), specifically developed for use when time is at a premium (Sriram and Greenwald 2009). The methodology of the shorter

and longer versions of the IAT is identical. In both, respondents are instructed to press a keyboard key as quickly as they can after seeing one of four different kinds of text or visual stimuli on a screen (a Black person’s face, a white person’s face, a positive word, or a negative word) in a series of repeated trials. Specifically, they are instructed to press the same key for white faces and for positive words and another key for anything else, or the same key for Black faces and for positive words and another key for anything else. The next round alternates, so participants must classify white faces with the positive category and Black faces with the negative category. Based on the difference in response times between white-good, Black-bad, and white-bad, Black-good, a D score is calculated on a scale of  $-2$  to  $2$ , where  $-2$  indicates maximum preference for Black people over white people and  $+2$  maximum preference for white people over Black people.

In the AMP, respondents are first shown a picture of a Black person’s face or a white person’s face on a screen for a fraction of a second, followed by a picture of a Chinese character displayed for a longer time. They are then asked to say whether the Chinese character appeared pleasant or unpleasant to them. Crucially, they are reminded that the photographs they saw prior to the Chinese character might bias their answers and are specifically instructed to guard against this. The resulting AMP scores are calculated on a scale that ranges from  $-1$  to  $1$ , where  $-1$  indicates that respondents classify all characters preceded by a Black person’s face as pleasant (maximum pro-Black preference) and those preceded by a white person’s face as unpleasant, and  $1$  indicates the opposite.

The AMP was administered online during waves nine and ten of the panel study; the IAT in wave nineteen (the fourth section of this paper examines concerns that this separation in time raises). About 10% of respondents ( $N = 158$ ) evaded valid AMP measurement by selecting either “unpleasant” or “pleasant” after every profile they viewed. Consistent with standard practice (see Kinder and Ryan 2017; Ryan 2017), these respondents are excluded from our analysis. Following standard practice for the IAT (e.g., Greenwald et al. 2009), we also excluded respondents who responded too rapidly, too slowly, or had an error rate above

35% on the IAT (7% of the full sample;  $N = 105$ ). The distinctive feature of our study is that the same respondents took both the IAT and the AMP. The final sample size for our main analysis of white respondents who completed both implicit measures is 1352.

## 4 Results

Both the IAT and the AMP purport to measure racial prejudice. Just so far as they do, scores on one will predict scores on the other well. If either one or both are not valid measures of implicit prejudice, scores on one will predict scores on the other poorly or not at all.

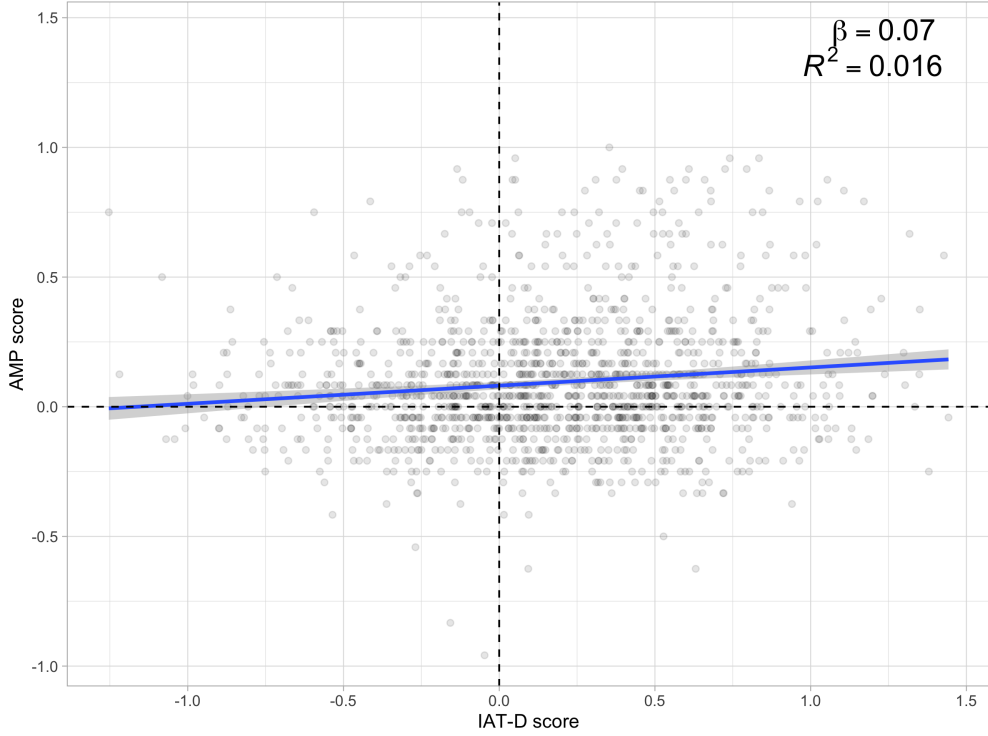
Figure 1 plots white respondents' IAT-D scores on the x-axis and their AMP scores on the y-axis ( $N = 1352$ ) in the 2008–2009 ANES panel study. For each implicit measure, higher values indicate a positive preference for white people over Black people; conversely, negative values indicate a positive preference for Black people over white people, and zero indicates indifference. The blue line is an OLS regression line and the gray shaded area around the line is the 95% confidence interval. The quantity of interest is the magnitude of the relationship between the IAT and the AMP. As Figure 1 shows, there is virtually no connection between them ( $b = 0.07$ ,  $R$ -squared = 0.016). A one-unit increase in IAT-D score (scaled  $-2$  to  $2$ ) is associated with an increase of less than one-tenth of a point in AMP score (scaled  $-1$  to  $1$ ). The relationship between the two main measures of implicit prejudice is substantively indistinguishable from zero.<sup>4</sup>

A necessary consequence of the virtually non-existent relationship between the two measures of implicit prejudice is misclassification. Figure 2 plots AMP scores conditional on IAT-D scores. If a respondent scores to the right of the zero point on the x-axis – i.e., exhibits some signs of implicit prejudice – she should score above the zero point on the y-axis,

---

<sup>4</sup>It is a standard practice among many researchers who use implicit measures to interpret statistical significance as being substantively meaningful (e.g., Greenwald, Banaji, and Nosek 2015). We do find that the association between the IAT and the AMP is statistically significant ( $p < .05$ ), but we view the magnitude of the regression coefficient – 0.07 – as more telling. An individual's score on the IAT tells us virtually nothing about her score on the AMP, and vice versa.

Figure 1: Relationship between IAT-D scores and AMP scores is virtually non-existent.



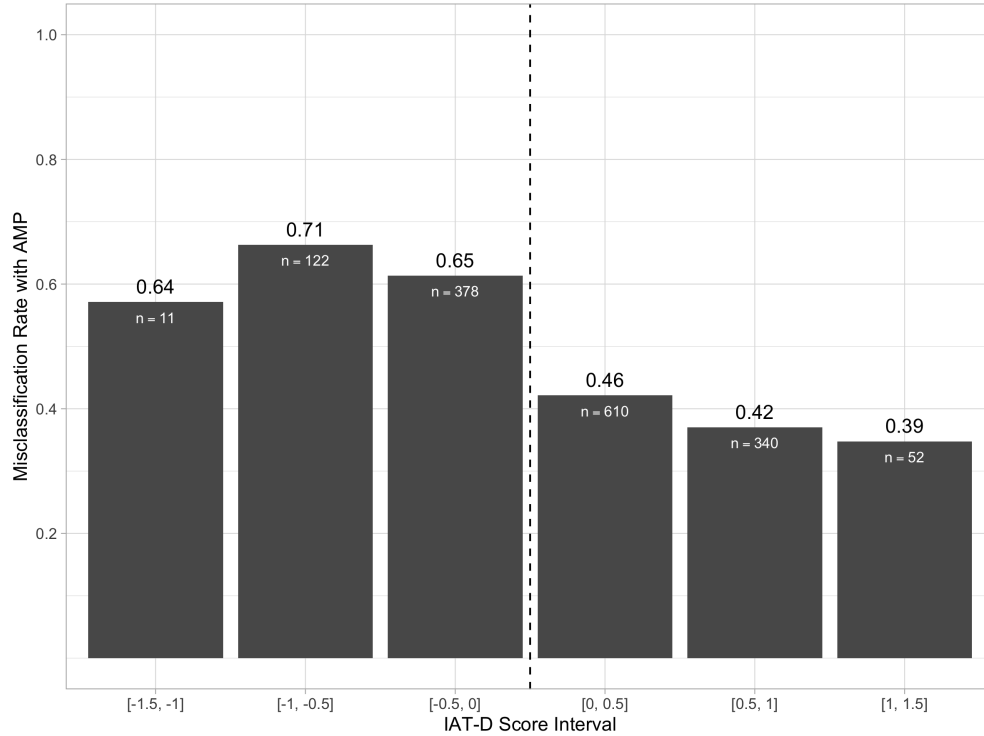
$N = 1352$  white respondents in the 2008–2009 ANES panel study. IAT scores are measured on a  $-2$  to  $2$  scale, and AMP scores on a  $-1$  to  $1$  scale (higher values indicating higher prejudice against Black people). The blue line is an OLS regression line and the gray shaded area is the 95% confidence interval.

or at the least not below it. Conversely, if she scores to the left of the zero point on the x-axis – i.e., does not show evidence of implicit prejudice, she should score below the zero point on the y-axis, or at the least not above it.

Figure 2 groups white respondents across intervals of 0.5 based on their IAT-D score.<sup>5</sup> The dashed line in the center of the plot separates these groups between those classified as prejudiced against Black people (right-hand side) and prejudiced against white people (left-hand side) on the IAT. For each interval of IAT response, the misclassification rate plotted on the y-axis represents the proportion of respondents in that group who are classified as prejudiced against Black people based on one implicit measure but not the other. This could

<sup>5</sup>We present the analogous figure that groups respondents across intervals of their AMP score in the Online Appendix (Figure B.1).

Figure 2: Misclassification based on the AMP and the IAT is pervasive across levels of implicit prejudice.



$N = 1352$  white respondents in the 2008–2009 ANES panel study. IAT scores are measured on a  $-2$  to  $2$  scale, and AMP scores on a  $-1$  to  $1$  scale (higher values indicating higher prejudice toward Black people). Respondents are grouped across intervals of their IAT-D score. The gray bars represent the proportion of respondents in each group who are classified as prejudiced against Black people according to the IAT but not the AMP, with sample sizes for each group included in the bars.

result from either a positive IAT-D score and a negative (or zero) AMP score, or a negative IAT-D score and a positive (or zero) AMP score. Sample sizes for each group are included within the bars.

Figure 2 reveals that the misclassification rates are the lowest for the most implicitly biased respondents on the IAT, but lowest is not a synonym for low. Even the most prejudiced individuals on the IAT are classified as non-prejudiced almost half of the time on the AMP. Misclassification is the highest for respondents classified as non-prejudiced on the IAT. Two in every three that the IAT classifies as free of implicit prejudice, the AMP classifies as exhibiting implicit prejudice. The problem may lie with the IAT or the AMP or both.



The strength of implicit measures of prejudice is asserted to be their ability to identify people who do not endorse explicit expressions of prejudice, presumptively because they want to present themselves in a socially desirable light and/or because they are not aware of being prejudiced, but are prejudiced nonetheless. If this is true, implicit measures of prejudice should more comprehensively identify members of a majority group who are prejudiced against a minority group than explicit measures.

Figure 3 shows density plots of white respondents’ IAT-D scores (left) and AMP scores (middle) in the ANES panel and, for cross-validation, the results from the AMP administered in the 2008 ANES cross-sectional study (right). The shaded area shows the proportion of respondents who are classified as more favorable toward Black people than white people. For both the IAT and the AMP in the panel, it is 34 percent. Similarly, for the AMP in the cross section, it is 30 percent (excluding those who failed to provide a valid AMP response as in the panel), indicating that respondents who took the AMP in the 2008 ANES cross-sectional study behaved similarly to those in the panel. These results contradict common sense and systematic evidence. 9.7 percent of white respondents rate their feelings toward Black people as warmer than their feelings toward white people in the 2008 ANES cross-sectional study. Moreover, in the 2008 ANES cross-sectional study, only 3.6 percent of white respondents rate Black people as harder working, and just 2.1 percent as more intelligent, than white people.<sup>6</sup>

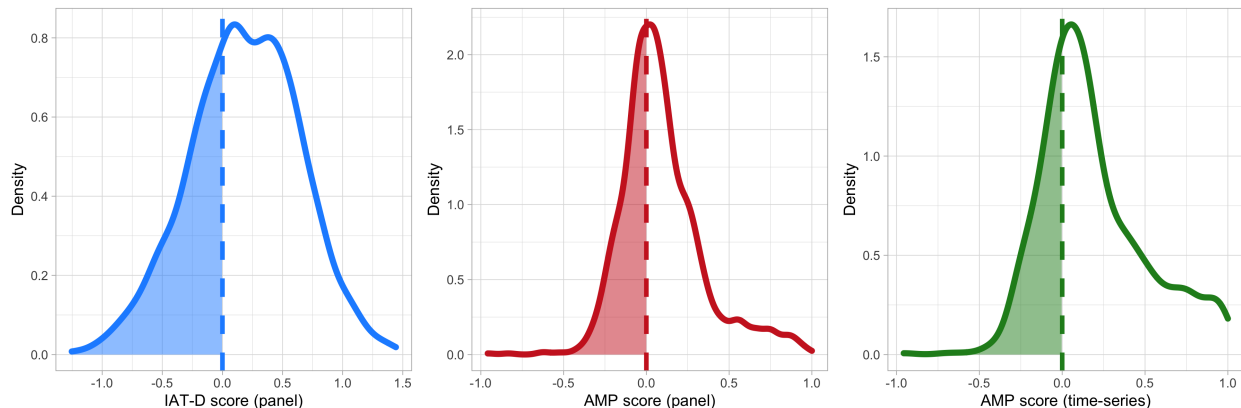
Table 1 therefore reports the percentages of white respondents who openly express derogatory evaluations of Black people *but are nonetheless classified by the IAT and the AMP as not prejudiced*.<sup>7</sup> The first row under the column headers includes respondents who rate most Black people as “lazier” than white people; the second row, those who rate most Black people as “less intelligent at school” than white people; and the third row, the those who

---

<sup>6</sup>The feeling thermometer question and the lazy/intelligent stereotype questions used in the ANES panel study were dissimilar to those standardly used in the ANES cross-sectional studies. We examine those used in the panel in more detail in Table 1; see the Online Appendix for more details on question wording.

<sup>7</sup>This includes both respondents who give Black people a more positive evaluation than white people or evaluate Black and white people equally favorably according to the implicit measures.

Figure 3: Both the IAT and the AMP overstate white respondents' favorability toward Black people.



$N = 1352$  white respondents in the 2008–2009 ANES panel study;  $N = 894$  white respondents in the 2008 ANES cross-sectional study. Each plot shows the density of scores on implicit measures of prejudice across all respondents. The dashed line is the zero line (indicating no prejudice) and the shaded area to the left of the line shows the fraction of white respondents who appear to be more favorable toward Black people than white people on each implicit measures.

feel “cooler” toward Black people than white people. As Table 1 shows, among those who describe Black people as lazier than white people, 23 percent are classified as free of prejudice on the IAT in the panel, 33 percent are classified as free of prejudice on the AMP in the panel, and 20 percent are classified as free of prejudice on the AMP in the 2008 ANES cross-sectional study. The comparable numbers for white respondents who say Black people are less intelligent than white people are 26 percent, 35 percent, and 19 percent. Finally, among those who feel “cooler” toward Black people than white people, 27 percent, 36 percent, and 23 percent in the IAT (panel), AMP (panel), and AMP (cross section), respectively, are classified as free of implicit prejudice.

It is not immediately obvious how to reconcile these results with the claim that implicit measures of prejudice, because they are free of social desirability bias or assess unconscious processes, provide a more comprehensive tally of majority members who are prejudiced. On the contrary, they classify as non-prejudiced large numbers of white respondents who state

Table 1: Significant proportions of white respondents who openly express prejudice toward Black people are classified as free of implicit anti-Black prejudice.

	<b>Non-prejudiced, IAT (panel)</b>	<b>Non-prejudiced, AMP (panel)</b>	<b>Non-prejudiced, AMP (x-section)</b>
Rate Black people lazier than white people	23%	33%	20%
Rate Black people less intelligent than white people	26%	35%	19%
Feel cooler toward Black people than white people	27%	36%	23%

Explicitly prejudiced white respondents only are included. In the first row,  $N = 225$  white respondents in the 2008–2009 ANES panel study and  $N = 410$  in the cross-sectional study rate Black people as lazier than white people; in the second row,  $N = 339$  in the panel and  $N = 361$  in the cross section rate Black people as less intelligent than white people; in the third row,  $N = 254$  in the panel and  $N = 297$  in the cross section feel cooler toward Black people than white people. Respondents with missing data for any of the three explicit measures we examine are excluded from the full set of white respondents ( $N = 77$  in the panel;  $N = 27$  in the cross section). The values in each cell are the percentages of these explicitly prejudiced respondents who are coded as non-prejudiced based on their IAT D-score in the panel (second column), AMP score in the panel (third column), or AMP score in the cross section (fourth column). See the Online Appendix for more details on differences in question wording for the explicit measures of prejudice in the panel and the cross section.

that Black people are inferior to white people – less intelligent, lazier.

## 5 Discussion

To our knowledge, this is the first study to assess the performance of the IAT and the AMP, administered to the same respondents, in a high-quality representative sample of white respondents in the general population. We want therefore to call attention to the only other study that we are aware of that has administered the IAT and the AMP to the same respondents, albeit in a sample that is highly unrepresentative in many respects, among

them significantly skewed in favor of politically liberal respondents (Greenwald et al. 2009).<sup>8</sup>

The study’s focus was the relationship between implicit attitudes and vote choice in the 2008 presidential election, but the authors’ use of both measures allows us to compare their results with ours. In the high-quality ANES panel study sample, the zero-order correlation between the IAT and the AMP is 0.12. In the Greenwald et al. (2009) sample, based on respondents who chose to go the Project Implicit site in order to learn about prejudice, the zero-order product moment correlation between the two measures is 0.22.<sup>9</sup> The IAT and the AMP thus share less than five percent of the variance in the Project Implicit sample,<sup>10</sup> suggesting that, regardless of who takes the tests, there is no substantively meaningful relationship between the IAT and the AMP.

It is all the more important, then, to call attention to potentially mitigating factors. One possibility is that implicit measures of prejudice perform satisfactorily in a controlled laboratory setting but not in an online interview, conceivably because their instructions are too complicated for older or lower-educated respondents.<sup>11</sup> Another potential concern is that the ANES panel study comprises twenty waves. The frequency of being interviewed and re-interviewed so many times possibly was off-putting to respondents. However, the results in Figure 3 and Table 1 demonstrate that the AMP as administered in the 2008 ANES cross-sectional study behaves similarly to that in the panel.<sup>12</sup> These results suggest that measures of implicit prejudice in a multi-wave panel are comparable to scores in a one-off interview.

---

<sup>8</sup>In addition, respondents in this study had to have made a decision whether to vote for Obama or McCain (“vote intention”). Further, the sample was weighted not to be representative of the American electorate but, instead, to have equal proportions of Obama and McCain “intended voters” (Greenwald et al. 2009, p. 243).

<sup>9</sup>We should like to credit Greenwald et al. with the discovery of the absence of a relationship of substantive significance between the IAT and the AMP, but they treat the result as raising no concerns about the validity of either or both measures. In systematic studies of general population beliefs and attitudes, a correlation of 0.22 would standardly be characterized as weak, or minimal, or modest.

<sup>10</sup>Since the size of a correlation coefficient is conditional on the variances in measures, it is an inappropriate statistic for comparison across difference samples (Achen 1982, p. 80).

<sup>11</sup>A report of the pilot test of the BIAT in the ANES panel study noted complaints about the tediousness of the testing procedure and high rates of break-off (Krosnick and Lupia 2008). In the actual study, no comparable problems were noted (DeBell, Krosnick, and Lupia 2010).

<sup>12</sup>For example, the correlation between AMP score and voting for Obama in the panel is  $-0.15$  and in the cross section is  $-0.16$ .

The fact that IAT and the AMP were administered in waves widely separated in time is also a potential concern. Is it likely that they would have been more highly correlated if they had been administered at the same time? Gawronski et al. (2017) report that test-retest reliability coefficients for the IAT and the AMP for assessment of racial attitudes are low – on the order of 0.44 and 0.38, respectively. In their most recent review of research on implicit social cognition, Greenwald and Lai (2020) report test-retest reliability coefficients for the BIAT of 0.43. Rather than the timing of the waves, the magnitude of measurement error for both the IAT and the AMP, particularly in the context of racial prejudice, is the most likely explanation for the lack of a relationship between the IAT and the AMP.

Indeed, the issue is not the validity of implicit measures of attitudes. Reliabilities of both the IAT and the AMP for the measurement of other objects, (e.g., self-concept, evaluative ratings of political figures) are satisfactory (Gawronski et al. 2017). The concern is the unreliability of test scores of the IAT and the AMP as measures of socially sensitive attitudes specifically.

## 6 Conclusion

This study’s focus is narrow – the validity of the two principal implicit measures of prejudice, the IAT and the AMP. Both claim to measure racial prejudice. Yet, the correlation between them is substantively indistinguishable from zero. One, or the other, or both, cannot be measuring what they purport to measure. It is fair to say that the scientific and legal implications of this result are considerable.

## References

- Achen, Christopher H. 1982. *Interpreting and Using Regression*. Thousand Oaks, CA: Sage Publications.
- Arkes, Hal R., and Philip E. Tetlock. 2004. “Attributions of Implicit Prejudice, or ‘Would Jesse Jackson ‘Fail’ the Implicit Association Test?’.” *Psychological Inquiry* 15(4): 257–278.
- Banaji, Mahzrin R., Brian Nosek, and Anthony G. Greenwald. 2004. “No Place for Nostalgia in Science: A Response to Arkes and Tetlock.” *Psychological Inquiry* 15: 279–289.
- Blaison, Christophe, Roland Imhoff, Isabell Hühnel, Ursula Hess, and Rainer Banse. 2012. “The Affect Misattribution Procedure: Hot or Not?” *Emotion* 12(2): 403–412.
- Blanton, Hart, James Jaccard, Jonathan Klick, Barbara Mellers, Gregory Mitchell, and Philip E. Tetlock. 2009. “Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT.” *Journal of Applied Psychology* 94(3): 567–582.
- Corneille, Olivier, and Mandy Hütter. 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.
- DeBell, Matthew, Jon A. Krosnick, and Arthur Lupia. 2010. “Methodology Report and User’s Guide for the 2008–2009 ANES Panel Study.” Available at: [https://electionstudies.org/wp-content/uploads/2009/03/anes\\_specialstudy\\_2008\\_2009panel\\_MethodologyRpt.pdf](https://electionstudies.org/wp-content/uploads/2009/03/anes_specialstudy_2008_2009panel_MethodologyRpt.pdf).
- Devine, Patricia G. 1989. “Stereotypes and Prejudice: Their Automatic and Controlled Components.” *Journal of Personality and Social Psychology* 56(1): 5–18.

- Ditonto, Tessa M., Richard R. Lau, and David O. Sears. 2013. "AMPing Racial Attitudes: Comparing the Power of Explicit and Implicit Racism Measures in 2008." *Political Psychology* 34(4): 487–510.
- Fazio, Russell H., and Michael A. Olson. 2003. "Implicit Measures in Social Cognition Research: Their Meaning and Use." *Annual review of psychology* 54(1): 297–327.
- Fazio, Russell H., David M. Sanbonmatsu, Martha C. Powell, and Frank R. Kardes. 1986. "On the Automatic Activation of Attitudes." *Journal of Personality and Social Psychology* 50(2): 229–238.
- Fazio, Russell H., Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams. 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.
- Finn, Christopher, and Jack Glaser. 2010. "Voter Affect and the 2008 U.S. Presidential Election: Hope and Race Mattered." *Analyses of Social Issues and Public Policy* 10(1): 262–275.
- Friese, Malte, Colin Tucker Smith, Marton Koeber, and Matthias Bluemke. 2016. "Implicit Measures of Attitudes and Political Voting Behavior." *Social and Personality Psychology Compass* 10(4): 188–201.
- Gawronski, Bertram, Mike Morrison, Curtis E. Phillips, and Silvia Galdi. 2017. "Temporal Stability of Implicit and Explicit Measures: A Longitudinal Analysis." *Personality and Social Psychology Bulletin* 43(3): 300–312.
- Greenwald, Anthony G., and Calvin K. Lai. 2020. "Implicit Social Cognition." *Annual Review of Psychology* 71: 419–495.
- Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological review* 102(1): 4.

- Greenwald, Anthony G., and Mahzarin R. Banaji. 2017. "The Implicit Revolution: Reconceiving the Relation Between Conscious and Unconscious." *American Psychologist* 72(9): 861.
- Greenwald, Anthony G., Colin Tucker Smith, N. Sriram, Yoav Bar-Anan, and Brian A. Nosek. 2009. "Implicit Race Attitudes Predicted Vote in the 2008 U.S. Presidential Election." *Analyses of Social Issues and Public Policy* 9(1): 241–253.
- Greenwald, Anthony G., Debbie E. McGhee, and Jordan L.K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74(6): 1464–1480.
- Greenwald, Anthony G., Mahzarin R. Banaji, and Brian A. Nosek. 2015. "Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects." *Journal of Personality and Social Psychology* 108(4): 553–561.
- Kalmoe, Nathan P., and Spencer Piston. 2013. "Is Implicit Prejudice against Blacks Politically Consequential? Evidence from the AMP." *Public Opinion Quarterly* 77(1): 305–322.
- Kam, Cindy D. 2007. "Implicit Attitudes, Explicit Choices: When Subliminal Priming Predicts Candidate Preference." *Political Behavior* 29(3): 343–367.
- Kinder, Donald R., and Timothy J. Ryan. 2017. "Prejudice and Politics Re-Examined: The Political Significance of Implicit Racial Bias." *Political Science Research and Methods* 5(2): 241–259.
- Krosnick, Jon A., and Arthur Lupia. 2008. "Decisions Made About Implicit Attitude Measurement in the 2008 American National Election Studies." Available at: [https://electionstudies.org/wp-content/uploads/2019/04/20090625\\_IAT.pdf](https://electionstudies.org/wp-content/uploads/2019/04/20090625_IAT.pdf).



- Lundberg, Kristjen B., and B. Keith Payne. 2014. "Decisions Among the Undecided: Implicit Attitudes Predict Future Voting Behavior of Undecided Voters." *PLoS ONE* 9(1): e105655.
- Messing, Solomon, Maria Jabon, and Ethan Plaut. 2016. "Bias in the Flesh." *Public Opinion Quarterly* 80(1): 44–65.
- Oswald, Frederick L., Gregory Mitchell, Hart Blanton, James Jaccard, and Philip E. Tetlock. 2013. "Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies." *Journal of Personality and Social Psychology* 105(2): 171–192.
- Pasek, Josh, Alexander Tahk, Yphtach Lelkes, Jon A. Krosnick, B. Keith Payne, Omair Akhtar, and Trevor Thompson. 2009. "Determinants of Turnout and Candidate Choice in the 2008 U.S. Presidential Election: Illuminating the Impact of Racial Prejudice and Other Considerations." *Public Opinion Quarterly* 73(5): 943–994.
- Payne, B. Keith, Clara Michelle Cheng, Olesya Govorun, and Brandon D. Stewart. 2005. "An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement." *Journal of Personality and Social Psychology* 89(3): 277–293.
- Payne, B. Keith, Jon A. Krosnick, Josh Pasek, Yphtach Lelkes, Omair Akhtar, and Trevor Thompson. 2010. "Implicit and Explicit Prejudice in the 2008 American Presidential Election." *Journal of Experimental Social Psychology* 46(2): 367–374.
- Payne, Keith, and Kristjen Lundberg. 2014. "The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms." *Social and Personality Psychology Compass* 8(12): 672–686.
- Pérez, Efrén O. 2010. "Explicit Evidence on the Import of Implicit Attitudes: The IAT and Immigration Policy Judgments." *Political Behavior* 32: 517–545.

- Pérez, Efrén O. 2016. *Unspoken Politics: Implicit Attitudes and Political Thinking*. New York: Cambridge University Press.
- Ryan, Timothy J. 2017. “How Do Indifferent Voters Decide? The Political Importance of Implicit Attitudes.” *American Journal of Political Science* 61(4): 892–907.
- Segura, Gary M., and Ali A. Valenzuela. 2010. “Hope, Tropes, and Dopes: Hispanic and White Racial Animus in the 2008 Election.” *Presidential Studies Quarterly* 40(3): 497–514.
- Sriram, N., and Anthony G. Greenwald. 2009. “The Brief Implicit Association Test.” *Experimental Psychology* 56(4): 283–294.

# Online Appendix

A	Explicit prejudice question wording	A2
B	Additional figure	A3

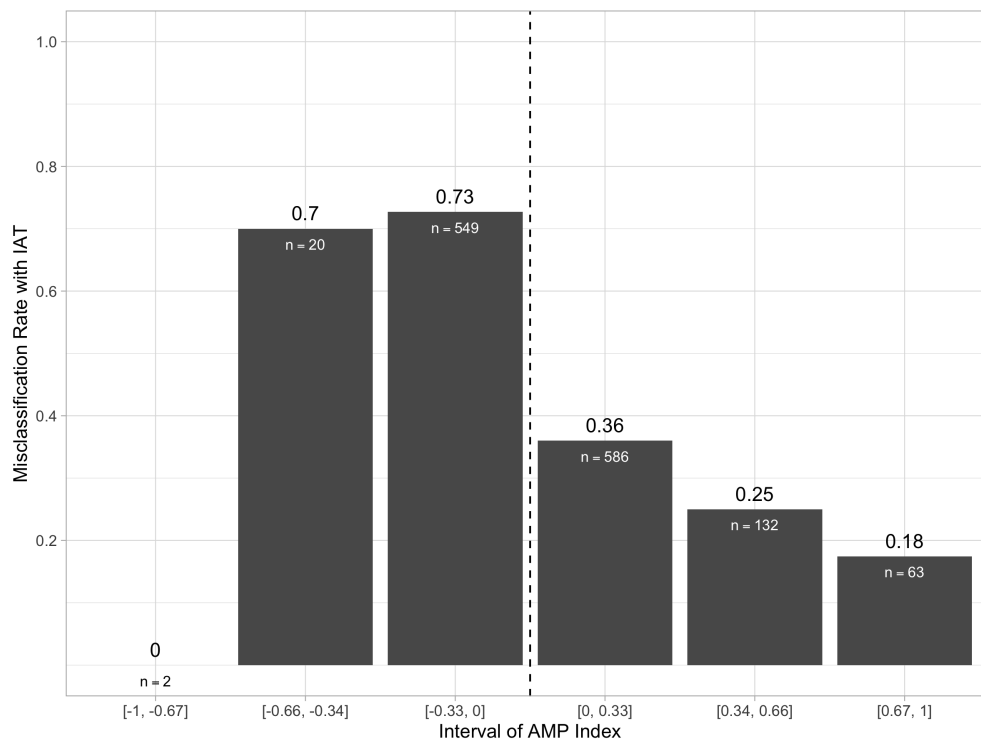
## A Explicit prejudice question wording

The question wording for the explicit measures of racial prejudice that were used in the ANES panel study was slightly different in the 2008 ANES cross-sectional study. Moreover, the measures in the cross section were all administered in the same wave. In the panel study, respondents were asked: “How well does the word ‘lazy’ describe most [whites/Blacks]?” and “How well does the word ‘intelligent at school’ describe most [whites/Blacks]?” on five-point scales (in a random order). In the cross section, the questions read: “Now I have some questions about different groups in our society. I’m going to show you a seven-point scale on which the characteristics of the people in a group can be rated. In the first statement a score of 1 means that you think almost all of the people in that group tend to be ‘hard-working.’ A score of 7 means that you think most people in the group are ‘lazy.’ A score of 4 means that you think that most people in the group are not closer to one end or the other, and of course, you may choose any number in between... The next set asks if people in each group tend to be ‘intelligent’ or ‘unintelligent’ ...Where you rate [WHITES/BLACKS] in general on this scale?” (randomly ordered among a series of other groups).

For the feeling thermometers in the panel, questions asked, “Do you feel warm, cold, or neither warm nor cold toward [whites/Blacks]?” (in a random order) and then asked whether respondents felt “extremely,” “moderately,” or “a little” warm or cold, creating a 7-point composite measure for each group. By contrast, the ANES cross section feeling thermometer asked: “I’d like to get your feelings toward some of our political leaders and other people who are in the news these days. I’ll read the name of a person and I’d like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don’t feel favorable toward the person and that you don’t care too much for that person. You would rate the person at the 50 degree mark if you don’t feel particularly warm or cold toward the person. Still using the thermometer, how would you rate the following groups: [WHITES/BLACKS]” (randomly ordered among a series of other groups).

## B Additional figure

Figure B.1: Misclassification rates on the IAT based on intervals of the AMP.



$N = 1352$  white respondents in the 2008–2009 ANES panel study. IAT scores are measured on a  $-2$  to  $2$  scale, and AMP scores on a  $-1$  to  $1$  scale (higher values indicating higher prejudice toward Black people). Respondents are grouped across intervals of their AMP score. The gray bars represent the proportion of respondents in each group who are classified as prejudiced against Black people according to the AMP but not the IAT, with sample sizes for each group included in the bars.