Supporting Information

Stanley et al. 10.1073/pnas.1014345108

SI Text

Exclusion Methodology and Criteria for Participants from Study 2. Examination of individuals' monetary offers revealed a distinct subset of individuals ($n = 14$) who used a rule-based strategy when making offers (e.g., always offer \$10), and therefore did not consider each interaction independently. These participants were excluded from further analysis, which had no significant effects on the results reported here. To quantify the visually observed differences in strategies used by the participants in study 2, we calculated the number of identical consecutive offers for each participant. Examination of the distribution of this metric across the population confirmed the existence of two distinct subgroups within the population (Fig. S3). To determine the criterion at which to split the data objectively, we performed a likelihood ratio (LR) test on this distribution. We took the ratio of the likelihood of the data given the "full model" to that given a "restricted model." For the full model, we split the data at a range of values and calculated the likelihood of the data for each subgroup independently (i.e., drawn from separate Gaussian distributions); we then multiplied the likelihoods of the two subgroups together. The restricted model calculated the likelihood of the data given that all points were drawn from a single Gaussian distribution. Maximum likelihood procedures were used in calculating the split value and the mean and variance for each Gaussian distribution. The LR test statistic $-2\log(\mathrm{LR})$ is asymptotically χ^2-distributed with 3 df for the three extra parameters (split value and the mean and SD of the second group). The maximum LR value corresponded to a split at 194 identical consecutive offers (more than two-thirds of trials) and an overall minimum P value of 2.12×10^{-18}. Those participants above the split value were excluded from subsequent analysis. The data of the remaining 43 participants (overall μ = \$3.77, SD = \$1.77; black μ = \$3.74, SD = \$1.99; white μ = 3.75, SD = \$1.72, other race μ = \$3.81, SD = \$1.87) did not differ significantly from the total pool of 57 participants (overall μ = \$4.02, SD = \$2.55; black μ = \$4.00, SD = \$2.69; white μ = \$4.00, SD = \$2.49; other race μ = \$4.06, SD = \$2.62). In addition, the correlation between IAT score and offer disparity remained significant even when the participants excluded from study 2 were included in the analysis ($r = 0.30$, $P = 0.025$).

Bootstrap Analysis. We were interested in the robustness of the correlation between rating/offer disparity and implicit race bias (IAT D score). To test this, we ran bootstrap analyses to determine the minimum number of trials required to replicate the positive correlation reliably. For each sample size (1–91 trials from each race category), we took 2,500 samples with replacement from each participant. For each sample, values were converted to z scores and each participant's rating/offer disparity was calculated [Mean(white z score) − Mean(black z score)]. Finally, the correlation (Pearson's r) between the rating/offer disparities and the implicit race biases of the population was calculated. For each experiment, this resulted in 91 distributions (1 per sample size) of 2,500 r values each, from which means and 95% confidence intervals were determined (Fig. S2). Individual differences in implicit race bias were positively correlated with rating disparity in more than 95% of random samples of three or more ratings from each race category. The same was true for random samples of five or more offers from each race category. The robustness of the rating/offer disparity was also evidenced when we correlated that of each sample with the results from the full experiment. Even at samples of 1 trial per condition, more than 95% of samples' rating/offer disparity correlated positively with the rating/offer disparity calculated using all trials.

Separate Contribution of Raw Black and White Ratings and Offers to the Relationship Between Implicit Race Bias and Trust Disparity. It is possible that the relationship between implicit race attitude and ratings/offers disparity was predominantly driven by evaluations of either black or white faces but not both. To investigate this, we examined the correlations between IAT score and mean raw black and white ratings/offers separately (Fig. S1). First, we established that we were able to replicate our main finding (IAT score correlates with trust disparity) with the raw response data. IAT score was significantly correlated with trust disparity [Mean (black) − Mean(white)] in ratings [$r(48) = 0.3742$, $P = 0.0073$] and offers [$r(41) = 0.3410$, $P = 0.0252$]. We then examined whether this effect was more a result of the black or white responses. Individual differences in IAT score were significantly correlated with mean ratings for white faces [$r(48) = 0.3078$, $P = 0.0297$] but not for black faces [$r(48) = -0.0180$, $P = 0.9014$)]; individuals whose IAT scores reflected a stronger pro-white implicit bias rated white faces as more trustworthy. Interestingly, the economic decision data did not show this pattern. The correlation between individual differences in IAT score and mean offers to black partners trended toward significance [$r(41) = -0.2772$, $P = 0.0720$], whereas that between IAT score and mean offers to white partners did not [$r(41) = -0.0932$, $P = 0.5522$]. Individuals whose IAT scores reflected a stronger pro-white implicit bias were less trusting of black partners when money was involved. That different behavioral components may drive the relationship between trust and implicit race attitude in the two experiments is interesting and suggests that the underlying factors contributing to trust evaluations in these situations may differ. Our study was not designed to address this question, however. Further research should establish that this same pattern is seen within subjects and that it is reliably obtained.

Analysis of White vs. Other-Race and Black vs. Other-Race Trust Disparity. To examine the relationship between IAT score and disparities in trustworthiness estimations of other-race faces and partners, we calculated similar disparity metrics as we did for our main analysis. For both ratings and offers, white/other-race disparity is defined as [Mean(white) − Mean(other race)]/SD(all) and black/other-race disparity is defined as [Mean(black) − Mean (other race)]/SD(all). In study 1, the correlations between IAT score and both white/other-race [$r(49) = 0.25$, $P = 0.08$] and black/other-race rating disparity [$r(49) = -0.26$, $P = 0.07$] were marginally significant and opposite in sign. In study 2, white/other-race offer disparity [$r(41) = 0.33$, $P < 0.05$] and subsequent rating disparity [$r(38) = 0.48$, $P < 0.01$] were both significantly correlated with IAT score; however, black/other-race offer disparity [$r(41) = -0.19$, $P = 0.23$] and subsequent rating disparity [$r(38) = 0.05$, $P = 0.76$] were not. These findings suggest that the relationship between black/white implicit race bias and trust disparity may generalize to biases with respect to other racial groups.

Analysis of Intersubject Agreement on Ratings and Offers. Previous studies that have collected trustworthiness ratings and economic decisions from participants using large sets of face stimuli have reported a high degree of intersubject agreement for individual faces (1, 2). We duplicated the analyses from those studies to compare the level of intersubject and interexperiment agreement in our data. First, for each face, we calculated the mean rating

across subjects within each of the three datasets (trustworthiness ratings from study 1, offers from study 2, and trustworthiness ratings from study 2). We then calculated the correlation between these mean ratings across the three datasets. Replicating previous findings, we found a high level of agreement between all three datasets. Mean trustworthiness ratings from study 1 correlated positively and significantly with both mean offers [Pearson's $r(289) = 0.82$, $P < 0.0001$] and mean ratings [Pearson's $r(289) = 0.86$, $P < 0.0001$] from study 2. Mean offers and ratings from study 2 were also positively and significantly correlated [Pearson's $r(289) = 0.80$, $P < 0.0001$]. We also calculated the correlation between individual participants' responses to each face in the Trust Game and the ratings portions of study 2. The resulting distribution of individual correlations was significantly different from zero [mean $r(289) = 0.23$, $P < 0.0001$].

Explicit Association Indices. In addition to the standard explicit measures of race attitude that we used, participants completed a series of explicit association ratings of our own design (e.g., how strongly do you associate black/white Americans with approach/avoid) that were combined to create positive and negative explicit association indices (EAIs). When these EAIs were included in the stepwise regression analyses, IAT score remained a significant predictor of race disparity in trustworthiness ratings but not in offers (Table S2). It is unclear how to interpret these results, however, because the relationship between participants' IAT score and their EAI in studies 1 and 2 varied. Further research investigating the relationship between these EAIs and race-related implicit associations must be conducted to establish their validity and reliability.

1. Engell AD, Haxby JV, Todorov A (2007) Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *J Cogn Neurosci* 19:1508–1519.

2. van't Wout M, Sanfey AG (2008) Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition* 108:796–803.

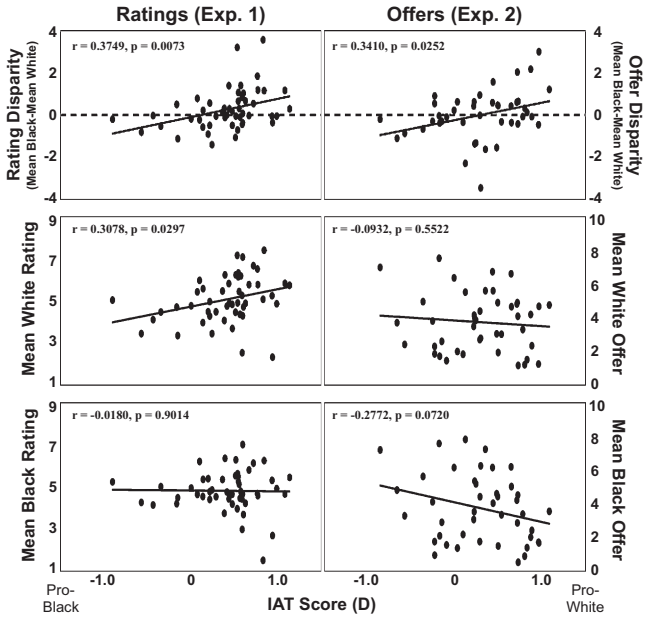

Fig. S1. Separate components of rating and offer disparity correlate with implicit race attitude (IAT score). The six panels depict the correlation between IAT scores and raw response data, including trust disparities [Mean(black response) − Mean(white response), *Top*; Mean(white response), *Middle*; and Mean(black response), *Bottom*] from experiments 1 (*Left*) and 2 (*Right*). In both experiments, participants' IAT scores were significantly correlated with trust disparity, confirming our main finding in the absence of *z*-score normalization. In study 1, this effect was driven by the trustworthiness ratings of white faces (*Middle Left*), with little to no contribution from ratings of black faces (*Lower Left*). Those with a stronger pro-white implicit bias rated whites as more trustworthy. In study 2, the trending correlation between offers to black partners may be indicative of their influence over the correlation between IAT score and trust disparity (*Lower Right*); the stronger an individual's pro-white implicit bias, the less they offered black partners. In contrast, offers to white partners were not correlated with IAT score (*Middle Right*).

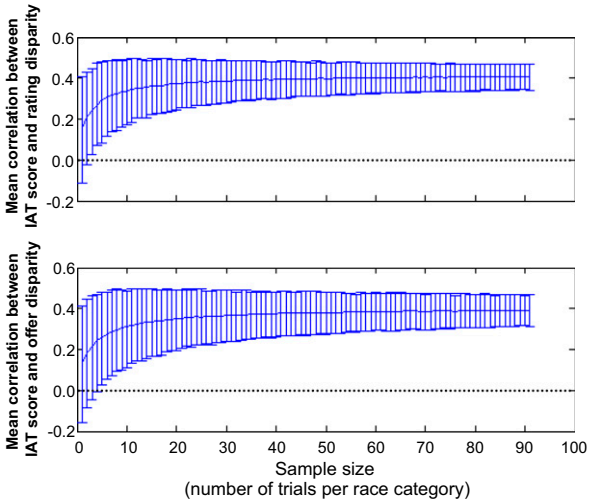

Fig. S2. Bootstrap analysis. Graphs of mean sample correlation between implicit race attitude (IAT score) and rating (*Upper*) or offer (*Lower*) disparity at each sample size. Error bars represent bootstrapped 95% confidence intervals.

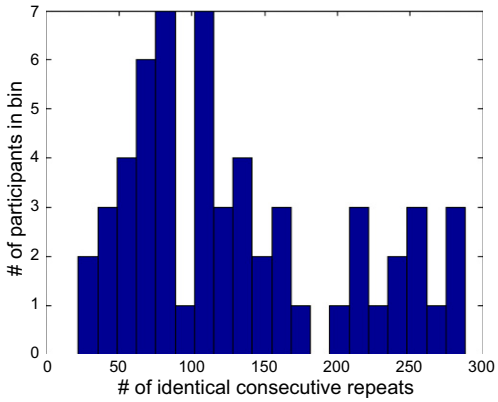

Fig. S3. Frequency distribution of total identical consecutive offers across participants in study 2. (*Left*) Notice how the large subgroup of data is normally distributed. Using a maximum likelihood procedure, the data were determined best fit by two Gaussian distributions when separating participants with total identical consecutive offers above 194 from the rest.

Table S1. Group statistics for implicit and explicit measures

	Mean	Median	SD	Minimum	Maximum
Experiment 1, *n* = 48					
IAT*	0.41	0.51	0.41	−2	2
IMS	36.48	37.50	6.77	5	45
EMS	20.38	19.50	10.50	4	45
MRS	−8.25	−9.50	4.55	−14	14
SRS	16.19	15.50	4.09	8	31
LIB/CON*	−2.05	−3.00	2.16	−5	5
Positive EAI	0.21	0.20	1.23	−14	14
Negative EAI	−0.16	0.00	1.24	−14	14
Experiment 2 (included), *n* = 43					
IAT	0.29	0.29	0.48		
IMS	36.16	37.00	6.00		
EMS	20.58	18.00	11.23		
MRS	−8.09	−8.00	3.99		
SRS	15.70	16.00	4.18		
LIB/CON	−2.48	−3.00	1.78		
Positive EAI	0.14	0.00	1.59		
Negative EAI	0.05	0.00	0.23		
Experiment 2 (all), *n* = 57					
IAT	0.31	0.36	0.44		
IMS	36.30	38.00	5.78		
EMS	20.44	18.00	11.07		
MRS	−8.02	−8.00	4.00		
SRS	15.82	16.00	3.86		
LIB/CON	−2.13	−3.00	1.94		
Positive EAI	0.13	0.00	1.40		
Negative EAI	0.12	0.00	1.16		

Statistics for the subgroup of included participants in study 2 are tabulated separately. Minimum and Maximum refer to the absolute minimum and maximum values each measure can have. *n = 50 for IAT and LIB/CON measures in study 1. EMS, External Motivation to Avoid Prejudice Survey; IMS, Internal Motivation to Avoid Prejudice Survey; LIB/CON, political leaning scale (Liberal/Conservative); MRS, Modern Racism Scale; SRS, Symbolic Racism Scale.

Table S2. Stepwise regression analyses for ratings/offers disparity with EAIs included

Experiment: Trustworthiness ratings, n = 48			
Dependent variable: Trust disparity			
Independent predictors: IAT, EMS, IMS, MRS, SRS, LIB/CON, positive EAI, negative EAI			
Final model: r^2 = 0.452, $P < 0.001$			
Factors	Standardized β	Significance in final model (P)	Change in r^2
Positive EAI	0.369	0.003	0.262
LIB/CON	0.348	0.004	0.107
IAT	0.295	0.013	0.083
Experiment: Modified Trust Game, n = 43			
Dependent variable: Offer disparity			
Independent predictors: IAT, EMS, IMS, MRS, SRS, LIB/CON, positive EAI, negative EAI			
Final model: r^2 = 0.246, P = 0.001			
Factors	Standardized β	Significance in final model (P)	Change in r^2
Negative EAI	−0.496	0.001	0.246

Separate stepwise regression analyses (probability of F to enter, P = 0.05; probability of F to remove, P = 0.10) for disparity in ratings (*Upper*) and offers (*Lower*) found that IAT scores independently accounted for a significant portion of the variance in rating disparity but not offer disparity. Political leaning and the positive EAI also remained in the final model as predictors of rating disparity. The negative EAI was the only factor to remain in the final model as a predictor of offer disparity. In addition to the overall predictive power of the EAI, it is interesting to note the association of trustworthiness ratings with the positive EAI and economic offers with the negative EAI. This could be indicative of a different mental focus induced in participants during each task; however, this conclusion is beyond the scope of the current study. EMS, External Motivation to Avoid Prejudice Survey; IMS, Internal Motivation to Avoid Prejudice Survey; LIB/CON, political leaning scale(Liberal/Conservative); MRS, Modern Racism Scale; SRS, Symbolic Racism Scale.

Other Supporting Information Files

SI Appendix A