PNAS
www.pnas.org

Supplementary Information for

The relationship between implicit intergroup attitudes and beliefs

Benedek Kurdi[1], Thomas C. Mann, Tessa E. S. Charlesworth, and Mahzarin R. Banaji[2]

[1] To whom correspondence should be addressed: kurdi@g.harvard.edu

[2] To whom correspondence should be addressed: mahzarin_banaji@harvard.edu

This PDF file includes:

- Supplementary text
- Tables S1 to S20

Supplementary Materials and Methods (Studies 1–2)

Participants. Participants in Studies 1–2 were American volunteers recruited via the Project Implicit educational website (http://implicit.harvard.edu). Demographic data on age, gender, educational attainment, and political orientation are available in the datasets published on the Open Science Framework (OSF; https://osf.io/xyhgu/).

Participants who did not complete both Implicit Association Tests (1), which constituted the main dependent measure in all studies ($N = 312$), as well as participants whose response latencies were below 300 ms on at least 10% of IAT trials ($N = 47$) were eliminated from all subsequent analyses (2). In addition, in Study 2, participants who failed the manipulation check probing explicit recollection of the learning phase were also excluded from consideration ($N = 302$). Participant exclusions resulted in a final sample size of $N = 2{,}325$. For a breakdown of participant exclusion and sample sizes by study, see Table S1.

Statistical power. Sufficient power is particularly pertinent in the context of the present studies: Underpowered studies could inaccurately support attitude–belief dissociation by failing to detect existing relationships. Therefore, in Table 1 we report the results of a sensitivity power analysis conducted for each study. Such analyses consist of two steps: (*a*) calculation of the smallest effect size detectable by a study (r_{min}) given its sample size and (*b*) comparison of the smallest detectable effect size with the effect size actually obtained (r_{obt}). For instance, the White/Black condition of Study 1A was conducted with 282 participants. This sample size provides sufficient (.80) power to detect a minimum correlation of $r_{min} = .17$. The effect size that was obtained in this condition ($r_{obt} = .31$) considerably exceeded the r_{min} of .17. In other words, the sample size of the study was sufficient to find even a much smaller correlation than the one that emerged empirically.

More generally, if $r_{obt} > r_{min}$, this should increase confidence in the robustness of significant associations because it indicates that the sample size of the study was large enough to detect the obtained effect. On the other hand, if $r_{obt} < r_{min}$, studies (*a*) may not be able to detect an existing association or (*b*) may produce false positives (3). In the current project, studies on average had sufficient power to detect the small effect of $r_{min} = .24$; obtained effect sizes (r_{obt}) exceeded r_{min} by a mean of $r_{diff} = +.12$ (see Table S17). That is, studies were adequately powered to find

even small correlations, whereas actually obtained correlations tended to be medium-sized, thus diminishing the possibility of erroneously supporting attitude–belief dissociations.

Implicit Association Test. The Implicit Association Test (IAT) (1) used to measure implicit attitudes and implicit beliefs consisted of five blocks in all conditions of Studies 1 and 2. Stimuli are listed in Tables S2–S11, S13, and S15, and are available for download from OSF (https://osf.io/xyhgu/).

To illustrate the procedure, we use a Laapian/Niffian attitude IAT (Studies 2A and 2B) as an example. The procedure was identical for all other IATs. In the first block (category practice), participants made 20 speeded single categorization judgments of Laapian and Niffian stimuli. In the second block (attribute practice), participants made 20 speeded single categorization judgments of good and bad stimuli. In the third block (first combined block), participants made 40 speeded double categorization judgments on which Laapians and good stimuli were assigned to one response key and Niffians and bad stimuli were assigned to the other response key (or the opposite, depending on counterbalancing, see below). In the fourth block (second category practice), participants made 20 speeded single categorization judgments of Laapian and Niffian stimuli with the assignment to response keys reversed compared to the first block. Finally, in the fifth block (second combined block), participants made 40 speeded double categorization judgments on which the mapping of categories to attributes was reversed compared to the third block. The order of the two combined blocks was counterbalanced across participants such that some participants completed the Laapian/good (Niffian/bad) block first and the Niffian/good (Laapian/bad) block second, whereas for other participants the order was reversed.

Procedure of Study 2. Study 2 consisted of (*a*) a learning phase in which participants were randomly assigned to an experimental condition, involving attitude induction via evaluative conditioning (i.e., pairing group members with valenced images), or to a control condition, involving the same number of stimulus presentations without attitude induction, and (*b*) a test phase involving measurement of implicit attitudes and beliefs. Unlike Study 1, which involved real-world targets, Study 2 used fictitious social groups as targets (4, 5). The stimuli used in the learning phase of the experiment are shown in Tables S12 and S14 and available for download from OSF (https://osf.io/xyhgu/).

Learning phase (experimental condition). In the experimental condition, attitudes toward Niffians and Laapians were shifted using an evaluative conditioning paradigm (6, 7). To counteract a slight baseline preference in favor of Laapians (5), Niffians were consistently paired with positive stimuli and Laapians were consistently paired with negative stimuli. The evaluative conditioning paradigm used in the experimental condition was identical to the REP condition in Study 1 of (5), with two exceptions: (*a*) participants were exposed to 40, rather than 37, stimulus pairings and (*b*) stimulus presentations were divided into 4 blocks. Each block featured 5 presentations of a randomly selected Niffian name paired with a randomly selected positive stimulus and 5 presentations of a randomly selected Laapian name paired with a randomly selected negative stimulus.

The order of Niffian–positive and Laapian–negative trials was also individually randomized. On each trial, a Niffian name and a positive stimulus or a Laapian name and a negative stimulus were simultaneously presented next to each other in the center of the screen for 2,500 ms, followed by an intertrial interval of 1,000 ms consisting of a blank screen. For further details on the procedure, see (5) and for the verbatim text of the instructions presented to participants, see Table S16.

Learning phase (control condition). The control condition was structurally identical to the experimental condition, with the exception that category members (Laapians or Niffians) were never paired with valenced images. Rather, participants were exposed to 4 blocks of stimulus presentations in individually randomized order: (*1*) 10 pairings of a randomly selected positive image with another randomly selected positive image, (*2*) 10 pairings of a randomly selected negative image with another randomly selected negative image, (*3*) 10 pairings of a randomly selected Niffian stimulus with another randomly selected Niffian stimulus, and (*4*) 10 pairings of a randomly selected Laapian stimulus with another randomly selected Laapian stimulus. As such, the control condition provided perfect control for frequency of stimulus exposure in the experimental condition without pairing unconditioned and conditioned stimuli with each other (8).

Test phase. In the test phase, participants completed (*a*) a Niffian/Laapian attitude IAT and (*b*) a Niffian/Laapian belief IAT with the attributes “American” and “foreign” (Study 2A) or “mental” vs. “physical” (Study 2B). In Study 2A, the order of the two IATs was randomized; in

Study 2B, all participants completed the belief IAT first to provide a more conservative test of evaluative learning effects on implicit beliefs.

Variance decomposition. Measurement error is known to attenuate the correlation between manifest measures compared to their underlying correlation at the level of latent constructs (9). Therefore, even if certain attitudes and beliefs were to be conceptually redundant, the correlation between the observed attitude IAT and belief IAT scores may still remain considerably below unity due to noise in both measures. To account for this possibility, a measure of variance decomposition was obtained that partitioned variance in implicit beliefs into (*a*) error variance, (*b*) variance accounted for by attitudes, and (*c*) residual true variance. The main hypothesis tests of interest, conducted via a bootstrapping approach, probed whether (*a*) implicit attitudes accounted for a significant (i.e., nonzero) portion of variance in implicit beliefs and (*b*) residual true variance in implicit beliefs significantly differed from zero after accounting for measurement error and the effect of implicit attitudes.

Error variance (*EV*) was calculated as $EV = 1 - R_{B\text{–}IAT}$, where an estimate of the internal consistency of the belief IAT ($R_{B\text{–}IAT}$) was obtained by taking the mean of 1,000 split-half correlations (10). *Variance accounted for by the attitude (VAF)* was calculated as the square of the disattenuated correlation between attitude IAT (A–IAT) and belief IAT (B–IAT) capped at 1 to account for the possibility that estimates of the disattenuated correlation may exceed unity, i.e., $VAF = \min\left(\frac{r_{A\text{–}IAT,\, B\text{–}IAT}^{2}}{R_{A\text{–}IAT} \times R_{B\text{–}IAT}}, 1\right)$. Finally, *residual true variance* (*RTV*) was calculated as $RTV = \min(1 - EV - VAF, 0)$, capped at a minimum of 0 to account for the possibility that the estimate of EV + VAF may exceed 1. Significance tests were conducted indirectly, i.e., by calculating 95-percent bootstrap confidence intervals around each measure and examined for overlap with zero. For further details, see the analysis script made available on OSF (https://osf.io/xyhgu/).

Supplementary Study

A supplementary study was conducted to assess relative implicit evaluations of the traits *book-smart* vs. *street-smart* (Study 1C) and *mental* vs. *physical* (Studies 1D and 1E). Based on the results of Study 1C, it was hypothesized that the trait *book-smart* would be implicitly preferred to the trait *street-smart*. Based on the results of Studies 1D and 1E, it was hypothesized that the trait *mental* would be implicitly preferred to the trait *physical*.

Method.

Participants and statistical power. Participants (total $N = 596$) were American volunteers recruited from the Project Implicit educational website (http://implicit.harvard.edu). Participants who did not complete at least one of the IATs, which constituted the main dependent measures of the study ($N = 42$), as well as participants whose response latencies were below 300 ms on at least 10% of IAT trials ($N = 13$) were eliminated from all subsequent analyses (2). For the purposes of this analysis, non-White participants ($N = 134$) were excluded from consideration to ensure comparability with Studies 2 and 3. Participant exclusions resulted in a final sample size of $N = 407$.

The main hypothesis tests of interest concerned whether D scores on each IAT significantly differed from zero. For the book-smart/street-smart IAT, the final sample size was sufficient to detect an effect of Cohen's $d = 0.14$ with .80 power; for the mental/physical IAT, the final sample size was sufficient to detect an effect of Cohen's $d = 0.15$ with .80 power. These minimum detectable effect sizes are considerably below the actual effect sizes obtained (see below), thus increasing confidence in the results of the study.

Overview of the procedure. In the study, participants completed two standard five-block Implicit Association Tests (IAT; 1) in counterbalanced order: a book-smart/street-smart IAT, implemented to provide an estimate of the relative automatic evaluation of the traits *book-smart* vs. *street-smart* and a mental/physical IAT, implemented to provide an estimate of the relative automatic evaluation of the traits *mental* vs. *physical*. The order of critical blocks was independently counterbalanced within each IAT. Following the implicit measures, self-report measures of evaluation of each item used on both IATs as well as additional self-report items were also administered but are not discussed here given our focus on implicit measures of cognition. However, data obtained with explicit measures are available for follow-up analyses (https://osf.io/xyhgu/).

Materials. All stimuli used in this study are available for download from OSF (https://osf.io/xyhgu/).

Attribute stimuli. Both IATs shared attribute labels and attribute stimuli. The attribute labels "pleasant" and "unpleasant" were used. Attribute stimuli were identical to the unconditioned

stimuli used in Studies 2A and 2B (see Tables S12 and S14). That is, line drawings of positive objects (a flower, a heart, an ice cream, a sun, and a beach) served as pleasant stimuli and line drawings of negative objects (a frowny face, a fleeing man, a snake, a terrorist, and an insect) served as unpleasant stimuli. Instead of the more commonly used positive and negative words, images were selected for use in the pleasant study in order to facilitate the discriminability of attribute vs. category stimuli.

Category stimuli. Category stimuli differed depending on the IAT implemented.

For the book-smart/street-smart IAT, the category labels "book-smart" and "street-smart" were used. Category stimuli for the book-smart category included "bookish," "book-smart," "brainy," and "educated." Category stimuli for the street-smart category included "cool," "hip," "savvy," and "street-smart." It should be noted that some of the stimuli used in Study 2 were replaced to provide a more conservative test of the hypothesis that the automatic evaluation of book smarts is more positive than that of street smarts. Specifically, in the "book-smart" category, "learned" was replaced with "brainy," and in the "street-smart" category, "practical" and "shrewd" and were replaced by the more positive items "cool" and "hip."

For the mental/physical IAT, the category labels "mental" and "physical" were used. Category stimuli for the mental category included "book," "educated," "smart," and "read." Category stimuli for the mental category included "agile," "athletic," "sports," and "strong." It should be noted that some of the stimuli used in Study 1D were removed and/or replaced with the aim to create a more evaluatively balanced set of stimuli across both categories, thus providing a more conservative test of the hypothesis that the automatic evaluation of the trait *mental* is more positive than that of the trait *physical*. For instance, the extremely positive item "genius" was removed from the *mental* category.

Results and Discussion.

The main hypothesis tests of interest concerned whether automatic evaluations of the traits *book-smart* vs. *street-smart* and *mental* vs. *physical* significantly differed from zero in the expected direction.

Book-smart/street-smart IAT. As hypothesized, automatic evaluations of the trait *book-smart* were significantly more positive than automatic evaluations of the trait *street-smart*, mean

$D = 0.30$ ($SD = 0.50$), $t(406) = 12.16$, $P < 0.0001$, $BF_{10} = 8.81 \times 10^{25}$, Cohen's $d = 0.60$. This result is in line with the findings of Study 1C in which we observed a positive correlation between participants' A–IAT and book-smart/street-smart B–IAT scores such that participants with more attitudinal ingroup preference associated the trait *book-smart* stronger with the White ingroup than with the Black, Hispanic, or Asian outgroup.

Mental/physical IAT. As hypothesized, automatic evaluations of the trait *mental* were significantly more positive than their automatic evaluations of the trait *physical*, mean $D = 0.33$ ($SD = 0.46$), $t(372) = 14.22$, $P < 0.0001$, $BF_{10} = 4.10 \times 10^{33}$, Cohen's $d = 0.74$. This result is in line with the findings of Studies 1D and 1E in which we observed a positive correlation between participants' A–IAT and mental/physical B–IAT scores such that participants with more attitudinal ingroup preference associated the trait *mental* stronger with the White ingroup than with the Black outgroup.

Supplementary Materials and Methods (Study 3)

Materials. Word embeddings for the analyses reported in the main text were obtained from a pre-trained set of vectors available from the open-source FastText project of Facebook Research (https://fasttext.cc/), accessed on 5/9/2018 ("crawl-300d-2M.vec.zip") at https://fasttext.cc/docs/en/english-vectors.html. The vectors were fit by training the FastText algorithm (11) on the Common Crawl (600 billion tokens). Supplementary analyses were performed on pre-trained vectors fit by training FastText on a corpus comprised of the Wikipedia 2017, UMBC webbase corpus and statmt.org news dataset also available from FastText ("wiki-news-300d-1M.vec.zip", accessed on 5/9/2018), as well as pre-trained vectors fit by training the GloVe algorithm on the Common Crawl corpus ("glove.840B.300d.zip" accessed on 4/27/2017 at https://nlp.stanford.edu/projects/glove/) (12).

All vectors used in the analyses reported both in the main text and SI text are available in the supplementary file "Kurdi, Mann, Charlesworth, & Banaji (2018) Vectors.csv" (https://osf.io/xyhgu/). These include only the subsets of the full pre-trained sets of vectors that correspond to words selected to represent the social group and attribute categories of interest. Social groups included *Asians*, *educated*, *Jews*, *men*, *professions*, *wealthy*, *disabled*, *elderly*, *homeless*, *lower-class*, *unemployed*, *Christians*, *middle-class*, *Whites*, *women*, *Blacks*, *blue-*

collar, *gay*, *Muslims*, *Native*, *young*, and *Hispanics*. Attribute categories included *positive*, *negative*, *warm*, *cold*, *competent*, *incompetent*, *high-arousal*, and *low-arousal*.

Table S20 provides the complete set of categories and selected word stimuli within each category. Warm and competent items were obtained from the Stereotype Content Model (SCM) (13) and supplemented with synonyms using online dictionaries. Cold and incompetent items were derived from the SCM items using the antonym function of online dictionaries. Finally, positive, negative, high-arousal, and low-arousal items were taken from (14). Positive and negative items were the 200 words highest (mean = 7.78, *SD* = 0.21) and lowest (mean = 1.96, *SD* = 0.21) on valence, respectively. High-arousal (mean = 7.22, *SD* = 0.29) and low-arousal (mean = 2.56, *SD* = 0.30) items were highly discrepant on arousal but were selected to be equal on valence (mean = 4.52, *SD* = 0.14, for high-arousal vs. mean = 4.50, *SD* = 1.01, for low-arousal) to avoid confounded results.

Procedure. For each social group category and each bipolar set of attributes (i.e., positive/negative, warm/cold, competent/incompetent, and high-arousal/low-arousal), we computed a standardized measure of the relative similarity of the word vectors within the social group category and the word vectors within each of the two opposing attribute categories. This procedure employed a modified version of an algorithm (the Word Embedding Association Test, or WEAT) previously developed to compare relative group-based evaluative and stereotype associations obtained from word embeddings to parallel responses obtained from the Implicit Association Test (15). The original algorithm was designed to compute relative associations between two bipolar dimensions (e.g., Black/White and positive/negative) to achieve conceptual correspondence to scores obtained from the IAT. For the present work, the algorithm was modified to produce a measure of the relative association of a *single* category to each attribute in an opposed pair (S–WEAT) so as to achieve conceptual correspondence to a single-category version of the IAT (16).

Description of the S-WEAT algorithm. Each S–WEAT calculation involved two bipolar attributes (e.g., "warm" and "cold") and a single social group category (e.g., "Asians"). First, we computed cosine similarities between each category vector (e.g., *asian*, *Asians*, *Asian*, *asians*) and each of the positive attribute vectors and each of the negative attribute vectors. To compute a single score reflecting the relative similarity of the social category to the two attributes, we cal-

culated the mean cosine similarity of the category to each attribute and then took the difference of the two mean cosine similarity scores. Difference scores were calculated in such a way that—depending on the test—positive values reflect greater similarity between the group category and the attribute *positive*, *warm*, *competent*, or *high-arousal* over *negative*, *cold*, *incompetent*, or *low-arousal*, respectively. Finally, to parallel the IAT scoring algorithm, these difference scores were normalized to a theoretical range of -2 to +2 by dividing by a pooled standard deviation of the average cosine similarities computed for each attribute word vector across both attribute categories. Formally:

Let A and B be the two sets of opposing attribute category word vectors of equal size, and let X be the set of social group category word vectors. Let $\cos(\vec{a}, \vec{x})$ express the cosine of the angle between the two vectors $\vec{a}$ and $\vec{x}$. The S–WEAT statistic is then defined by:

$$s(X, A, B) = \frac{\text{mean}_{a \in A} s(a, X) - \text{mean}_{b \in B} s(b, X)}{\text{std\_dev}_{w \in A \cup B} s(w, X)}$$

where

$$s(a, X) = \text{mean}_{x \in X} \cos(\vec{a}, \vec{x})$$

Results with alternate corpus/training algorithm. The main text reports the pairwise correlations of the S–WEAT similarities of each of the 22 social groups to the attributes positive (vs. negative), warm (vs. cold), competent (vs. incompetent), and high-arousal (vs. low-arousal), using the vectors trained on the Common Crawl with the FastText algorithm. Below, we report the same analyses using vectors fit via an alternate training algorithm (GloVe) trained on the same corpus (Common Crawl), and using vectors fit with the same algorithm (FastText) trained on a different corpus (Wiki-News). In all cases, the findings were highly similar to those reported in the main text.

GloVe vectors trained on Common Crawl. Warmth and competence were significantly positively correlated with one another, $r = 0.85$, 95% CI: [0.67; 0.94], $P < 0.0001$, $BF_{10} = 3.18 \times 10^4$, as were warmth and valence, $r = 0.77$, 95% CI: [0.51; 0.90], $P < 0.0001$, $BF_{10} = 8.86 \times 10^2$, and competence and valence, $r = 0.86$, 95% CI: [0.68; 0.94], $P < 0.0001$, $BF_{10} = 4.39 \times 10^4$.

Once again, no correlation was found between arousal and warmth, r = -0.07, 95% CI: [-0.47; 0.37], P = 0.775, BF_{01} = 3.64, or arousal and competence, r = 0.11, 95% CI: [-0.33; 0.50], P = 0.641, BF_{01} = 3.42.

***FastText vectors trained on Wiki-News*.** Warmth and competence were significantly positively correlated with one another, r = 0.70, 95% CI: [0.40; 0.87], P = 0.0003, $BF_{10} = 1.38 \times 10^2$, as were warmth and valence, r = 0.89, 95% CI: [0.75; 0.95], $P < 0.0001$, $BF_{10} = 4.92 \times 10^5$, and competence and valence, r = 0.56, 95% CI: [0.18; 0.79], $P < 0.0001$, BF_{10} = 7.90. Once again, no correlation was found between arousal and warmth, r = -0.28, 95% CI: [-0.63; 0.16], P = 0.203, BF_{01} = 1.77, or arousal and competence, r = 0.07, 95% CI: [-0.36; 0.47], P = 0.674, BF_{01} = 3.62.

Internal consistency of high-arousal/low-arousal vectors. To ensure that a lack of correlation between warmth and arousal and competence and arousal was not due to unsatisfactory reliability of the high-arousal and low-arousal vectors, internal consistency of each was calculated in the following way. The set of high-arousal vectors and low-arousal vectors was split into two equally sized random halves. For each half, the mean distance of low-arousal vectors from each social group category was subtracted from the mean distance of high-arousal vectors from each social group category, yielding two vectors of length 22, each providing an independent measure of the relative distance of each social group category from high vs. low arousal. To calculate the internal consistency of the arousal measure, these two vectors of length 22 were correlated with each other. This calculation was repeated 1,000 times and the median of the resulting distribution of correlation coefficients is reported in the main paper as the measure of internal consistency.

Methodological Implications of the Present Project

The present results suggest that a valence confound may pose challenges to the interpretation of stereotype IAT results in terms of nonevaluative (purely semantic) differences even when (*a*) the attributes used do not blatantly differ from each other in valence (book-smart/street-smart; Study 1C), or (*b*) have even been normed to be evaluatively equal using explicit measures of attitude (mental/physical; Study 1E). In fact, the present results have revealed robust implicit preference for one attribute over the other in the absence of any explicit preference (mental/physical; Study 3). As such, it may be tempting to conclude that IAT stimuli across the two

belief attributes should be normed to be evaluatively equal using implicit, rather than explicit, measures. However, the current project, along with other work conducted in our lab, suggests that finding a pair of concepts with equivalent implicit evaluations may be prohibitively challenging. Therefore, depending on the goals of the project, researchers may choose to (*a*) norm the stereotype attributes to be evaluatively equal on explicit measures, which we have shown to substantially decrease the attitude–belief correlation, (*b*) include an additional measure of attitudes towards the target groups to statistically isolate the unique effects of the stereotype measure, or (*c*) use evaluatively discrepant attributes but recognize that the measure may not reveal much beyond generalized group evaluations.

Whichever option is selected, our results also draw attention to the importance of using analytic techniques that take into account the internal consistency of the measures used. This is true in projects, such as the current one, where the interest is in investigating relationships among implicit measures of multiple constructs. Indeed, without removing error variance from the measures using the variance decomposition measure, we would have been unable to detect that some implicit attitudes and implicit beliefs are fully redundant, rather than merely associated, with each other. Similar considerations apply when the focus is on establishing the unique relationships of correlated implicit attitudes and implicit beliefs with a third variable: Without taking into account the reliability of implicit measures, spurious evidence of incremental predicative validity may be obtained (17).

References

1. Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring individual differences in implicit cognition: The Implicit Association Test. *J Pers Soc Psychol* 74(6):1464–1480.

2. Greenwald AG, Nosek BA, Banaji MR (2003) Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *J Pers Soc Psychol* 85(2):197–216.

3. Asendorpf JB, et al. (2013) Recommendations for increasing replicability in psychology. *Eur J Pers* 27(2):108–119.

4. Gregg AP, Seibt B, Banaji MR (2006) Easier done than undone: Asymmetry in the malleability of implicit preferences. *J Pers Soc Psychol* 90(1):1–20.

5. Kurdi B, Banaji MR (2017) Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *J Exp Psychol Gen* 146(2):194–213.

6. Martin I, Levey AB (1978) Evaluative conditioning. *Advances in Behaviour Research and Therapy* 1(2):57–101.

7. Levey AB, Martin I (1975) Classical conditioning of human "evaluative" responses. *Behav Res Ther* 13(4):221–226.

8. Field AP, Davey GCL (1997) Conceptual conditioning: Evidence for an artifactual account of evaluative learning. *Learn Motiv* 28(3):446–464.

9. Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15(1):72–30.

10. Kurdi B, et al. (2018) Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *Am Psychol*:1–50.

11. Mikolov T, Grave E, Bojanowski P, Puhrsch C, Joulin A (2017) Advances in pre-training distributed word representations. 1–4.

12. Pennington J, Socher R, Manning CD (2014) GloVe: Global Vectors for Word Representation.

13. Fiske ST, Cuddy AJC, Glick P, Xu J (2002) A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *J Pers Soc Psychol* 82(6):878–902.

14. Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav Res Meth* 45(4):1191–1207.

15. Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186.

16. Karpinski A, Steinman RB (2006) The Single Category Implicit Association Test as a measure of implicit social cognition. *J Pers Soc Psychol* 91(1):16–32.

17. Westfall J, Yarkoni T (2016) Statistically controlling for confounding constructs is harder than you think. *PLoS ONE* 11(3):e0152719–22.

18. Amodio DM, Devine PG (2006) Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *J Pers Soc Psychol* 91(4):652–661.

Supplementary Tables

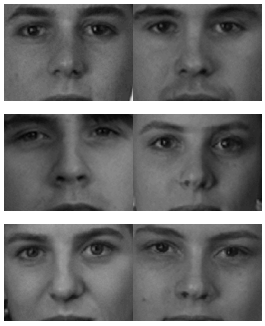
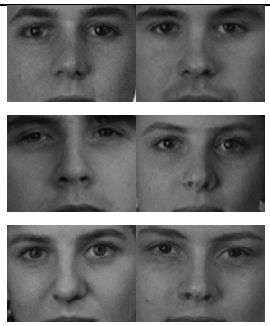
Table S1. Participant exclusions by study. IAT = Implicit Association Test, *N* = sample size.

Study	Participants	IAT categories	Belief IAT attributes	Total *N*	Did not complete both IATs	Went too fast	Failed manipulation check	Final *N*
1A	White	White Black	Smart Dumb	308	25	1	–	282
		White Hispanic		135	13	1	–	121
		White Asian		183	11	2	–	170
1B	Black	White Black		186	29	3	–	164
	Hispanic	White Hispanic		153	19	4	–	130
	Asian	White Asian		117	16	6	–	95
1C	White	White Black	Book-smart Street-smart	290	21	2	–	267
		White Hispanic		145	20	1	–	124
		White Asian		160	11	4	–	145
1D	White	White Black	Mental Physical	272	32	1	–	239
1E				225	20	–	–	205
2A	Control	Niffian Laapian	American Foreign	230	20	5	68	137
	Experimental			202	10	2	103	87
2B	Control		Mental Physical	186	34	6	47	99
	Experimental			184	31	9	84	60

Table S2. Stimulus materials for Studies 1A and 1B (White/Black name IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	African American	Good	Bad
Amy	Aliyah	Glory	Agony
Claire	Ebony	Happy	Bitter
Emily	Imani	Joy	Bomb
Katie	Jada	Love	Devil
Madeleine	Shanice	Lucky	Grief
Molly	Tiara	Peace	Hate
Cody	Darnell	Sweet	War
Connor	Malik		
Jake	Marquis		
Justin	Terell		
Tanner	Trevon		
Wyatt	Tyrone		
B–IAT			
Categories		Attributes	
European American	African American	Smart	Dumb
Amy	Aliyah		
Claire	Ebony		
Emily	Imani		
Katie	Jada		
Madeleine	Shanice	Bright	Dumb
Molly	Tiara	Clever	Ignorant
Cody	Darnell	Intelligent	Stupid
Connor	Malik	Smart	Unintelligent
Jake	Marquis		
Justin	Terell		
Tanner	Trevon		
Wyatt	Tyrone		

Table S3. Stimulus materials for Studies 1A and 1B (White/Black face IATs)*. A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	African American	Good	Bad
		Glory	Agony
		Happy	Bitter
		Joy	Bomb
		Love	Devil
		Lucky	Grief
		Peace	Hate
		Sweet	War
B–IAT			
Categories		Attributes	
European American	African American	Smart	Dumb
		Bright	Dumb
		Clever	Ignorant
		Intelligent	Stupid
		Smart	Unintelligent

* The images for this and all other studies are available separately for download from the Open Science Framework (https://osf.io/xyhgu/).

Table S4. Stimulus materials for Studies 1A and 1B (White/Hispanic IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	Hispanic American	Good	Bad
Baker	Flores	Glory	Agony
Clark	Garcia	Happy	Bitter
Davis	Gonzalez	Joy	Bomb
Hall	Hernandez	Love	Devil
Johnson	Lopez	Lucky	Grief
Jones	Martinez	Peace	Hate
Martin	Perez	Sweet	War
Miller	Ramirez		
Moore	Rivera		
Smith	Rodriguez		
Taylor	Sanchez		
Wilson	Torres		
B–IAT			
Categories		Attributes	
European American	Hispanic American	Smart	Dumb
Baker	Flores		
Clark	Garcia		
Davis	Gonzalez		
Hall	Hernandez		
Johnson	Lopez	Bright	Dumb
Jones	Martinez	Clever	Ignorant
Martin	Perez	Intelligent	Stupid
Miller	Ramirez	Smart	Unintelligent
Moore	Rivera		
Smith	Rodriguez		
Taylor	Sanchez		
Wilson	Torres		

Table S5. Stimulus materials for Studies 1A and 1B (White/Asian IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	Asian American	Good	Bad
Baker	Chen	Glory	Agony
Clark	Choi	Happy	Bitter
Davis	Chung	Joy	Bomb
Hall	Kang	Love	Devil
Johnson	Kim	Lucky	Grief
Jones	Li	Peace	Hate
Martin	Lin	Sweet	War
Miller	Liu		
Moore	Nguyen		
Smith	Pham		
Taylor	Wang		
Wilson	Yang		
B–IAT			
Categories		Attributes	
European American	Asian American	Smart	Dumb
Baker	Chen		
Clark	Choi		
Davis	Chung		
Hall	Kang		
Johnson	Kim	Bright	Dumb
Jones	Li	Clever	Ignorant
Martin	Lin	Intelligent	Stupid
Miller	Liu	Smart	Unintelligent
Moore	Nguyen		
Smith	Pham		
Taylor	Wang		
Wilson	Yang		

Table S6. Stimulus materials for Study 1C (White/Black name IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	African American	Good	Bad
Amy	Aliyah	Glory	Agony
Claire	Ebony	Happy	Bitter
Emily	Imani	Joy	Bomb
Katie	Jada	Love	Devil
Madeleine	Shanice	Lucky	Grief
Molly	Tiara	Peace	Hate
Cody	Darnell	Sweet	War
Connor	Malik		
Jake	Marquis		
Justin	Terell		
Tanner	Trevon		
Wyatt	Tyrone		
B–IAT			
Categories		Attributes	
European American	African American	Book-smart	Street-smart
Amy	Aliyah		
Claire	Ebony		
Emily	Imani		
Katie	Jada		
Madeleine	Shanice	Book-smart	Street-smart
Molly	Tiara	Educated	Practical
Cody	Darnell	Learned	Savvy
Connor	Malik	Bookish	Shrewd
Jake	Marquis		
Justin	Terell		
Tanner	Trevon		
Wyatt	Tyrone		

Table S7. Stimulus materials for Study 1C (White/Black face IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

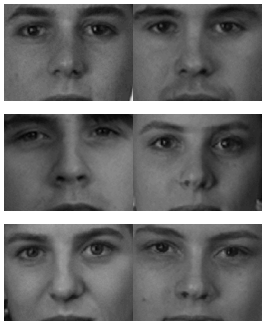
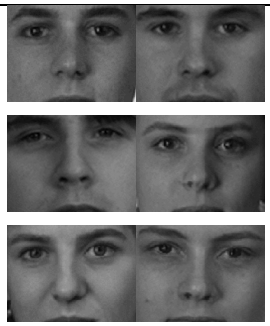
A–IAT			
Categories		Attributes	
European American	African American	Good	Bad
		Glory	Agony
		Happy	Bitter
		Joy	Bomb
		Love	Devil
		Lucky	Grief
		Peace	Hate
		Sweet	War
B–IAT			
Categories		Attributes	
European American	African American	Book-smart	Street-smart
		Book-smart	Practical
		Bookish	Savvy
		Educated	Shrewd
		Learned	Street-smart

Table S8. Stimulus materials for Study 1C (White/Hispanic IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	Hispanic American	Good	Bad
Baker	Flores	Glory	Agony
Clark	Garcia	Happy	Bitter
Davis	Gonzalez	Joy	Bomb
Hall	Hernandez	Love	Devil
Johnson	Lopez	Lucky	Grief
Jones	Martinez	Peace	Hate
Martin	Perez	Sweet	War
Miller	Ramirez		
Moore	Rivera		
Smith	Rodriguez		
Taylor	Sanchez		
Wilson	Torres		
B–IAT			
Categories		Attributes	
European American	Hispanic American	Book-smart	Street-smart
Baker	Flores		
Clark	Garcia		
Davis	Gonzalez		
Hall	Hernandez		
Johnson	Lopez	Book-smart	Practical
Jones	Martinez	Bookish	Savvy
Martin	Perez	Educated	Shrewd
Miller	Ramirez	Learned	Street-smart
Moore	Rivera		
Smith	Rodriguez		
Taylor	Sanchez		
Wilson	Torres		

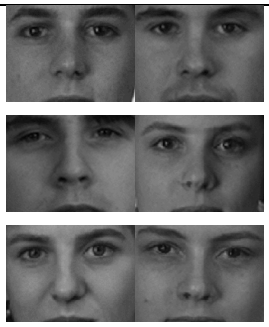
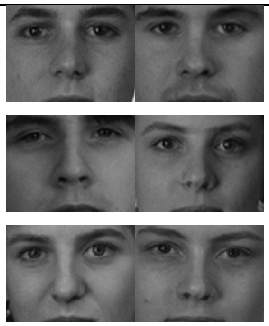
Table S9. Stimulus materials for Study 1C (White/Asian IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	Asian American	Good	Bad
Baker	Chen	Glory	Agony
Clark	Choi	Happy	Bitter
Davis	Chung	Joy	Bomb
Hall	Kang	Love	Devil
Johnson	Kim	Lucky	Grief
Jones	Li	Peace	Hate
Martin	Lin	Sweet	War
Miller	Liu		
Moore	Nguyen		
Smith	Pham		
Taylor	Wang		
Wilson	Yang		
B–IAT			
Categories		Attributes	
European American	Asian American	Book-smart	Street-smart
Baker	Chen		
Clark	Choi		
Davis	Chung		
Hall	Kang		
Johnson	Kim	Book-smart	Practical
Jones	Li	Bookish	Savvy
Martin	Lin	Educated	Shrewd
Miller	Liu	Learned	Street-smart
Moore	Nguyen		
Smith	Pham		
Taylor	Wang		
Wilson	Yang		

Table S10. Stimulus materials for Study 1D (White/Black IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
European American	African American	Pleasant	Unpleasant
		Diamond	Abuse
		Freedom	Bomb
		Heaven	Cancer
		Honest	Disaster
		Honor	Evil
		Love	Murder
		Loyal	Poverty
		Lucky	Rotten
		Peace	Sickness
		Rainbow	Vomit
B–IAT			
Categories		Attributes	
European American	African American	Mental	Physical
		Aptitude	Agile
		Book	Athletic
		Brainy	Basketball
		College	Boxing
		Educated	Dance
		Genius	Football
		Math	Jump
		Read	Rhythmic
		Science	Run
		Smart	Track

Table S11. Stimulus materials for Study 1E (White/Black IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.*

A–IAT			
Categories		Attributes	
European American	African American	Pleasant	Unpleasant
		Glory	Agony
		Happy	Bitter
		Joy	Bomb
		Love	Devil
		Sweet	Hate
B–IAT			
Categories		Attributes	
European American	African American	Mental	Physical
		Book	Agile
		Brainy	Athletic
		Educated	Baseball
		Read	Dance
		Smart	Run

* The attribute labels were identical to Study 1D. However, some of the attribute stimuli used on the A–IAT by (18), and thus in Study 1D, were stereotypically associated with the target categories (e.g., "murder" and "poverty" as negative attributes are also associated with Black rather than White Americans). Moreover, the stimuli used to represent the mental attribute on the belief IAT were explicitly evaluated by Project Implicit participants from Study 1D as more positive ($M = 6.30$, $SD = 1.27$) than the stimuli used to represent the physical attribute ($M = 5.55$, $SD = 1.21$), $t(235) = 9.16$, $P < 0.0001$, $BF_{10} = 1.85 \times 10^{14}$, Cohen's $d = 0.60$. These two aspects of the stimuli may have inflated the correlation between the attitude IAT and belief IAT in Study 1D.

To offer a cleaner test of one-type vs. two-type theories, Study 1E included only positive and negative attribute stimuli without stereotypic association with the target categories on the attitude IAT. The belief IAT used a subset of the stimuli created by (18) that were rated as equally positive across the two attributes in Study 1D ($M_{mental} = 6.06$, $SD_{mental} = 1.34$, vs. $M_{physical} = 6.07$, $SD_{physical} = 1.33$), $t(235) = 0.09$, $P = 0.925$, $BF_{01} = 13.67$, Cohen's $d = 0.01$.

Table S12. Stimulus materials for Study 2A (learning phase).

Conditioned stimuli (CSs)		**Unconditioned stimuli (USs)**	
Niffians	Laapians	Pleasant	Unpleasant
Ibbonif	Caalap		
Jabbunif	Feelslap		
Lebbunif	Gabeelap		
Mettanif	Ineelap		
Oballnif	Maasolap		
			

Table S13. Stimulus materials for Study 2A (Niffian/Laapian IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
Niffians	Laapians	Pleasant	Unpleasant
		Glory	Agony
Ibbonif	Caalap	Happy	Bitter
Jabbunif	Feelslap	Joy	Bomb
Lebbunif	Gabeelap	Love	Devil
Mettanif	Ineelap	Peace	Failure
Oballnif	Maasolap	Success	Hate
		Sweet	War
B–IAT			
Categories		Attributes	
Niffians	Laapians	American	Foreign
			
Ibbonif	Caalap		
Jabbunif	Feelslap		
Lebbunif	Gabeelap		
Mettanif	Ineelap		
Oballnif	Maasolap		
			

Table S14. Stimulus materials for Study 2B (learning phase).

Conditioned stimuli (CSs)		**Unconditioned stimuli (USs)**	
Niffians	Laapians	Pleasant	Unpleasant
Ibbonif	Caalap		
Jabbunif	Feelslap		
Lebbunif	Gabeelap		
Mettanif	Ineelap		
Oballnif	Maasolap		

Table S15. Stimulus materials for Study 2B (Niffian/Laapian IATs). A–IAT = attitude Implicit Association test, B–IAT = belief Implicit Association Test.

A–IAT			
Categories		Attributes	
Niffians	Laapians	Pleasant	Unpleasant
		Glory	Agony
Ibbonif	Caalap	Happy	Bitter
Jabbunif	Feelslap	Joy	Bomb
Lebbunif	Gabeelap	Love	Devil
Mettanif	Ineelap	Peace	Failure
Oballnif	Maasolap	Success	Hate
		Sweet	War
B–IAT			
Categories		Attributes	
Niffians	Laapians	Mental	Physical
Ibbonif	Caalap	Book	Agile
Jabbunif	Feelslap	Brainy	Athletic
Lebbunif	Gabeelap	Educated	Baseball
Mettanif	Ineelap	Read	Dance
Oballnif	Maasolap	Smart	Run

Table S16. Verbatim text of the instructions from Studies 2A and 2B by condition. Double slashes indicate that the given section of the instructions was displayed on a new screen.

Condition	Instructions
Control	In this experiment you will see two groups of people (Laapians and Niffians) as well as pleasant things and unpleasant things. Laapians have names that end with the syllable -lap, such as Eloolap, Le-boolap, or Ufaalap. Niffians have names that end with the syllable -nif, such as Albonnif, Nollinif, or Wubbonif. // Here is the full set of names that you will see: [*table with names of Laapians in the top row and names of Niffians in the bottom row*]. // Here is the full set of pleasant and unpleasant things that you will see: [*table with positive US in the top row and negative US in the bottom row*]. // This is one of four blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Block 1 of stimulus presentations]* // This is one of four blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Block 2 of stimulus presentations]* // This is one of four blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Block 3 of stimulus presentations]* // This is one of four blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Block 4 of stimulus presentations]*
Experimental	In this experiment you will learn that a certain group of people (Laapians or Niffians) is associated with pleasant things or unpleasant

	things. Laapians have names that end with the syllable -lap, such as Eloolap, Leboolap, or Ufaalap. Niffians have names that end with the syllable -nif, such as Albonnif, Nollinif, or Wubbonif. // Here is the full set of names that you will see: [*table with names of Laapians in the top row and names of Niffians in the bottom row*]. // Here is the full set of pleasant and unpleasant things that you will see: [*table with positive US in the top row and negative US in the bottom row*]. // This is block 1 out of 4 blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[First block of EC]* // This is block 2 out of 4 blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Second block of EC]* // This is block 3 out of 4 blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Third block of EC]* // This is block 4 out of 4 blocks of learning. After you hit SPACE, sit back and watch the pairings. As you watch, make sure that you learn the relationship between the names and the things you see. This part of the study will take about 30 seconds to complete. // *[Fourth block of EC]*

Table S17. Summary of the results of Studies 1–2. N = final sample size; r_{min} = smallest correlation for which the sample size provides adequate (.80) power; IAT = Implicit Association Test; A–IAT = attitude IAT; B–IAT = belief IAT; BF_{10} = Bayes Factor in favor of the alternative hypothesis (i.e., correlation between A–IAT and B–IAT). IATs are scored such that positive D scores indicate implicit preference in favor of the first category or association of the first category with the first attribute. 95-percent confidence intervals for comparing means, correlations, and variance components to zero are reported in square brackets.

Study	Subjects	N	r_{min}	IAT categories	B–IAT attributes	A–IAT mean (D)	B–IAT mean (D)	A–B IAT Correlation	BF_{10}	B–IAT variance components		
										Error variance	Variance accounted for by attitude	Residual true variance
1A	White	282	.17	White Black	Smart Dumb	0.39 [0.34; 0.44]	0.33 [0.28; 0.38]	.31 [.20; .41]	8.64×10^{4}	34.70 [26.84; 45.07]	23.66 [9.23; 44.45]	41.65 [16.12; 59.27]
		121	.25	White Hispanic		0.36 [0.29; 0.43]	0.29 [0.21; 0.37]	.61 [.49; .71]	8.80×10^{10}	33.07 [22.43; 46.69]	90.99 [55.70; 100.00]	0.24 [0.00; 14.10]
		170	.21	White Asian		0.27 [0.21; 0.33]	0.19 [0.13; 0.25]	.34 [.20; .47]	2.58×10^{3}	40.24 [30.64; 52.74]	32.57 [14.68; 64.05]	27.20 [0.00; 49.39]
1B	Black	164	.22	White Black		-0.06 [-0.14; 0.01]	-0.04 [-0.10; 0.02]	.37 [.22; .49]	9.29×10^{3}	37.50 [26.46; 50.93]	29.07 [10.15; 56.32]	33.43 [0.44; 56.34]
	Hispanic	130	.24	White Hispanic		-0.07 [-0.15; 0.01]	0.00 [-0.07; 0.07]	.47 [.33; .60]	9.94×10^{5}	34.83 [23.52; 50.19]	49.99 [21.98; 82.95]	15.18 [0.00; 45.10]
	Asian	95	.28	White Asian		-0.09 [-0.18; -0.01]	-0.06 [-0.13; 0.01]	.21 [.00; .39]	9.08×10^{-1}	50.23 [33.03; 68.81]	13.23 [0.06; 52.69]	36.54 [0.00; 60.41]
1C	White	267	.17	White Black	Book-smart Street-smart	0.42 [0.37; 0.47]	0.33 [0.28; 0.38]	.27 [.16; .38]	2.38×10^{3}	33.10 [25.30; 43.95]	17.45 [5.13; 36.88]	49.45 [25.62; 64.77]
		124	.25	White Hispanic		0.37 [0.30; 0.44]	0.36 [0.29; 0.43]	.15 [-.02; .32]	4.67×10^{-1}	41.72 [27.24; 59.04]	6.48 [0.02; 30.76]	51.80 [18.31; 69.33]
		145	.23	White Asian		0.35 [0.29; 0.42]	0.22 [0.15; 0.29]	.25 [.09; 40]	1.01×10^{1}	32.47 [22.98; 44.53]	15.71 [2.31; 37.69]	51.83 [23.76; 68.84]
1E	White	239	.18	White Black	Mental Physical	0.31 [0.25; 0.36]	0.27 [0.22; 0.32]	.28 [.16; .39]	1.28×10^{3}	21.98 [16.37; 28.52]	13.78 [4.05; 27.29]	64.25 [48.95; 75.51]
1D		205	.19			0.30 [0.24; 0.36]	0.24 [0.19; 0.29]	.15 [.02; .28]	9.63×10^{-1}	23.69 [17.06; 32.68]	4.67 [0.31; 16.40]	71.65 [57.27; 80.58]
2A	Control	137	.24	Niffian Laapian	American Foreign	-0.16 [-0.23; -0.09]	-0.18 [-0.24; -0.11]	.39 [.23; .52]	4.77×10^{3}	36.81 [24.72; 52.31]	36.05 [12.75; 68.27]	27.14 [0.00; 54.25]
	Experimental	87	.30			0.22 [0.11; 0.33]	0.05 [-0.05; 0.14]	.60 [.45; .72]	1.75×10^{7}	28.82 [17.47; 43.67]	64.82 [38.53; 94.07]	6.36 [0.00; 37.64]
2B	Control	99	.28		Mental Physical	-0.13 [-0.22; -0.04]	0.02 [-0.08; 0.11]	.50 [.33; .63]	9.84×10^{4}	29.65 [19.27; 43.65]	51.53 [28.40; 82.28]	18.82 [0.00; 45.33]
	Experimental	60	.35			0.22 [0.09; 0.34]	0.26 [0.15; 0.37]	.46 [.23, .64]	1.20×10^{2}	32.38 [17.99; 53.98]	42.22 [14.57; 94.36]	25.39 [0.00; 59.74]

Table S18. Summary of the results obtained using explicit measures (Studies 1–2). *N* = final sample size; r_{min} = smallest correlation for which the sample size provides adequate (.80) power; A–B correlation = attitude–belief correlation; BF_{10} = Bayes Factor in favor of the alternative hypothesis (i.e., correlation between A–IAT and B–IAT). Measures are scored such that positive difference scores indicate explicit preference in favor of the first category or association of the first category with the first attribute. 95-percent confidence intervals for comparing means and correlations to zero are reported in square brackets.

Study	Subjects	N	r_{min}	Categories	Attributes	Explicit attitude mean	Explicit belief mean	A–B correlation	BF10
1A[†]	White	297	.16	White Black	Smart Dumb	-0.48 [-0.62; -0.35]	0.33 [0.24; 0.41]	.45 [.35; .54]	1.59×10^{13}
		316	.16	White Hispanic		0.33 [0.19; 0.47]	0.38 [0.29; 0.48]	.34 [.23; .43]	8.19×10^{6}
		316	.16	White Asian		0.04 [-0.08; 0.17]	-0.41 [-0.51; -0.32]	.09 [-.01; .20]	2.59×10^{-1}
1B	African American	95	.28	White Black		-0.13 [-0.33; 0.06]	-0.17 [-0.36, 0.03]	.33 [.14; .50]	2.47×10^{1}
	Hispanic	93	.29	White Hispanic		-0.27 [-0.50; -0.04]	0.05 [-0.15; 0.25]	.49 [.32; .63]	3.02×10^{4}
	Asian	84	.30	White Asian		-0.30 [-0.51; -0.09]	-0.60 [-0.80; -0.41]	.18 [-.03; .38]	5.47×10^{-1}
1C[‡]	White	264	.17	White Black	Book-smart Street-smart	–	0.90 [0.75; 1.07]	–	–
		123	.25	White Hispanic		–	0.85 [0.60; 1.11]	–	–
		143	.23	White Asian		–	-1.14 [-1.38; -0.91]	–	–
1D	White	216	.19	White Black	Mental Physical	0.19 [0.05; 0.35]	0.76 [0.61; 0.91]	.40 [.28; .51]	1.02×10^{7}

[†] Due to a coding error, these data were not collected as part of Study 1A but rather as part of the Supplementary Study available on the Open Science Framework (https://osf.io/xyhgu/).

[‡] Due to a coding error, no explicit attitude measures were collected in Study 1C.

1E		187	.20			0.34 [0.18; 0.49]	0.80 [0.63; 0.97]	.48 [.36; .58]	6.62×10^{10}
2A	Control	118	.25	Niffian Laapian	American Foreign	-0.04 [-0.14; 0.06]	-0.19 [-0.39; 0.00]	.19 [.01; .36]	9.83×10^{-1}
	Experimental	79	.31			1.90 [1.39; 2.41]	0.31 [-0.17; 0.80]	.34 [.13; .52]	1.55×10^{1}
2B	Control	87	.30		Mental Physical	-0.31 [-0.60; -0.02]	-0.09 [-0.33; 0.15]	.41 [.22; .57]	4.12×10^{3}
	Experimental	57	.36			2.07 [1.45; 2.68]	0.86 [0.36; 1.37]	.03 [-.24; .28]	1.44×10^{-1}

Table S19. Comparison of implicit vs. explicit measures (Studies 1–2). To enable easy comparison, means of both implicit and explicit measures are reported as Cohen's *d* effect sizes (with a reference point of zero). A–B correlation = attitude–belief correlation. Measures are scored such that positive difference scores indicate implicit or explicit preference in favor of the first category or association of the first category with the first attribute. 95-percent confidence intervals for comparing to zero are reported in square brackets.

Study	Subjects	Categories	Attributes	Implicit attitude mean	Explicit attitude mean	Implicit belief mean	Explicit belief mean	Implicit A–B correlation	Explicit A–B correlation
1A	White	White Black	Smart Dumb	0.98	-0.40	0.80	0.43	.31 [.20; .41]	.45 [.35; .54]
		White Hispanic		0.89	0.27	0.67	0.46	.61 [.49; .71]	.34 [.23; .43]
		White Asian		0.69	0.04	0.47	-0.48	.34 [.20; .47]	.09 [-.01; .20]
1B	African American	White Black		-0.13	-0.14	-0.11	-0.17	.37 [.22; .49]	.33 [.14; .50]
	Hispanic	White Hispanic		-0.15	-0.24	0.00	0.05	.47 [.33; .60]	.49 [.32; .63]
	Asian	White Asian		-0.22	-0.31	-0.17	-0.67	.21 [.00; .39]	.18 [-.03; .38]
1C	White	White Black	Book-smart Street-smart	1.06	–	0.78	0.70	.27 [.16; .38]	–
		White Hispanic		0.92	–	0.95	0.59	.15 [-.02; .32]	–
		White Asian		0.91	–	0.51	-0.80	.25 [.09; 40]	–
1D	White	White Black	Mental Physical	0.69	0.17	0.66	0.66	.28 [.16; .39]	.40 [.28; .51]
1E				0.69	0.31	0.66	0.66	.15 [.02; .28]	.48 [.36; .58]
2A	Control	Niffian Laapian	American Foreign	-0.38	0.07	-0.45	-0.17	.39 [.23; .52]	.19 [.01; .36]
	Experimental			0.42	0.82	0.10	0.14	.60 [.45; .72]	.34 [.13; .52]

2B	Control		Mental Physical	-0.28	0.22	0.04	-0.07	.50 [.33; .63]	.41 [.22; .57]
	Experimental			0.45	0.87	0.60	0.45	.46 [.23, .64]	.03 [-.24; .28]

Table S20. List of all word stimuli for which vector embeddings were used to represent social group and attribute categories (Study 3).

Social group category	Word set
Asians	asian, Asians, Asian, asians
Educated	educated, intellectual, academic, learned
Jews	Jew, Jewish, Jews, jewish
men	men, man, male, he
professionals	professional, professionals, white-collar, executive
wealthy	rich, wealthy, affluent, upper-class
disabled	disabled, handicapped, retarded, incapacitated
elderly	elderly, old, older, senior
homeless	homeless, supplicant, beggar, mendicant
lower-class	poor, low-income, low-class, lower-class
unemployed	unemployed, unwaged, jobless, unoccupied
Christians	Christian, Christians, Catholic, Protestant
middle-class	middle-class, average-income*, middle-income, Middle-Class
Whites	White, Whites, European, Caucasian
women	women, woman, female, she
Blacks	Black, Blacks, African, African-American

* This stimulus did not occur in the fastText Common Crawl corpus; therefore, for the middle-class category, S–WEAT scores were calculated on the basis of three stimuli.

blue-collar	blue-collar, working-class, laborer, Blue-collar
gay	gay, lesbian, LGBT, homosexual
Muslims	Muslims, Muslim, muslims, muslim
Native	Native, indigenous, native, aboriginal
young	young, child, youth, teenager
Hispanics	Hispanics, Hispanic, Latin, Latino
Attribute category	**Word set**
Warm	friendly, well-intentioned, trustworthy, warm, good-natured, sincere, nice, kind, dependable, agreeable, supportive
Cold	unfriendly, hateful, untrustworthy, cold, mean, dishonest, vicious, deceitful, disloyal, selfish, hostile
Competence	competent, confident, capable, efficient, intelligent, skillful, skilled, qualified, proficient, able, smart
Incompetence	incompetent, uncertain, ignorant, inefficient, unintelligent, unskilled, helpless, inept, unqualified, dumb, foolish
Positive	new, good, love, free, give, play, music, live, care, fun, pretty, kind, million, create, beautiful, song, happy, enjoy, rest, thanks, amazing, travel, excellent, safe, awesome, relationship, loved, successful, thank, funny, positive, healthy, sweet, comfortable, cute, friendly, tree, beauty, spring, mom, glad, peace, lovely, fantastic, excited, marriage, achieve, exciting, vacation, entertainment, freedom, award, smart, chocolate, winner, honest, smile, cake, pleasure, musical, victory, pleased, incredible, gorgeous, savings, laugh, joy, bonus, resort, celebrate, intelligence, joke, comedy, healing, fabulous, confident, talented, golden, romantic, prize, relax, girlfriend, wisdom, entertaining, pizza, intelligent, humor, happiness, intellectual, kiss, cure, enjoyable, relaxing,

	creativity, excitement, peaceful, wildlife, lover, determination, christmas, friendship, payday, angel, hilarious, praise, puppy, magical, sunny, courage, exotic, delighted, excellence, amazed, delight, inexpensive, luxurious, treasure, thankful, faithful, grandmother, knowledgeable, relaxation, bargain, enjoyment, princess, harmony, laughter, sunshine, compassion, breeze, enthusiastic, wellness, orgasm, hug, energetic, abundant, literacy, feast, pleasing, kindness, amusing, aroma, affection, mama, playful, splendid, accomplishment, daytime, humorous, heavenly, grin, kitten, cheerful, compassionate, w, greatness, comedian, bliss, waterfall, sweetheart, admiration, honorable, joyful, seaside, courageous, oasis, spirited, pleasurable, cheesecake, sweetie, giggle, prosper, panda, lovable, sincerity, pancake, serenity, harmonious, fudge, smiley, tranquility, smoothie, winnings, excite, cuddle, gratification, companionship, liberating, macaroni, fondness, blissful, springtime, snuggle, lullaby, seashore, euphoric, lovemaking, sundae, starlight, motherly
Negative	death, pain, dead, cancer, disease, traffic, debt, attack, hate, die, kill, worry, abuse, stress, sad, failure, crime, crisis, criminal, killing, prison, murder, terrible, enemy, virus, infection, suffer, victim, illness, disaster, harm, fraud, poverty, jail, HIV, suicide, killer, acne, assault, rape, steal, funeral, foreclosure, AIDS, slave, pollution, racist, torture, terrorism, deadly, homeless, tragedy, rude, fatal, tumor, headache, racism, nightmare, devastating, slavery, unhappy, asshole, sue, poison, bombing, neglect, wreck, abusive, worthless, incest, misery, epidemic, nausea, greedy, diarrhea, traumatic, genocide, rob, horrific, bankrupt, assassination, disastrous, insulting, catastrophic, massacre, bury, attacker, hopeless, homicide, dreaded, murderer, menopause, herpes, mourning, leukemia, STD, kidnapping, suicidal, coma, interrogation, alcoholism, scum, hateful, vile, vomit, unethical, starvation, assassin, breakup, parasite, unbearable, negligent, vandalism, perish, sabotage, murderous, chemo, fag, humiliating, abduction, incarcera-

	tion, feces, rabies, destroyer, parasitic, manslaughter, puke, nigger, infidelity, rapist, kidnap, motherfucker, dishonesty, hijack, extortion, excruciating, unhappiness, heartbroken, repulsive, pedophile, satanic, pimple, destitute, felon, pollute, cirrhosis, smallpox, overworked, mistrust, syphilis, amputation, incurable, gunpoint, nauseous, roach, vengeful, deathly, foreclose, morgue, moldy, embolism, guillotine, excrement, homicidal, deathbed, lynching, gonorrhea, comatose, appendicitis, mortuary, molester, executioner, penitentiary, bloodbath, molest, frostbite, unsanitary, maggot, kidnapper, hijacker, sadist, mutilate, tapeworm, asphyxiation, seasick, mugger, bubonic, flunk, castrate, hellhole, amputate, motherless, asphyxia, parkinsons, stomachache, cellmate, drunk-driving, blackmailer, heart-disease, breaking-and-entering
High-arousal	party, war, speed, military, heat, impact, attack, hate, exciting, birth, worry, sexual, environmental, winner, evil, discover, nuclear, electrical, investigation, blonde, adventure, warning, fantasy, complicated, diamond, lesbian, horrible, dramatic, rapid, succeed, violent, suicide, gambling, glory, excitement, bomb, revolution, passionate, intensity, scholarship, collapse, panic, spouse, cinema, desirable, audit, bullet, vaccine, frustration, jealous, flame, rat, reasoning, bust, nightmare, lottery, blowjob, scare, compliment, freezing, hype, lightning, freak, overtime, asshole, stimulate, dominate, embarrassing, immense, lick, disgusting, advancement, pistol, jerk, thrill, outgoing, contamination, stimulating, invasive, accelerate, caffeine, confront, poisoning, thief, daring, doom, genocide, striker, plague, pornography, furious, paycheck, assassination, nightlife, penetrate, cruelty, countdown, fury, badass, irresistible, sinful, agony, genital, mating, burglary, mindful, hustle, millionaire, victorious, famine, hateful, craze, prosper, illuminate, individuality, horrendous, oppressive, sneaky, motorbike, cringe, bestseller, miscarriage, blitz, faggot, goddamn, martini, treacherous, adamant, stalker, landslide, pornographic, cheery, overkill, gunfire, stimulant, jackass, wickedness, defiant, scandal-

	ous, birthing, euphoria, flirty, fanatical, avenge, agonizing, scorpion, superstitious, energize, extraterrestrial, amphibious, infuriating, terminator, typhoon, bulletproof, contaminate, hyperactive, wail, claustrophobic, machete, revolting, swat, telepathy, invader, aggressor, foreclose, drunkenness, confiscate, frisky, implosion, vasectomy, impeach, matrimony, manhunt, flatulence, promiscuity, penniless, agitate, shitload, goddamned, screamer, kryptonite, flamethrower, abduct, necromancer, adulterer, incriminate, showgirl, frisk, shithead, mooch, unbutton, swordplay, gunboat, asphyxia, blowtorch, pisser, carjack, breaking-and-entering
Low-arousal	back, low, code, prior, foot, concept, panel, district, medium, trailer, copyright, percentage, generic, lock, conservative, dose, operator, sec, blank, paste, par, coal, preserve, elderly, brick, comply, spokesman, trace, distant, lung, opt, fare, doctrine, conclude, bush, bucket, sigh, tin, drawer, parliament, colonial, cardboard, tenant, canned, vacant, lever, cubic, stall, damp, businessman, mediocre, hallway, gravel, memo, cane, duct, stool, digest, cabbage, axle, solvent, magnesium, limb, shareholder, valium, mosque, fiberglass, potty, casing, clergy, derive, solitary, inflatable, crate, slab, monk, salesman, chimney, monastery, catholic, plywood, longitude, mower, loader, tack, hog, summarize, needless, rag, constituency, peptide, hanger, syllabus, managerial, berth, tram, blurry, clipboard, remover, omission, minivan, dormant, concede, solemn, sac, appendix, synagogue, postpone, pamphlet, mule, canister, pancreas, defer, moth, ointment, winger, rein, bonnet, hind, wiper, incomparable, sitter, picket, chromium, sod, lukewarm, brood, chute, beak, consulate, rudimentary, bran, syllable, selenium, statesman, daft, uneventful, lint, tarp, bale, dork, gallbladder, hairline, flue, musk, bard, mantel, tartar, pail, janitor, swab, monogrammed, mime, deft, prawn, dictation, boron, annals, drowsy, schooner, washroom, silo, bromide, remiss, motley, linoleum, barium, ordnance, burr, protestant, menthol, whitewash, necktie, celluloid, downfield, insipid, platonic, pestle,

	doorknob, weatherman, untie, gavel, humdrum, caddie, brown-stone, retractor, seltzer, newscaster, undersecretary, birthmark, plainclothes, nunnery, pillbox, selectman, midshipman, dustpan, drainpipe, bagman, decimeter