Supplemental Materials:

Facing One's Implicit Biases: From Awareness to Acknowledgment

Adam Hahn[1]

Bertram Gawronski[2]

Word count: 20,754 excluding acknowledgements, abstract, references, figures, and tables

[1] Social Cognition Center Cologne, University of Cologne, Germany

[2] University of Texas at Austin, Austin, TX, USA

Correspondence regarding this article should be sent to Adam Hahn, Social Cognition Center Cologne, Department of Psychology, University of Cologne, Richard-Strauss-Str. 2, 50931 Köln, Germany, Adam.Hahn@uni-koeln.de

Supplemental Materials Section A

To investigate potential downstream consequences of IAT score prediction and IAT completion, participants in Studies 1-3 completed a series of exploratory measures at the end of each study (see Figure A1). To streamline the presentation of hypotheses-relevant measures, the results of the exploratory measures are presented here instead of the main article.

Study 1

In Study 1, all measures of downstream consequences were completed at the end of the study before participants provided demographic information (see Figure A1). First, expanding on findings that egalitarian participants react with negative affect to information that their thoughts or behaviors may be out of line with their egalitarian standards (Czopp, Monteith, & Mark, 2006; Monteith, Devine, & Zuwernik, 1993),[1] participants completed measures of their current emotional state. Second, participants were asked to indicate whether they believed they should trust their "hunches" when interacting with different groups of people. Lastly, participants completed the Internal and External Motivation to Respond without Prejudice Scale (MRWP, Plant & Devine, 1998), a scale assessing nonprejudicial goals (Gawronski, Peters, Brochu & Strack, 2008), the rational-experiential inventory (REI; Epstein, Pacini, Denes-Raj, & Heier, 1996), and four items assessing their beliefs that the IAT is a measure of "true attitudes" or "culturally-learned associations." The order in which participants completed these measures is depicted in Figure A1.

[1] In discussing participants' emotional reactions as a function of participants' egalitarian standards, Monteith and colleagues (1993) use the terms *high-prejudice* and *low-prejudice* to differentiate between participants who score high or low on different explicit scales measuring attitudes towards minorities. However, in line with other research, we use the term *prejudice* to describe evaluative intergroup biases. To avoid potential confusion between our dependent measures, we use the terms *high egalitarian* and *low egalitarian* to describe participants who would presumably react negatively to finding out that their behavior is biased.

Method

Intergroup intuitions. One item asked participants to rate their agreement with the statement *It is best to go with one's 'hunches' when interacting with different groups of people* on a 7-point item ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Nonprejudicial goals. On 5-point scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) participants rated their agreement with 10 items assessing non-prejudicial goals (Cronbach's α = .76, Gawronski et al., 2008).

Emotions. On 6-point scales ranging from 1 (*not at all*) to 6 (*extremely*), participants were asked to indicate how much they experienced various emotions in individually randomized orders. We combined all negative emotions (irritated, bothered, angry, uneasy, sad, depressed, ashamed, guilty, and disappointed with myself) into one measure of negative affect (Cronbach's α = .89).

Rational-experiential inventory. The REI (Epstein et al., 1996) consists of two subscales with five items each that separately assess participants' Need for Cognition (Cronbach's α = .74) and Faith in Intuition (α = .77). Participants rated whether these items described them on 5-point scales ranging from 1 (*completely false*) to 5 (*completely true*).

Motivation to respond without prejudice scale. Participants completed Plant and Devine's (1998) MRWP scale, using 9-point scales ranging from 1 (*strongly disagree*) to 9 (*strongly agree*) in individually randomized orders. We averaged all items into a single factor with higher scores indicating stronger motivation to respond without prejudice (Cronbach's α = .77).[2]

[2] The original conceptualization of the MRWP scale includes two subscales: internal motivation (IMS, α = .84), and external motivation (EMS, α = .80). Because the patterns of results were similar for the two subscales, we combined the two subscales in a single factor for the sake of simplicity.

IAT beliefs. To measure participants' subjective beliefs about the IAT, we adapted a scale from Hahn, Judd, Hirsh, and Blair (2014). Participants were asked to indicate whether they believed the IAT measured their "true attitudes" with two items (Cronbach's $\alpha = .66$) and their "culturally-learned associations" with two items (Cronbach's $\alpha = .61$).

Results

Unless otherwise indicated, all measures of downstream consequences were analyzed with 2 (IAT Score Prediction: prediction vs. no prediction) × 2 (IAT Completion Order: IATs completed before second thermometer ratings vs. IATs after second thermometer ratings) between-subjects ANOVAs. The means of all measures of downstream consequences in the four experimental conditions are presented in Table A1.

Intergroup intuitions. The measure of intergroup intuitions revealed a significant main effect of IAT Score Prediction, $F(1, 146) = 7.48$, $p = .007$, $\eta_p^2 = .049$. Participants who predicted their IAT scores agreed significantly less with the statement that one should trust one's hunches when interacting with different groups of people than participants who did not predict their IAT scores. IAT Completion Order had no effect, $F(1, 146) = 1.99$, $p = .16$, $\eta_p^2 = .013$, and neither did the interaction of IAT Completion Order and IAT Score Prediction, $F(1, 146) = .43$, $p = .54$, $\eta_p^2 = .514$.

Nonprejudicial goals. This scale showed no effects, all Fs < 0.60, $p > .40$, and was therefore not included in the subsequent studies.

Negative affect. We found a significant main effect of IAT Score Prediction on negative affect, $F(1, 146) = 6.71$, $p = .011$, $\eta_p^2 = .044$. Participants experienced greater negative affect when they predicted their IAT scores that when they did not. There was no significant effect of

IAT Completion Order, $F(1, 146) = 2.44$, $p = .120$, $\eta_p^2 = .016$, or the interaction between IAT Completion Order and IAT Score Prediction, $F(1, 146) = .00$, $p = .95$, $\eta_p^2 = .000$.

Rational-experiential inventory. Neither Need for Cognition nor Faith in Intuition showed any effects, all Fs < 1.1, all ps > .30, and were therefore not included in the subsequent studies.

Motivation to respond without prejudice. The MRWP scale showed a significant two-way interaction of IAT Score Prediction and IAT Completion Order, $F(1, 146) = 5.36$, $p = .022$, $\eta_p^2 = .035$. Simple effect analyses revealed that predicting IAT scores increased participants' motivation to respond without prejudice when they completed the IATs after they provided their second thermometer ratings, $F(1, 146) = 7.43$, $p = .007$, $\eta_p^2 = .048$. However, there was no significant effect of IAT Score Prediction when participants completed the IATs before they provided their second thermometer ratings, $F(1, 146) = .33$, $p = .57$, $\eta_p^2 = .002$.

IAT beliefs. To test whether our manipulations influenced participants' beliefs about the IAT, we conducted a 2 (Beliefs: IAT measures "culturally-learned associations" vs. "true attitudes") × 2 (IAT Score Prediction) × 2 (IAT Completion Order) mixed ANOVA with repeated measures on the first factor (see Table 2 for difference scores in relative agreement between the two beliefs). The ANOVA revealed a significant two-way interaction of Beliefs and IAT Score Prediction, $F(1, 146) = 9.85$, $p = .002$, $\eta_p^2 = .063$. Simple effect analyses revealed that participants who did not predict their IAT scores agreed significantly more with the idea that the IAT measures culturally-learned associations rather than true attitudes, $F(1, 146) = 17.93$, $p < .001$, $\eta_p^2 = .11$. In contrast, participants who predicted their IAT scores did not show any difference between the two types of belief, $F(1, 146) = .04$, $p = .84$, $\eta_p^2 = .000$.

Study 2

In Study 2, we implemented three changes in the measurement of downstream consequences. First, we modified the procedure in Study 2, so that we could test the downstream consequences of IAT completion more directly with a less ambiguous control condition. Rather than completing the IATs right after the second thermometer ratings (Step 5 in Figure A1), participants in the no-completion group now completed the IATs at the end of the study, after responding to the questions about their emotions, intergroup intuitions, and the MRWP scale (after Step 8 in Figure 1). Second, we included only those measures of downstream consequences that had shown significant effects in Study 1, and hence dropped the non-prejudicial goals scale and the REI scale. Third, participants in the no-completion condition completed the measures of downstream consequences before they completed the IATs (with the exception of the IAT beliefs scale). Using the graphical depiction of the procedure in Figure 1, Step 5 was moved below Step 8 (the MRWP scale), and Steps 6a and 7a were deleted.

Method

As depicted in Figure A1, after completing the second thermometer ratings, all participants completed the measure of intergroup intuitions, supplemented with two additional items: (1) *In interactions with people of different backgrounds, one should always go with one's gut*; (2) *When I interact with people who have backgrounds that are different than my own, I trust my initial gut reaction* (Cronbach's $\alpha = .80$). Then, all participants completed the same negative emotions items from Study 1 (Cronbach's $\alpha = .86$), followed by the MRWP Scale (Cronbach's $\alpha = .67$, Plant & Devine, 1998). At this point, participants in the no-completion condition were asked to complete the five IATs. Finally, all participants responded to the items on their beliefs about the IAT ("true attitudes", Cronbach's $\alpha = .50$; "culturally-learned associations", Cronbach's $\alpha = .25$,

Hahn et al., 2014), including two additional exploratory items not analyzed here. We also administered a series of new exploratory items. Because psychometric and exploratory factor analyses showed poor psychometric properties of these items, they are not discussed further.

Results

The effects of IAT Score Prediction and IAT Completion on the measures of downstream consequences were again analyzed with 2 (IAT Score Predictions) × 2 (IAT Completion) ANOVAs. Note that, in the current study, effects of IAT Completion on downstream consequences have a different meaning than in Study 1, in that they capture the effects of actually completing IATs (the only exception being the measure of IAT beliefs). The mean scores of all measures of downstream consequences are presented in Table A2.

Intergroup intuitions. For our scale measuring the belief that one should trust one's intuitions in intergroup encounters, the ANOVA yielded no significant effects, all Fs < 1, all ps > .35. Using only the item that we used in Study 1 showed no significant effects of IAT Score Prediction either, all Fs < 2, all ps > .15.

Negative affect. In contrast to Study 1, IAT Score Prediction did not show a significant effect on negative affect, $F(1, 189) = 2.15$, $p = .144$, $\eta_p^2 = .011$. Instead, negative affect showed a significant main effect of IAT Completion, $F(1, 189) = 5.77$, $p = .017$, $\eta_p^2 = .030$, in that participants expressed more negative affect when completed IATs than when they did not complete IATs.

Motivation to respond without prejudice. Analyses on the MRWP scale replicated the two-way interaction of IAT Completion and IAT Score Prediction in Study 1, $F(1, 189) = 4.86$, $p = .029$, $\eta_p^2 = .025$. When participants did not complete IATs, predicting their IAT scores led to a marginal increase in the motivation to respond without prejudice, $F(1, 189) = 3.26$, $p = .073$, η_p^2

= .017. However, when participants completed IATs, this effect was eliminated, $F(1, 189) = 1.72$, $p = .19$, $\eta_p^2 = ..009..$

IAT beliefs. A 2 (Beliefs: IAT measures "culturally-learned associations" vs. "true attitudes") × 2 (IAT Score Prediction) × 2 (IAT Completion) mixed ANOVA did not replicate the two-way interaction of Beliefs and IAT Score Prediction obtained in Study 1. Instead, we found a significant three-way interaction of Beliefs, IAT Score Prediction, and IAT Completion, $F(1, 189) = 3.18$, $p = .032$, $\eta_p^2 = .024$. Difference scores for the two belief scales are shown in Table 4. Simple effect analyses revealed that participants in the no-completion, no-prediction condition agreed more with the idea that the IAT measures "culturally-learned associations" rather than "true attitudes", $F(1, 189) = 5.69$, $p = .018$, $\eta_p^2 = .029$. Participants in all other conditions agreed with both ideas to the same degree, all Fs < 2.2, all ps < .10.

Study 3

Method

Study included an additional IAT completion manipulation, IAT completion with feedback, for a 2 (IAT score prediction vs. no prediction) by 3 (no IAT completion vs. IAT completion without feedback vs. IAT completion with feedback) design (see main paper). Measures of downstream consequences were similar to Study 2 and were administered in the same way at the same time points. The measures again included the trust in intergroup intuitions scale (Cronbach's α = .79), negative affect (Cronbach's α = .83), and the MRWP scales (Cronbach's α = .58, Plant & Devine, 1998). As in Study 2, participants in the no-IAT completion condition completed the IATs toward the end of the study before completing the measure of IAT beliefs ("true attitudes", Cronbach's α = .58; "culturally-learned associations",

Cronbach's α = .28, Hahn et al., 2014), again including two additional exploratory items not analyzed here.

Results

We analyzed all effects on downstream consequences with 2 (IAT score prediction) × 3 (IAT completion) ANOVAs unless otherwise indicated, and followed up with more specific pairwise contrasts for the different IAT completion conditions when the results with the 2-degrees-of-freedom factor afforded further analyses. The means of all variables are presented in Table A3.

Intergroup Intuitions. There were no significant effects in the overall ANOVA on our three-item scale measuring trust in intuitions for intergroup encounters, all *F*s < 2.3, all *p*s >.10, and no effects when only the item administered in Study 1 was used, all *F*s < 1.8, all *p*s >.15.

Negative affect. The manipulations had no effects on negative affect in this study, all *F*s < 1.05, all *p*s > .35.

Motivation to respond without prejudice. In contrast to Studies 1 and 3, where the MRWP scale showed an interaction effect of IAT Score Prediction and IAT Completion, the MRWP scale showed a significant main effect of IAT Score Prediction, $F(1, 237) = 4.74$, $p = .030$, $\eta_p^2 = .020$, indicating that participants who predicted their IAT scores were more motivated to respond without prejudice than participants who did not predict their IAT scores. IAT Completion Order had no effect, $F(2, 237) = 1.96$, $p = .143$, $\eta_p^2 = .016$, and neither did the interaction of IAT Completion Order and IAT Score Prediction, $F(2, 237) = .09$, $p = .913$, $\eta_p^2 = .001$

IAT beliefs. A 2 (Beliefs: IAT measures "culturally-learned associations" vs. "true attitudes") × 2 (IAT Score Prediction) × 3 (IAT Completion) ANOVA with repeated measures

on the first factor yielded a three-way interaction between all factors, $F(2, 237) = 4.57$, $p = .011$, $\eta_p^2 = .024$. Participants in both the no-completion condition and the IATs-with-feedback condition agreed with the idea that the IAT measures "cultural associations" more than "true attitudes" when they had not predicted their IAT scores, simple effect associations over attitudes in no IAT condition: $F(1, 237) = 10.60$, $p = .001$, $\eta_p^2 = .043$, simple effect associations over attitudes in IAT-with-feedback condition: $F(1, 237) = 20.66$, $p < .001$, $\eta_p^2 = .080$. However, this difference disappeared when they made IAT score predictions, both Fs < 1.6, both ps > .2. The pattern was reversed in the IAT completion without feedback condition. Here, participants who predicted their IAT scores preferred the associations over the attitudes explanation, $F(1, 237) = 6.68$, $p = .010$, $\eta_p^2 = .027$, whereas participants who did not predict their scores showed no preference, $F(1, 237) = 0.11$, $p = .75$, $\eta_p^2 = .000$.

Combined Analysis

Because small sample sizes can increase the likelihood of false negatives (Cumming, 2014) as well as false positives (Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafo, 2013), we decided to combine the samples from Studies 1-3 to obtain a larger sample for the identification of reliable effects. We combined the data from all participants except for participants in the IAT completion plus feedback conditions of Study 3 (because that condition was only run in Study 3) for a sample of $N = 503$ participants. All analyses also included Study as a factor with three levels.

Results

The effects of IAT Score Prediction and IAT Completion on the measures of downstream consequences were analyzed with 2 (IAT Score Predictions) × 2 (IAT Completion) × 3 (Study)

ANOVAs, we followed up the last factor with more specific contrasts in case of significant results. The mean scores of all measures of downstream consequences are presented in Table A4.

Intergroup intuitions. The one item of our "trust in intuitions" measure that was used in all three samples showed a significant main effect of IAT Score Prediction, $F(1, 491) = 4.14$, $p = .042$, $\eta_p^2 = .008$, showing that participants trusted their intuitions less when they predicted their IAT scores than when they did not. This main effect was qualified by a marginal interaction with Study, $F(2, 491) = 2.85$, $p = .056$, $\eta_p^2 = .011$. Follow-up analyses showed an interaction of Prediction with a contrast code contrasting Study 1 from Studies 2 and 3, $F(1, 491) = 3.93$, $p = .048$, $\eta_p^2 = .008$, indicating that this effect was significant only for the sample in Study 1, $F(1, 146) = 7.48$, $p = .007$, $\eta_p^2 = .049$, but not in the two samples in Studies 2 and 3, both Fs < 1.4, both ps > .2.[3]

Negative Affect. The negative affect measure showed an overall main effect of IAT completion, $F(1, 491) = 4.03$, $p = .045$, $\eta_p^2 = .008$. The interaction of this effect with Study was not significant, $F(2, 491) = 2.21$, $p = .111$, $\eta_p^2 = .009$. The same analysis also showed an interaction between IAT score prediction and Study, $F(2, 491) = 4.92$, $p = .008$, $\eta_p^2 = .020$. Follow-up analyses showed an interaction of this factor with a contrast comparing Study 1 with Studies 2 and 3. Only participants in Study 1, but not participants in Studies 2 and 3, reacted to predicting IAT scores with negative affect, $F(1, 491) = 7.36$, $p = .007$, $\eta_p^2 = .015$. In trying to make sense of this finding, we noticed one procedural difference between the studies. Participants in the Study 1 completed Gawronski et al.'s (2008) non-prejudicial goals scale right

[3] There was also an interaction of IAT completion with Study, $F(2, 491) = 3.36$, $p = .036$, $\eta_p^2 = .013$, indicating that differences on intergroup intuitions between the IAT Completion and the No-completion conditions went in different directions in the different studies. However, none of the simple effects of IAT Completion in any individual study approached significant, such that this interaction is not interpreted further.

before completing the affect measure (see Figure A1). This scale includes items such as "I feel guilty when I have negative thoughts or feelings about the members of disadvantaged minority groups" and "I get angry with myself when I have a thought or feeling that might be considered prejudiced." Rating these items may have caused participants to report negative affect in response to IAT score predictions to conform to the item content of the previous scale. If this explanation is true, we reasoned, then only participants who did in fact agree with those items should show negative affect in response to IAT score predictions, whereas participants who did not agree with these items should not show negative affect in response to IAT score predictions.

To test this explanation we regressed negative affect onto a z-standardized score of a three-item subscale of the non-prejudicial goal scale (henceforth called "internal non-prejudicial goals scale", INPG)[4], the experimental conditions, and all interactions. Results are depicted in Figure A2. Consistent with our reasoning, we found a significant interaction between INPG and IAT score prediction, $b = 22$, $SE = .07$, $t\,(142) = 2.92$, $p = .004$. As can be seen in Figure 5, only participants who agreed with the items suggesting one should feel bad after noticing prejudiced feelings also reported feeling worse after IAT score prediction as opposed to no prediction, $F(1, 142) = 16.05$, $p < .001$, $\eta_p^2 = .102$. Participants who did not agree with those items showed no effect, $F(1, 142) = 0.25$, $p = .86$. Although we did not predict this particular pattern, it is compatible with Monteith and colleague's (Czopp et al., 2006; Monteith et al., 1993) findings that negative affect in response to confrontations with prejudiced thoughts or behaviors is a

[4] An exploratory factor analysis with principal component analysis and Oblimin rotation suggested this factor comprised of the two items above and the item "Negative evaluations of disadvantaged minority members are wrong." (all three loadings > .68, loadings of all other items < .4). Repeating the reported analyses with only the two items reported above shows essentially equivalent results. Repeating analyses with the whole scale also replicates the crucial IAT score prediction by goals interaction, but to a weaker degree, $b = 15$, $SE = .08$, $t\,(142) = 2.00$, $p = .048$.

typical reaction only for high egalitarians. However, our result cannot differentiate whether there are also cultural differences, and whether egalitarian values need to be made salient (by having participants fill out a scale) to obtain this effect.

Motivation to respond without prejudice. The MRWP scale showed a significant main effect of IAT Score Prediction, $F(1, 491) = 5.08$, $p = .025$, $\eta_p^2 = .010$, which was qualified by an interaction with IAT Completion, $F(1, 491) = 7.84$, $p = .005$, $\eta_p^2 = .016$. There were no interactions of the manipulations with Study, all Fs < 1.5, all ps > .24. Hence, the overall analysis continued to show that predicting IAT scores led to increased motivation to respond without prejudice when participants did not complete IATs, $F(1, 491) = 12.98$, $p < .001$, $\eta_p^2 = .026$, but not when they completed IATs, $F(1, 491) = 1.47$, $p = .702$, $\eta_p^2 = .000$.

IAT beliefs. A 2 (Beliefs: IAT measures "culturally-learned associations" vs. "true attitudes") × 2 (IAT Score Prediction) × 2 (IAT Completion) × 3 (Study) ANOVA with repeated measures on the first factor showed a significant main effect of Beliefs, showing that, in general, participants preferred cultural over personal explanations of IAT results, $F(1, 491) = 22.30$, $p < .001$, $\eta_p^2 = .043$. This main effect was qualified by a marginal interaction of Beliefs with IAT Score Prediction, indicating that this effect tended to be weaker when participants predicted their IAT scores, $F(1, 491) = 3.16$, $p = .076$, $\eta_p^2 = .006$. However, this marginal two-way interaction was qualified by two independent three-way interactions. First, it interacted with Study, $F(2, 491) = 3.30$, $p = .038$, $\eta_p^2 = .013$. Second, it interacted with the IAT completion manipulation, $F(1, 491) = 10.17$, $p = .002$, $\eta_p^2 = .020$. Decomposing these interactions with dummy codes showed the following patterns. All participants showed a reduction in preference for cultural over personal explanations when IATs were completed towards the end of the study, $F(1, 491) = 12.55$, $p < .001$, $\eta_p^2 = .025$, and this simple effect did not interact with a contrast comparing

Study 1 to Studies 2 and 3, $F(1, 491) = 0.89$, $p = .35$. However, there was a simple three-way interaction between Beliefs, IAT score prediction and a Study 1 vs. Studies 2 and 3 contrast within the condition where IATs were completed earlier, $F(1, 491) = 6.78$, $p = .009$, $\eta_p^2 = .014$. Participants in Study 1 showed the same reduction in preference for cultural over personal explanations in response to predicting IAT scores regardless of when they completed the IATs. However, the preferences for cultural over personal explanations in Studies 2 and 3 showed the three-way interaction with Prediction and IAT Completion order explained above in the respective sections.

In sum, all participants showed a reduction in preference for cultural over personal explanations for IAT biases when IATs were completed towards the end of the study. But results were ambiguous with respect to whether the order of IAT completion modified this effect (Studies 2 and 3), or not (Study 1). One potential non-theoretical explanation for this asymmetry is that the "cultural associations" scale showed lower reliabilities in Studies 2 and 3 ($\alpha = .25$ and .21 in Studies 2 and 3, respectively) as opposed to the Study 1 ($\alpha = .61$).

General Discussion

Analyses across the three samples with over 500 participants confirmed and refined some of our findings from the individual studies. First, analyses confirmed that predicting IAT scores increases motivation to respond without prejudice, but that this effect diminished when participants were asked to re-evaluate their biases after completing IATs. They further showed that only participants in Study 1, but not participants in Studies 2 and 3, reacted with less trust in intuitions for intergroup encounters to predicting IAT scores. It is unclear at this point whether this result reflects a contextual difference between our racially diverse Canadian sample in Study

1 and our racially more homogenous German samples in Studies 2 and 3, or simply a false positive in the Canadian sample.

Another difference between the samples, more negative affect in response to predicting IAT scores in Study 1 as opposed to Studies 2 and 3, may have reflect a response to a procedural difference that unintentionally made egalitarian norms more salient in Study 1. Although not predicted, these results indicate that reactions to acknowledgement of bias depend on egalitarian norms rather than leading to similar reactions across situations and people. Lastly, all participants tended to show a reduction in preference for cultural over personal explanations of bias in response to predicting IAT scores, although the role of completing IATs for this effect remained ambiguous. Because these analyses were exploratory we refrain from post-hoc explanations at this point. In sum, however, these results indicate that the downstream consequences of acknowledgement of bias may depend heavily on contextual factors and other moderators.

References

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14,* 1-12.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7-29.

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, *90*, 784-803.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, *71*, 390-405.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin, 34*, 648-665.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*, 1369.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology*, *64*, 198-210.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811-832.

Table A1

Means (and standard errors) of trust in intergroup intuitions, emotions, motivation to respond without prejudice, and IAT beliefs as a function of IAT Score Prediction and IAT Completion, Study 1.

	IATs Completed Before Second Thermometer Ratings		IATs Completed After Second Thermometer Ratings	
Dependent Variables	IAT Score Predictions	No IAT Score Predictions	IAT Score Predictions	No IAT Score Predictions
Trust in intergroup intuitions	3.16 (.21)	3.89 (.22)	3.61 (.21)	4.05 (.21)
Non-prejudicial goals	3.68 (.10)	3.81 (.10)	3.77 (.10)	3.80 (.10)
Negative Affect	2.32 (.15)	1.93 (.25)	2.08 (.15)	1.71 (.15)
REI: Need for Cognition	3.57 (.13)	3.61 (.13)	3.38 (.12)	3.60 (.12)
REI: Faith in Intuition	3.21 (.11)	3.41 (.11)	3.33 (.11)	3.34 (.11)
Motivation to Respond Without Prejudice	5.84 (.18)	5.99 (.19)	6.38 (.18)	5.69 (.18)
Belief that the IAT measures "true attitudes" rather than "culturally-learned associations" (difference score)	.03 (.18)	.44 (.18)	-.08 (.18)	.64 (.18)

Note. Standard errors of predicted condition means are calculated from a 2 (IAT Completion) x 2 (IAT Score Prediction) between-subjects ANOVA. REI = Rational Experiential Inventory (Epstein et al., 1996).

Table A2

Means (and standard errors) of trust in intergroup intuitions, emotions, motivation to respond without prejudice, and IAT beliefs as a function of IAT Score Prediction and IAT Completion, Study 2.

Dependent Variables	IATs Completed Before Second Thermometer Ratings		IATs Completed at the End of the Study	
	IAT Score Predictions	No IAT Score Predictions	IAT Score Predictions	No IAT Score Predictions
Trust in Intergroup Intuitions	4.79 (.19)	4.80 (.19)	5.17 (.20)	4.96 (.19)
Negative Affect	1.97 (.11)	2.32 (.11)	1.88 (.11)	1.87 (.11)
Motivation to Respond Without Prejudice	5.94 (.22)	6.21 (.22)	6.32 (.23)	5.95 (.22)
Belief that the IAT measures "true attitudes" rather than "culturally-learned associations" (difference score)	.25 (.17)	-.16 (.17)	.09 (.17)	.40 (.17)

Note. Standard errors of predicted condition means are calculated from a 2 (IAT Completion) x 2 (IAT Score Prediction) between-subjects ANOVA.

Table A3

Means (and standard errors) of trust in intergroup intuitions, emotions, motivation to respond without prejudice, and IAT beliefs as a function of IAT Score Prediction and IAT Completion, Study 3.

Dependent Variables	IATs Completed with feedback		IATs Completed without feedback		IATs Completed at the End of the Study	
	IAT Score Predictions	No IAT Score Predictions	IAT Score Predictions	No IAT Score Predictions	IAT Score Predictions	No IAT Score Predictions
Trust in Intergroup Intuitions	5.08 (.17)	4.63 (.17)	4.81 (.17)	5.10 (.18)	4.73 (.17)	4.69 (.17)
Negative Affect	1.98 (.11)	2.00 (.11)	1.82 (.11)	1.89 (.12)	2.04 (.11)	1.80 (.11)
Motivation to Respond Without Prejudice	6.28 (.15)	6.06 (.15)	6.00 (.16)	5.74 (.16)	6.14 (.15)	5.79 (.15)
Belief that the IAT measures "true attitudes" rather than "culturally-learned associations" (difference score)	.24 (.19)	.89 (.20)	.51 (.20)	.07 (.20)	.13 (.19)	.65 (.20)

Note. Standard errors of predicted condition means are calculated from a 3 (IAT Completion) x 2 (IAT Score Prediction) between-subjects ANOVA.

Table A4

Means (and standard errors) of trust in intergroup intuitions, emotions, motivation to respond without prejudice, and IAT beliefs as a function of IAT Score Prediction and IAT Completion, Combined data of Studies 1-3.

	IATs Completed Before Second Thermometer Ratings		IATs Completed After Second Thermometer ratings	
Dependent Variables	IAT Score Predictions	No IAT Score Predictions	IAT Score Predictions	No IAT Score Predictions
Trust in Intergroup Intuitions	4.36 (.13)	4.70 (.13)	4.55 (.13)	4.64 (.13)
Negative Affect	2.03 (.07)	2.07 (.07)	1.99 (.07)	1.80 (.07)
Motivation to Respond Without Prejudice	5.93 (.09)	6.00 (.09)	6.28 (.09)	5.82 (.09)
Belief that the IAT measures "true attitudes" rather than "culturally-learned associations" (difference score)	.27 (.11)	.09 (.11)	.05 (.10)	.55 (.10)

Note. Standard errors of predicted condition means are calculated from a 2 (IAT Completion) x 2 (IAT Score Prediction) between-subjects ANOVA (interactions with predictors contrasting the different samples are not included in these models).

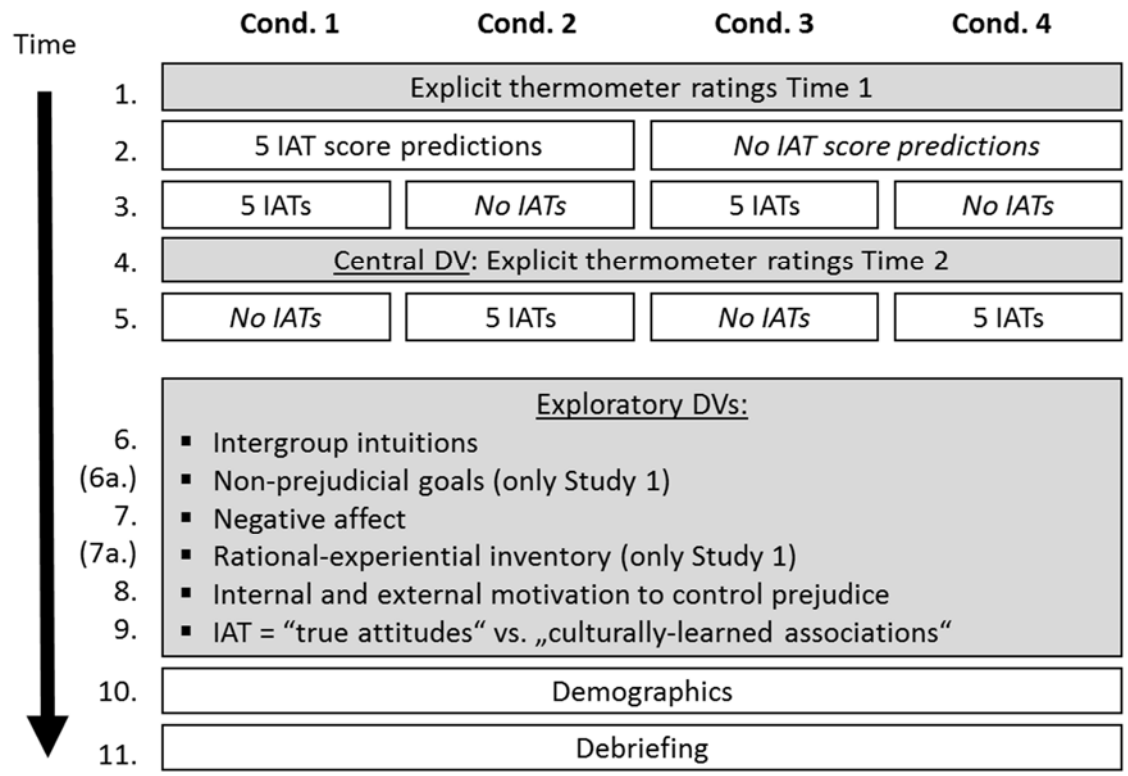

Figure A1. Procedure of Study 1: Four between-subjects conditions a 2 (IAT Score Prediction: prediction vs. no prediction) × 2 (IAT Completion before Time-2 thermometer ratings vs. after Time-2 thermometer ratings) between-subjects design. In Studies 2 and 3, participants in the no-completion conditions completed the IATs after the exploratory DVs before answering questions about their beliefs about the IAT.

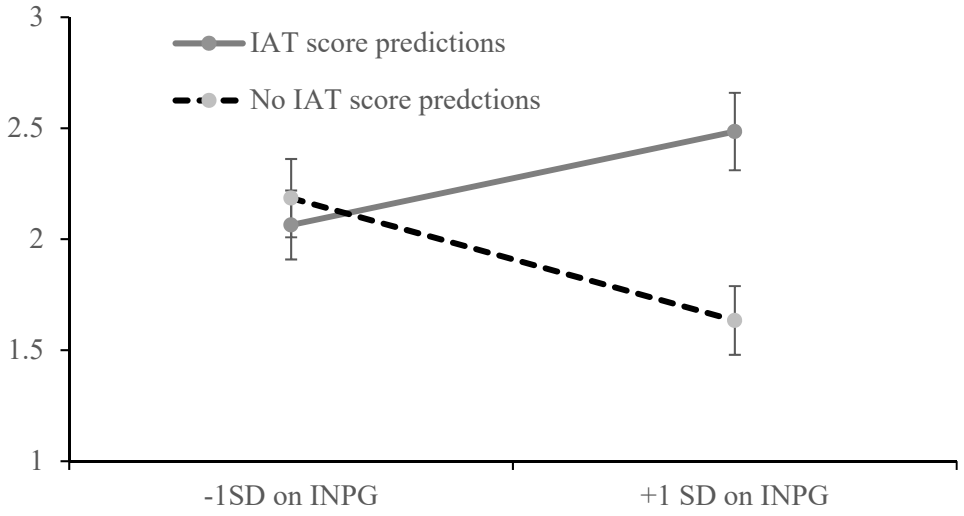

Figure A2. Negative affect as a function of IAT score prediction condition and agreement with the internal non-prejudicial goal subscale (INPG) of Gawronski et al.'s (2008) non-prejudicial goal scale, Study 1.

Supplemental Materials Section B

Table B1

Average feeling thermometer ratings for different target groups as a function of IAT Score Prediction, IAT Completion, and Time (Study 1).

	IATs Completed				No IATs Completed			
	IAT Score Predictions		No IAT Score Predictions		IAT Score Predictions		No IAT Score Predictions	
Target group	t1	t2	t1	t2	t1	t2	t1	t2
Asians	64	61	57	60	64	64	64	65
Blacks	64	62	59	62	64	64	59	60
Latinos/Hispanics	64	59	59	60	62	59	57	58
Whites	66	66	64	64	65	67	65	64
Celebrities	51	64	55	59	54	58	52	51
Regular people (non-celebrities)	64	62	64	64	69	64	63	64
Children	75	77	78	78	81	79	75	75
Adults	66	64	63	64	65	65	63	62

Note. Feeling thermometer ratings were completed on scales ranging from 0 (*very cool*) to 100 (*very warm*); t1 = Time-1, t2 = Time-2.

Table B2

Average feeling thermometer ratings for different target groups as a function of IAT Score Prediction, IAT Completion, and Time (Study 2).

	IATs Completed				No IATs Completed			
	IAT Score Predictions		No IAT Score Predictions		IAT Score Predictions		No IAT Score Predictions	
Target group	t1	t2	t1	t2	t1	t2	t1	t2
Asians	61	59	61	59	65	62	56	57
Blacks	68	64	65	63	65	66	65	65
Latinos/Hispanics	66	63	66	63	51	55	51	50
Whites	68	67	66	62	68	68	64	63
Celebrities	84	53	47	48	51	55	51	50
Regular people (non-celebrities)	66	64	63	59	67	66	61	62
Children	80	82	81	78	86	84	86	85
Adults	67	66	65	62	68	67	64	64

Note. Feeling thermometer ratings were completed on scales ranging from 0 (*very cool*) to 100 (*very warm*); t1 = Time-1, t2 = Time-2.

Table B3

Average feeling thermometer ratings for different target groups as a function of IAT Score Prediction, IAT Completion, and Time (Study 3).

	IATs Completed with feedback				IATs completed without feedback				No IATs Completed			
	IAT Score Predictions		No IAT Score Predictions		IAT Score Predictions		No IAT Score Predictions		IAT Score Predictions		No IAT Score Predictions	
Target group	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
Asians	60	60	59	60	62	61	61	60	60	60	64	65
Blacks	68	67	64	65	66	64	61	59	68	65	64	64
Latinos/Hispanics	66	63	64	63	68	61	64	61	67	65	70	69
Whites	69	70	60	61	65	64	66	65	66	66	70	68
Celebrities	52	56	49	48	49	55	42	45	50	55	54	52
Regular people (non-celebrities)	64	65	60	62	66	62	59	59	66	65	64	67
Children	90	89	83	82	84	82	84	83	88	89	82	81
Adults	69	66	62	63	66	66	61	62	67	67	69	69

Note. Feeling thermometer ratings were completed on scales ranging from 0 (*very cool*) to 100 (*very warm*); t1 = Time-1, t2 = Time-2.