Supplementary Materials

Revising Mental Representations of Faces Based on New Diagnostic Information

Samuel A. W. Klein, Ryan J. Hutchings, & Andrew R. Todd

University of California, Davis

Public repository of data, materials, and analysis code: https://osf.io/7u6cd/

Table of Contents

Linear Mixed-Effects Models (LMEMs)

Fixed-Effects Tables

Table S1

Trait Impressions of Robert by Fixed Effects of Time, Time 1 Induction, Time 2 Information, and All Interactions (Image-Generation Experiment)

Fixed Effects	*b*	*SE*	*t*	*df*
Intercept	3.64***	0.13	28.16	7.85
Time	0.18**	0.04	4.00	8.18
Time 1 Induction	-1.25***	0.18	-7.10	6.91
Time 2 Information	-0.23***	0.06	-3.90	35.07
Time × Time 1 Induction	-0.46***	0.03	-18.15	280.97
Time × Time 2 Information	0.20***	0.03	7.77	280.97
Time 1 Induction × Time 2 Information	0.34**	0.08	4.21	13.28
Time × Time 1 Induction × Time 2 Information	-0.37***	0.03	-14.36	280.97

Note. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$; degrees of freedom (*df*) were calculated using Satterthwaite approximation.

Table S2

Trait Ratings of Group Classification Images by Fixed Effects of Time, Time 1 Induction, Time 2 Information, and All Interactions (Image-Assessment Experiment 1)

Fixed Effects	*b*	*SE*	*t*	*df*
Intercept	3.98***	0.26	15.37	6.17
Time	0.02	0.01	1.23	8358.00
Time 1 Induction	-0.34***	0.04	-8.87	154.00
Time 2 Information	-0.09***	0.01	-5.93	8358.00
Time × Time 1 Induction	-.015***	0.01	-10.20	8358.00
Time × Time 2 Information	0.05***	0.01	3.64	8358.00
Time 1 Induction × Time 2 Information	0.11***	0.01	7.17	8358.00
Time × Time 1 Induction × Time 2 Information	-0.11***	0.01	-7.23	8358.00

Note. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$; degrees of freedom (*df*) were calculated using Satterthwaite approximation.

Table S3

Trait Ratings of Subgroup Classification Images by Fixed Effects of Time, Time 1 Induction, Time 2 Information, and All Interactions (Image-Assessment Experiment 2A)

Fixed Effects	*b*	*SE*	*t*	*df*
Intercept	3.97***	0.08	49.43	153.49
Time	-0.01	0.03	-0.21	45.04
Time 1 Induction	-0.55***	0.05	-10.61	81.29
Time 2 Information	-0.10*	0.04	-2.31	45.95
Time × Time 1 Induction	-0.29***	0.04	-8.12	58.67
Time × Time 2 Information	0.07*	0.03	2.23	44.26
Time 1 Induction × Time 2 Information	0.15***	0.04	3.41	48.40
Time × Time 1 Induction × Time 2 Information	-0.11**	0.03	-3.27	45.27

Note. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$; degrees of freedom (*df*) were calculated using Satterthwaite approximation.

Table S4

Trait Ratings of Individual Classification Images by Fixed Effects of Time, Time 1 Induction, Time 2 Information, and All Interactions (Image-Assessment Experiment 2B)

Fixed Effects	*b*	*SE*	*t*	*df*
Intercept	3.74***	0.05	74.76	315.60
Time	0.02	0.01	1.18	314.38
Time 1 Induction	-0.22***	0.02	-10.19	395.34
Time 2 Information	-0.04	0.02	-1.97	281.41
Time × Time 1 Induction	-0.13***	0.01	-8.78	314.36
Time × Time 2 Information	0.02	0.01	1.17	278.53
Time 1 Induction × Time 2 Information	0.07***	0.02	3.65	291.52
Time × Time 1 Induction × Time 2 Information	-0.03*	0.01	-2.43	280.36

Note. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$; degrees of freedom (*df*) were calculated using Satterthwaite approximation.

Random-Effects Structures

Image-Generation Experiment

In this and all other linear mixed-effects models (LMEMs) reported below, we aimed to maintain the largest possible random-effects structure without convergence failure or singularity. The two sources of variance accounted for in the image-generation experiment were participants (i.e., image generators) and traits (i.e., image generators' trait impressions of Robert). We reverse-scored the negatively-valenced traits (*mean*, *dominant*, and *aggressive*), ensuring that the traits can be interpreted as a sample drawn from a population of positive traits. The maximal random-effects structure included intercepts for participants and traits, a by-participant slope for Time, and by-trait slopes for Time, Time 1 induction, Time 2 information, and all possible interactions among those variables. Time is the only within-participant variable. Time, Time 1 induction, and Time 2 information are all within-trait variables.

This maximal model failed to converge. Thus, we downsized the random-effects structure until the model converged. The higher-order random-effects interactions involving traits proved most problematic, so we removed the three-way interaction, along with the Time × Time 1 and Time × Time 2 interactions. The final random-effects structure included intercepts for participants and traits, a by-participant slope for Time, and by-trait slopes for Time, Time 1 induction, Time 2 information, and the Time 1 induction × Time 2 information interaction.

Image-Assessment Experiment 1

The two sources of variance accounted for in image-assessment Experiment 1 were participants (i.e., image raters) and traits (i.e., image raters' trait impressions of the group classification images of Robert). As before, we reverse-scored the negative-valenced traits. Because each image rater rated all 8 group images on each of the 7 traits, Time, Time 1

induction, and Time 2 information are within-participant and within-trait variables. Therefore, the maximal random-effects structure includes intercepts for participants and traits, and both by-participant and by-trait slopes for Time, Time 1 induction, Time 2 information, as well as all possible interactions involving those variables.

This maximal model failed to converge without singular fit. Therefore, we downsized the random-effects structure until singular fit was eliminated. All by-trait slopes riddled the model with singular fit. So too did all by-participant slopes, except a slope for Time 1 induction. Therefore, the final random-effects structure included intercepts for participants and traits, a by-participant slope for Time, and by-trait slopes for Time, Time 1 induction, Time 2 information, and the Time 1 induction × Time 2 information interaction.

Image-Assessment Experiment 2A

The two sources of variance accounted for in image-assessment Experiment 2A were participants (i.e., image raters) and stimuli. Here, "stimuli" refers to the subgroup of image generators whose data were used to generate subgroup classification images at Time 1 and Time 2. With 96 subgroup images, there were 48 subgroup stimuli IDs, with each ID corresponding to the pair of Time 1 and Time 2 subgroup images generated by the same subgroup of image generators. Because each image rater rated all 96 subgroup images, Time, Time 1 induction, and Time 2 information were all within-participant variables. Because each subgroup image ID contained a Time 1 and Time 2 image, Time was also a within-stimuli variable. Therefore, the maximal random-effects structure included intercepts for participants and stimuli, by-participant slopes for Time, Time 1 induction, and Time 2 information, and all possible interactions involving those variables, as well as a by-stimuli slope for Time. This maximal model converged without any concerns over singular fit. Therefore, we reported the maximal model.

Image-Assessment Experiment 2B

The two sources of variance accounted for in image-assessment Experiment 2B were participants (i.e., image raters) and stimuli. Here, "stimuli" refers to the image generators of the individual classification images. These image generator IDs correspond to the 285 pairs of images (Time 1 and Time 2) created by each image generator. Because each image rater was randomly assigned images to rate from all possible conditions, Time, Time 1 induction, and Time 2 information were within-participant variables. With two time points of image generations per image generator, Time was a within-stimuli variable. Therefore, the maximal random-effects structure included intercepts for participants and stimuli, by-participant slopes for Time, Time 1 induction, and Time 2 information, and all possible interactions involving those variables, as well as a by-stimuli slope for Time.

This maximal model failed to converge without singular fit. Therefore, we downsized the random-effects structure until singular fit was no longer a concern. After removing the by-participant three-way interaction slope, singular fit was eliminated. Therefore, the final random-effects structure included intercepts for participants and traits, by-participant slopes for Time, Time 1 induction, Time 2 information, and all two-way interactions between the variables, as well as a by-stimuli slope for Time.

Analyses of Variance (ANOVAs)

Based on recommendations received during the editorial process, we reported LMEMs that account for additional sources of variance (traits in both the image-generation experiment and image-assessment Experiment 1; stimuli in both image-assessment Experiments 2A and 2B) in the main text. Here, we report analyses of variance (ANOVAs) that do not account for those additional sources of variance.

Image-Generation Experiment

Data Reduction

To limit the dimensionality of our data, we used the *psych* 1.8.12 package (Revelle, 2018) to conduct an exploratory factor analysis (EFA) with a *promax* rotation. We arrived at a two-factor solution, $\chi^2(570) = 6.69$, $p < .001$, accounting for 63% of the total variance (Factor 1 = 44%; Factor 2 = 27%; Fig. S1). Although a three-factor solution also fit the data, the two-factor solution made more theoretical sense. Similar to other two-factor solutions in the face-impression literature (e.g., Oosterhof & Todorov, 2008), four traits loaded onto a *positivity* factor (attractive, caring, intelligent, and trustworthy), and three traits loaded onto a *dominance* factor (aggressive, dominant, and mean). Each item loaded onto its primary factor at $|\lambda| > .58$, with all but *mean* loading at $|\lambda| > .65$. All items loaded onto the other factor at $|\lambda| < .40$. Composite indices were generated for each factor with its primary loadings; higher scores reflect greater positivity ($\alpha = .90$) and dominance ($\alpha = .83$), respectively. Although the positivity and dominance indices were highly correlated ($r = -.77$, $p < .001$), the EFA suggests they should be treated as two distinct variables.

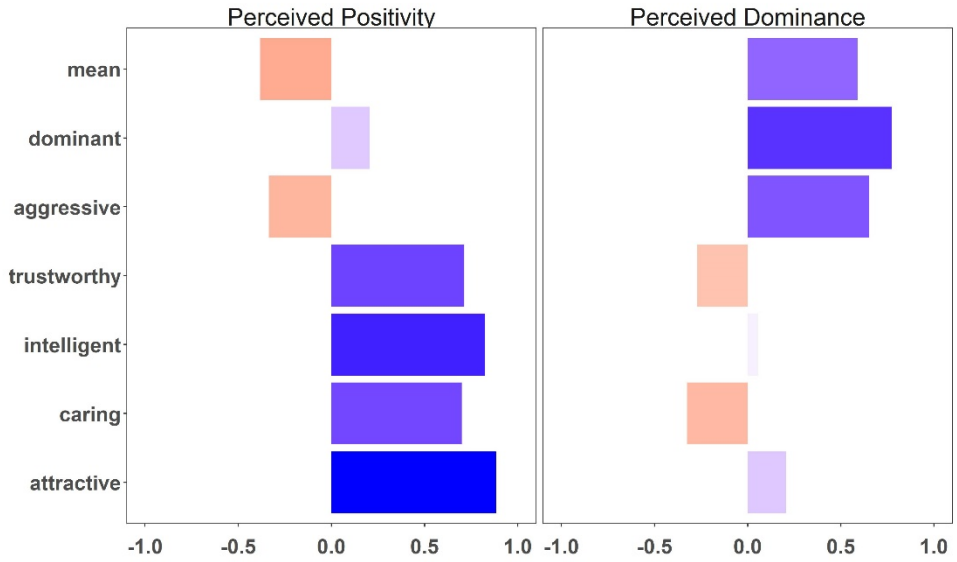

Figure S1. Results from the EFA on the trait ratings of Robert in the image-generation experiment. The *x*-axis depicts the positive loading strength for items on each factor. Factors are separated by panel and labeled by panel heading. Horizontal bars range from blue to red, with the greater appearance of blue representing higher positive load strength.

Positivity

A 2 (Time: 1 vs. 2; within-participants) × 2 (Time 1 induction: positive vs. negative; between-participants) × 2 (Time 2 information: control vs. counter-attitudinal; between-participants) mixed ANOVA on the positivity of the group images revealed a significant three-way interaction, $F(1, 560) = 73.20$, $p < .001$, $\eta_p^2 = .12$ (Fig. S2). We decomposed this interaction by conducting separate Time × Time 2 information ANOVAs in the positive-induction and negative-induction conditions.

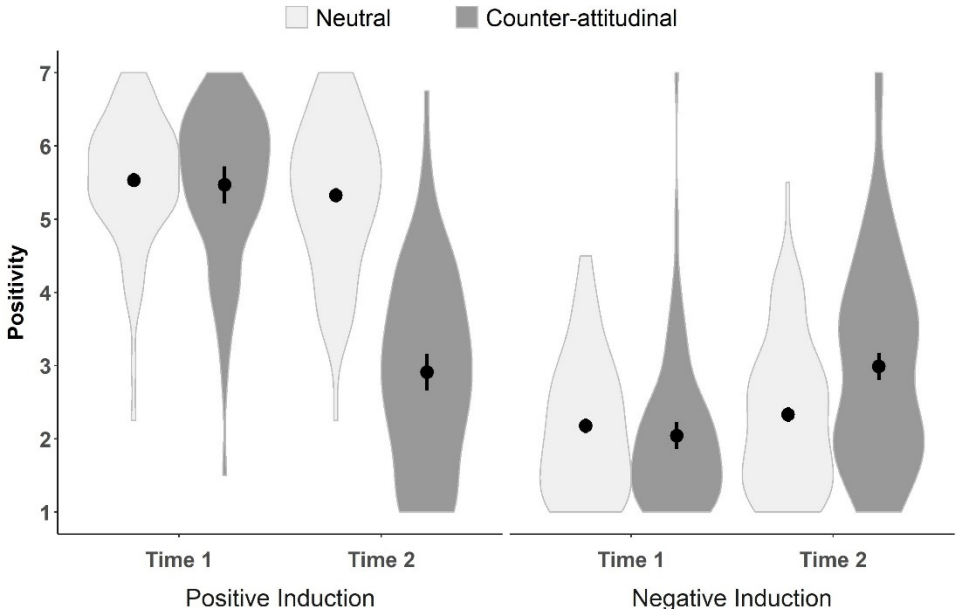

Figure S2. Positivity impressions of Robert by Time, Time 1 induction, and Time 2 information in the image-generation experiment. Error bars represent 95% confidence intervals. The surrounding violin plots are mirrored density distributions of the composite indices for the positivity factor (i.e., composite scores of trustworthy, intelligent, caring, and attractive) after a smoothing function was applied.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 280) = 83.77$, $p < .001$, $\eta_p^2 = .24$. Simple-effects tests indicated that learning about Robert's child molestation conviction prompted negative revision (Time 1: $M = 5.47$, $SD = 1.07$; Time 2: $M = 2.91$, $SD = 1.30$), $t(72) = 14.37$, $p < .001$, $d = 1.68$. Learning neutral information also prompted negative revision (Time 1: $M = 5.53$, $SD = 0.90$; Time 2: $M = 5.33$, $SD = 1.04$), $t(69) = 3.04$, $p = .003$, $d = 0.36$; however, this effect was smaller than that in the counter-attitudinal condition.

Evidence of revision also emerged in the negative-induction condition—Time × Time 2 information interaction, $F(1, 278) = 9.35$, $p = .002$, $\eta_p^2 = .03$. Learning about Robert's kidney donation prompted positive revision (Time 1: $M = 2.04$, $SD = 1.04$; Time 2: $M = 2.99$, $SD = 1.37$), $t(68) = -7.22$, $p < .001$, $d = -0.87$. Learning neutral information also prompted positive

revision (Time 1: $M = 2.18$, $SD = 0.93$; Time 2: $M = 2.33$, $SD = 1.02$), $t(72) = -2.16$, $p = .034$, $d = -0.25$; however, this effect was smaller than that in the counter-attitudinal condition.

Dominance

An identical 2 (Time) × 2 (Time 1 condition) × 2 (Time 2 condition) mixed ANOVA on the dominance of the group images also revealed a significant three-way interaction, $F(1, 560) = 46.38$, $p < .001$, $\eta_p^2 = .08$ (Fig. S3). We again decomposed this interaction by conducting separate 2 (Time) × 2 (Time 2 information) ANOVAs in the positive-induction and negative-induction conditions.

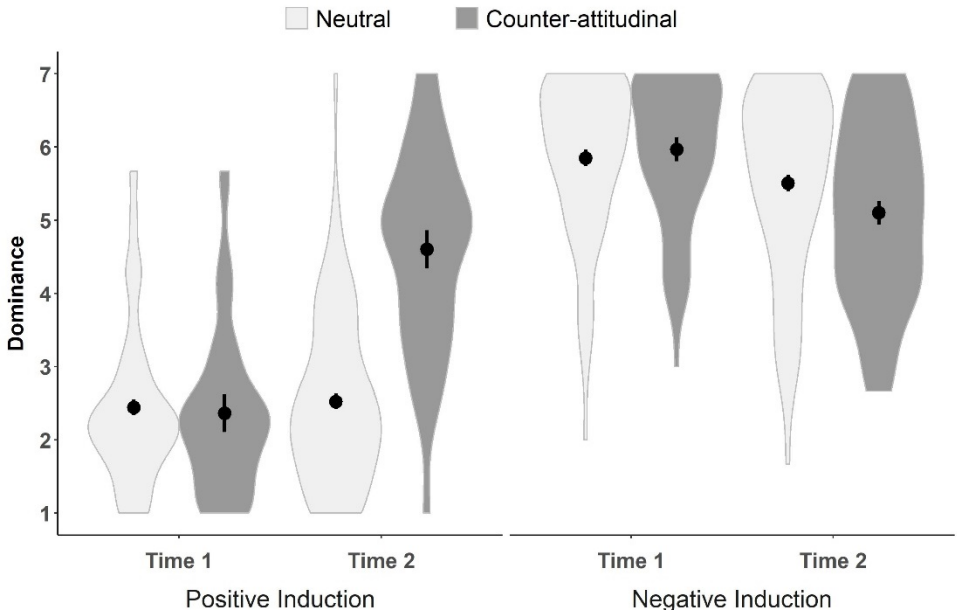

Figure S3. Dominance impressions of Robert images by Time, Time 1 induction, and Time 2 information in the image-generation experiment. Error bars represent 95% confidence intervals. The surrounding violin plots are mirrored density distributions of the composite indices for the dominance factor (i.e., composite scores of aggressive, dominant, and mean) after a smoothing function was applied.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 280) = 60.27$, $p < .001$, $\eta_p^2 = .18$. Learning about Robert's child molestation conviction prompted revision in the more dominant direction (Time 1: $M = 2.37$, $SD = 1.10$; Time 2: $M = 4.60$, $SD = 1.35$), $t(154) = -12.72$, $p < .001$, $d = -1.43$, whereas learning

neutral information did not (Time 1: $M = 2.44$, $SD = 1.09$; Time 2: $M = 2.52$, $SD = 1.18$), $t(69) = -1.07$, $p = .287$, $d = -0.13$.

Evidence of revision also emerged in the negative-induction condition; however, it was not restricted to the counter-attitudinal condition—Time × Time 2 information interaction, $F(1, 278) = 3.78$, $p = .053$, $\eta_p^2 = .03$. Learning about Robert's kidney donation prompted revision in the less dominant direction (Time 1: $M = 5.97$, $SD = 0.99$; Time 2: $M = 5.10$, $SD = 1.22$), $t(68) = 7.59$, $p < .001$, $d = 0.91$, as did learning neutral information (Time 1: $M = 5.85$, $SD = 1.11$; Time 2: $M = 5.51$, $SD = 1.28$), $t(72) = 4.32$, $p < .001$, $d = 0.51$.

Image-Assessment Experiment 1

Data Reduction

An EFA with *promax* rotation again revealed a two-factor solution, $\chi^2(1240) = 4.69$, $p = .086$, that accounted for 56% of the total variance (Factor 1 = 34%, Factor 2 = 22%; Fig. S4). Although a three-factor solution also fit the data, the two-factor solution made more theoretical sense. As before, four traits loaded onto a *positivity* factor (attractive, caring, intelligent, and trustworthy), and three traits loaded onto a *dominance* factor (aggressive, dominant, and mean). Each item loaded onto its primary factor at $\lambda > .60$ and the other factor at $\lambda < .15$. Composite indices were generated for each factor, with higher scores reflecting greater positivity ($\alpha = .85$) and dominance ($\alpha = .76$), respectively. Although the positivity and dominance indices were weakly correlated ($r = .09$, $p = .002$), the EFA suggested they should be treated separately.

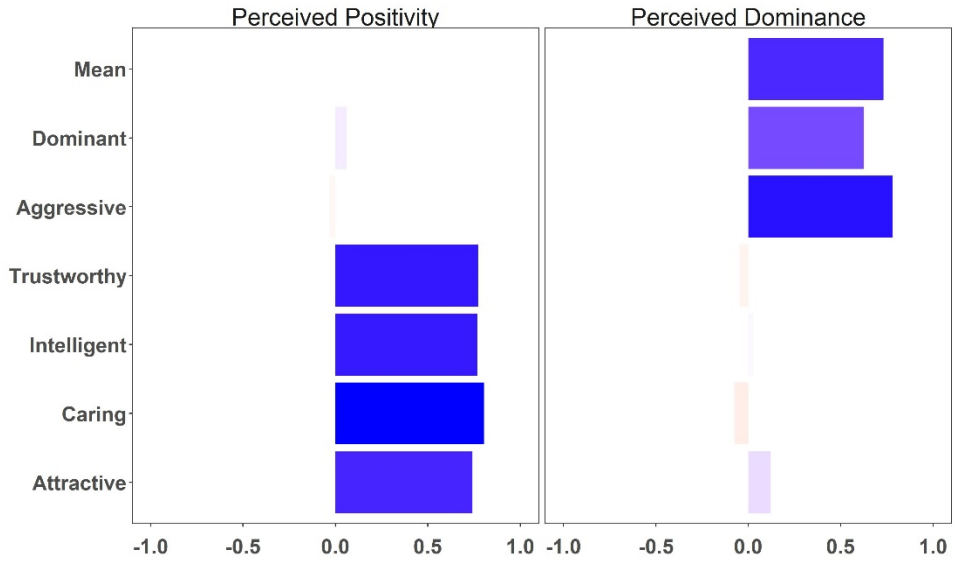

Figure S4. Results from the EFA on the trait ratings of group images in image-assessment Experiment 1. The *x*-axis depicts the positive loading strength for items on each factor. Factors are separated by panel and labeled by panel heading. Horizontal bars range from blue to red, with the greater appearance of blue representing higher positive load strength.

Positivity

A 2 (Time: 1 vs. 2) × 2 (Time 1 induction: positive vs. negative) × 2 (Time 2 information: control vs. counter-attitudinal) repeated-measures ANOVA on the positivity of the group images revealed a three-way interaction, $F(1, 154) = 58.62$, $p < .001$, $\eta_p^2 = .28$ (Fig. S5). We decomposed this interaction by conducting separate Time × Time 2 information ANOVAs in the positive-induction and negative-induction conditions.

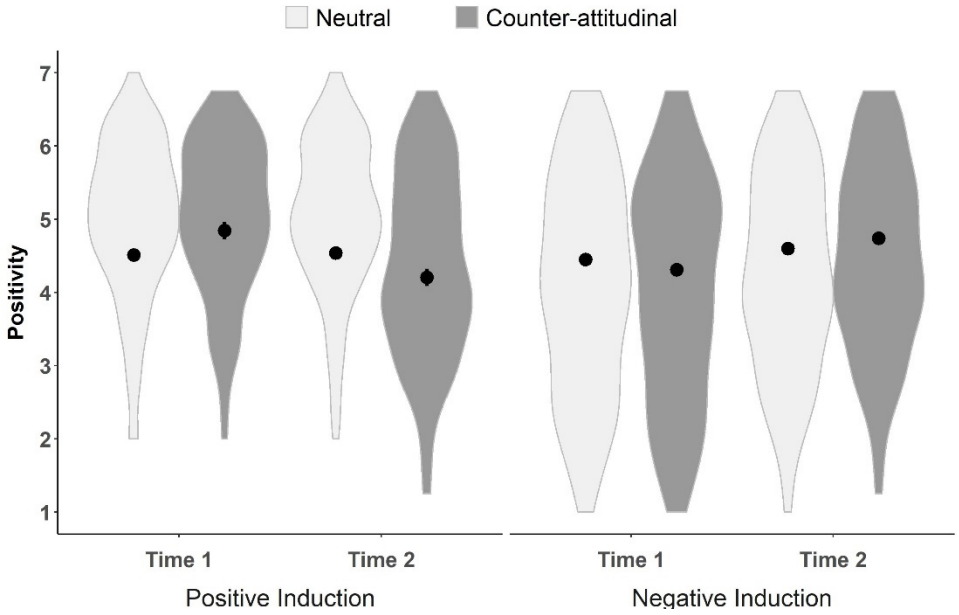

Figure S5. Positivity impressions of group classification images by Time, Time 1 induction, and Time 2 information in image-assessment Experiment 1. Error bars represent 95% confidence intervals. The surrounding violin plots are mirrored density distributions of the composite indices for the positivity factor (i.e., composite scores of trustworthy, intelligent, caring, and attractive) after a smoothing function was applied.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 154) = 48.96$, $p < .001$, $\eta_p^2 = .24$. Simple-effects tests indicated that learning about Robert's child molestation conviction prompted negative revision (Time 1: $M = 5.00$, $SD = 1.04$; Time 2: $M = 4.36$, $SD = 1.25$), $t(154) = 7.58$, $p < .001$, $d = 0.61$, whereas learning neutral information did not (Time 1: $M = 5.00$, $SD = 1.07$; Time 2: $M = 5.02$, $SD = 1.04$), $t(154) = -0.61$, $p = .546$, $d = -0.05$.

Evidence of revision also emerged in the negative-induction condition—Time × Time 2 information interaction, $F(1, 154) = 13.39$, $p < .001$, $\eta_p^2 = .08$. Learning about Robert's kidney donation prompted positive revision (Time 1: $M = 4.03$, $SD = 1.49$; Time 2: $M = 4.46$, $SD = 1.27$), $t(154) = -6.47$, $p < .001$, $d = -0.52$. Learning neutral information also prompted positive revision (Time 1: $M = 4.10$, $SD = 1.46$; Time 2: $M = 4.25$, $SD = 1.33$), $t(154) = -2.76$, $p = .007$, $d = -0.22$; however, this effect was smaller than that in the counter-attitudinal condition.

Dominance

A 2 (Time) × 2 (Time 1 condition) × 2 (Time 2 condition) repeated-measures ANOVA on the dominance of the group images also revealed a significant three-way interaction, $F(1, 154) = 24.55$, $p < .001$, $\eta_p^2 = .14$ (Fig. S6). We again decomposed this interaction by conducting separate 2 (Time) × 2 (Time 2 information) ANOVAs in the positive-induction and negative-induction conditions.

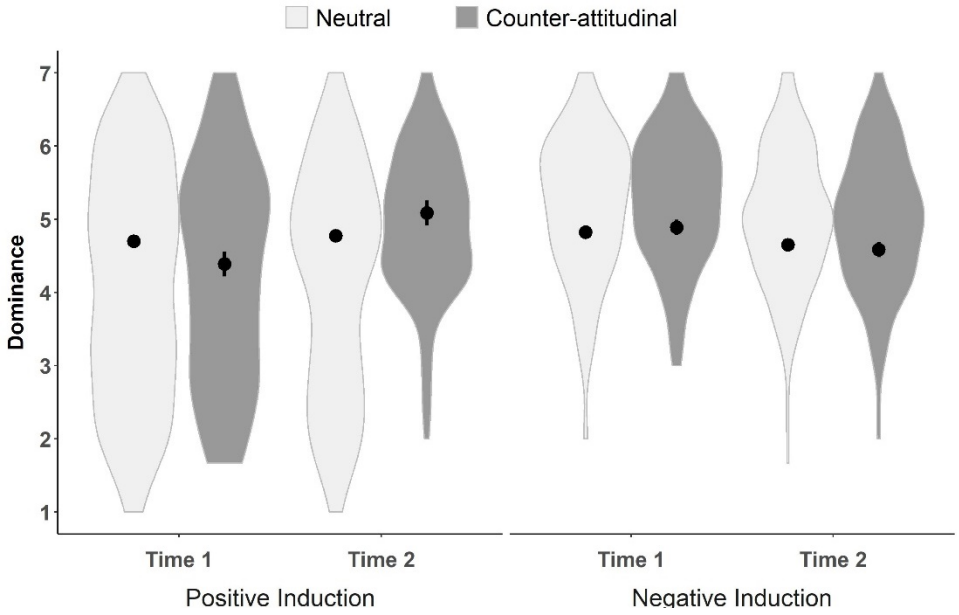

Figure S6. Dominance impressions of group classification images by Time, Time 1 induction, and Time 2 information in image-assessment Experiment 1. Error bars represent 95% confidence intervals. The surrounding violin plots are mirrored density distributions of the composite indices for the dominance factor (i.e., composite scores of aggressive, dominant, and mean) after a smoothing function was applied.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 154) = 22.58$, $p < .001$, $\eta_p^2 = .13$. Learning about Robert's child molestation conviction prompted revision in the more dominant direction (Time 1: $M = 4.25$, $SD = 1.46$; Time 2: $M = 4.95$, $SD = 0.96$), $t(154) = -5.79$, $p < .001$, $d = -0.47$, whereas learning neutral information did not (Time 1: $M = 4.10$, $SD = 1.54$; Time 2: $M = 4.17$, $SD = 1.50$), $t(154) = -1.38$, $p = .169$, $d = -0.11$.

Evidence of revision also emerged in the negative-induction condition; however, it was not restricted to the counter-attitudinal condition—Time × Time 2 information interaction, $F(1, 154) = 2.53$, $p = .114$, $\eta_p^2 = .02$. Learning about Robert's kidney donation prompted revision in the less dominant direction (Time 1: $M = 5.26$, $SD = 0.87$; Time 2: $M = 4.96$, $SD = 0.92$), $t(154) = 4.03$, $p < .001$, $d = 0.32$, as did learning neutral information (Time 1: $M = 5.19$, $SD = 0.97$; Time 2: $M = 5.02$, $SD = 0.94$), $t(154) = 2.71$, $p = .008$, $d = 0.22$.

Image-Assessment Experiment 2A

A 2 (Time) × 2 (Time 1 induction) × 2 (Time 2 information) repeated-measures ANOVA on the trustworthiness impressions of the subgroup images revealed the expected three-way interaction, $F(1, 113) = 129.20$, $p < .001$, $\eta_p^2 = .53$ (Fig. S7).[1] We decomposed this interaction by conducting separate 2 (Time) × 2 (Time 2 information) ANOVAs in the positive-induction and negative-induction conditions.

[1]An unexpected difference between Time 2 information conditions at Time 1 emerged in the positive-induction condition in Experiment 2A, $t(113) = -4.73$, $p < .001$, $d = -0.44$, and Experiment 2B, $t(241) = -5.75$, $p < .001$, $d = -0.37$. This difference did not emerge in the negative-induction condition in Experiment 2A, $t(113) = 0.96$, $p = .338$, $d = 0.09$, or Experiment 2B, $t(241) = 0.88$, $p = .382$, $d = 0.06$; nor did it emerge in either Time 1 induction condition in Experiment 1.

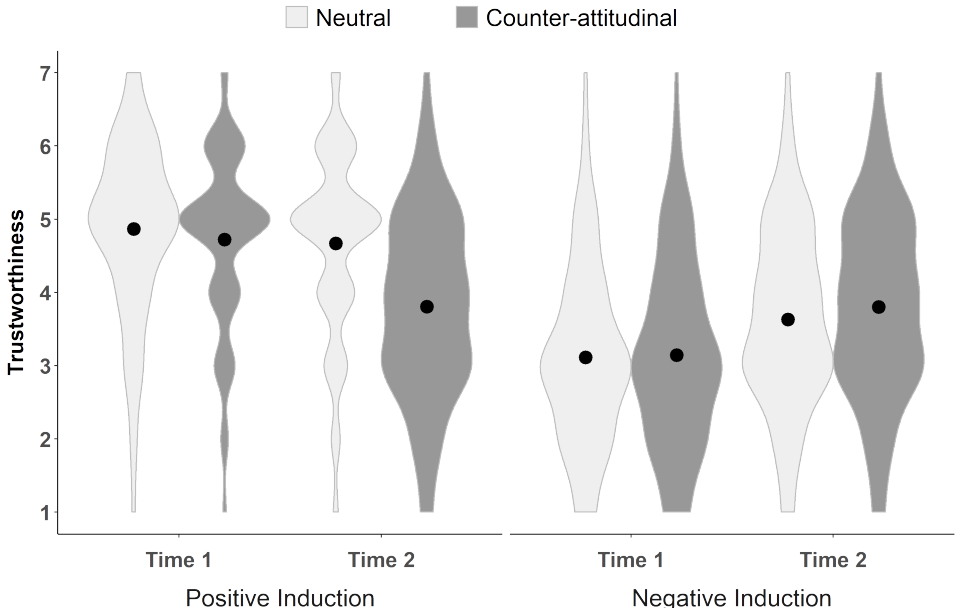

Figure S7. Trustworthiness impressions of subgroup classification images by Time, Time 1 induction, and Time 2 information in image-assessment Experiment 2A. Error bars represent 95% confidence intervals. The surrounding violin plots are mirrored density distributions of image raters' responses after a smoothing function was applied.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 113) = 158.64$, $p < .001$, $\eta_p^2 = .58$. Learning about Robert's child molestation conviction prompted negative revision (see Table S7 for descriptive statistics in Experiments 2A and 2B), $t(113) = 15.94$, $p < .001$, $d = 1.49$. Although learning neutral information about Robert also prompted negative revision, $t(113) = 6.02$, $p < .001$, $d = 0.56$, this effect was smaller than that in the counter-attitudinal condition.

Revision was also evident in the negative-induction condition—Time × Time 2 information interaction, $F(1, 113) = 9.40$, $p = .003$, $\eta_p^2 = .08$. Learning about Robert's kidney donation prompted positive revision, $t(113) = -14.20$, $p < .001$, $d = -1.33$. Learning neutral information also prompted positive revision, $t(113) = -13.31$, $p < .001$, $d = -1.25$, but again this effect was smaller than that in the counter-attitudinal condition.

Image-Assessment Experiment 2B

A 2 (Time) × 2 (Time 1 induction) × 2 (Time 2 information) repeated-measures ANOVA on the trustworthiness impressions of individual images revealed the expected three-way interaction, $F(1, 241) = 60.00$, $p < .001$, $\eta_p^2 = .19$ (Fig. S8). We again decomposed this interaction by examining the underlying patterns separately in the positive-induction and negative-induction conditions.

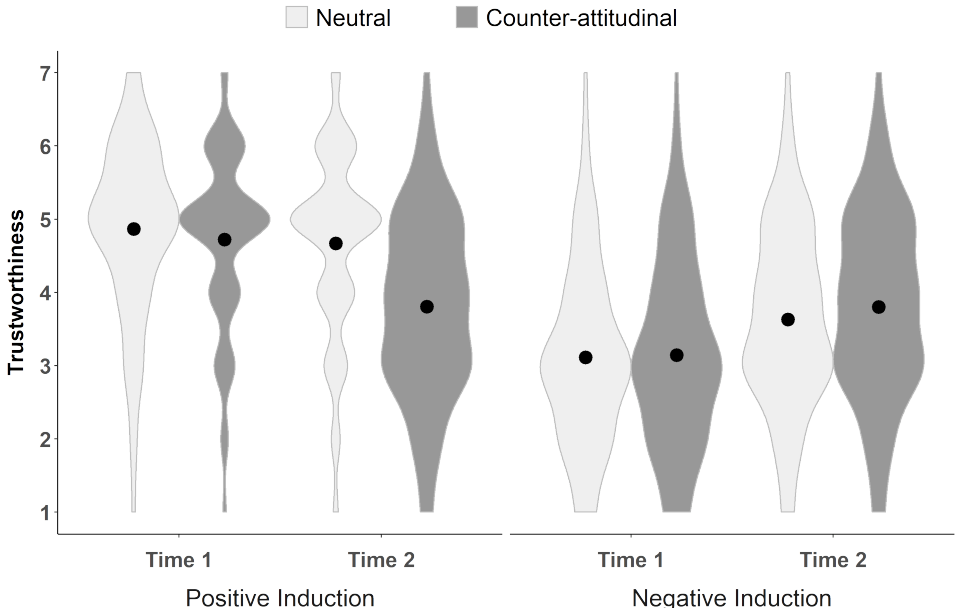

Figure S8. Trustworthiness impressions of individual classification images by Time, Time 1 induction, and Time 2 information in image-assessment Experiment 2B. Error bars represent 95% confidence intervals. The surrounding violin plots are mirrored density distributions of image raters' responses after a smoothing function was applied.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 241) = 56.72$, $p < .001$, $\eta_p^2 = .19$. Learning about Robert's child molestation conviction prompted negative revision, $t(241) = 16.17$, $p < .001$, $d = 1.04$. Learning neutral information also prompted negative revision, $t(241) = 9.31$, $p < .001$, $d = 0.60$; however, this effect was smaller than that in the counter-attitudinal condition.

Revision was also evident in the negative-induction condition—Time × Time 2 information interaction, $F(1, 241) = 11.68$, $p < .001$, $\eta_p^2 = .05$. Learning about Robert's kidney donation prompted positive revision, $t(241) = -12.55$, $p < .001$, $d = -0.81$. Learning neutral information also prompted positive revision, $t(241) = -8.75$, $p < .001$, $d = -0.56$, but again this effect was smaller than that in the counter-attitudinal condition.

Descriptive Statistics Tables

Table S5

Descriptive Statistics for Trait Ratings of Robert (Image-Generation Experiment)

	Positive-induction Condition			
	Neutral		Counter-attitudinal	
Trait	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)
Aggressive	1.93 (1.47)	2.21 (1.42)	1.74 (1.11)	4.84 (1.68)
Attractive	4.27 (1.46)	4.10 (1.58)	4.42 (1.59)	2.64 (1.66)
Caring	6.39 (1.09)	6.17 (1.26)	6.30 (1.14)	2.95 (1.50)
Dominant	3.83 (1.51)	3.43 (1.56)	3.53 (1.60)	4.63 (1.74)
Intelligent	5.49 (1.13)	5.31 (1.20)	5.26 (1.35)	3.85 (1.66)
Mean	1.57 (1.04)	1.93 (1.24)	1.82 (1.38)	4.34 (1.91)
Trustworthy	5.99 (1.19)	5.71 (1.34)	5.89 (1.37)	2.21 (1.56)

	Negative-induction Condition			
	Neutral		Counter-attitudinal	
Trait	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)
Aggressive	6.03 (1.38)	5.74 (1.55)	6.28 (1.21)	5.33 (1.56)
Attractive	2.14 (1.28)	2.34 (1.45)	2.26 (1.54)	2.75 (1.71)
Caring	1.67 (1.08)	1.85 (1.05)	1.51 (1.13)	3.26 (1.63)
Dominant	5.37 (1.66)	5.04 (1.80)	5.39 (1.61)	5.09 (1.56)
Intelligent	3.18 (1.41)	3.18 (1.47)	2.84 (1.45)	3.35 (1.60)
Mean	6.15 (1.43)	5.74 (1.61)	6.23 (1.15)	4.88 (1.71)
Trustworthy	1.73 (1.15)	1.96 (1.26)	1.57 (1.22)	2.59 (1.68)

Table S6

Descriptive Statistics for Trait Ratings of Group Classification Images (Image-Assessment Experiment 1)

Positive-induction Condition

	Neutral		Counter-attitudinal	
Trait	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)
Aggressive	3.97 (1.81)	4.09 (1.77)	4.18 (1.78)	4.97 (1.21)
Attractive	4.62 (1.52)	4.68 (1.57)	4.86 (1.47)	4.38 (1.62)
Caring	5.32 (1.27)	5.28 (1.38)	5.19 (1.31)	4.29 (1.53)
Dominant	4.23 (1.56)	4.33 (1.64)	4.40 (1.64)	4.90 (1.33)
Intelligent	5.07 (1.36)	5.06 (1.25)	4.97 (1.31)	4.50 (1.39)
Mean	4.09 (1.90)	4.10 (1.83)	4.17 (1.73)	4.98 (1.36)
Trustworthy	4.97 (1.33)	5.07 (1.30)	4.95 (1.26)	4.25 (1.54)

Negative-induction Condition

	Neutral		Counter-attitudinal	
Trait	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)
Aggressive	5.29 (1.21)	5.14 (1.19)	5.30 (1.17)	4.90 (1.25)
Attractive	4.02 (1.76)	4.26 (1.65)	3.84 (1.71)	4.34 (1.53)
Caring	4.10 (1.86)	4.14 (1.67)	3.93 (1.81)	4.47 (1.62)
Dominant	5.06 (1.35)	5.00 (1.32)	5.28 (1.27)	5.05 (1.34)
Intelligent	4.25 (1.54)	4.41 (1.44)	4.28 (1.54)	4.66 (1.37)
Mean	5.21 (1.27)	4.91 (1.34)	5.21 (1.24)	4.93 (1.30)
Trustworthy	4.02 (1.68)	4.19 (1.62)	4.06 (1.69)	4.36 (1.65)

Table S7

Descriptive Statistics for Trustworthiness Impressions of Subgroup Classification Images (Image-Assessment 2A) and Individual Classification Images (Image-Assessment Experiment 2B)

	Time 2 Information			
	Neutral		Counter-attitudinal	
Time 1 Induction	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)
Experiment 2A				
Positive	4.87 (0.92)	4.67 (0.95)	4.72 (0.88)	3.80 (0.75)
Negative	3.11 (0.85)	3.63 (0.79)	3.15 (0.81)	3.80 (0.78)
Experiment 2B				
Positive	4.16 (0.81)	3.98 (0.76)	4.05 (0.76)	3.65 (0.76)
Negative	3.41 (0.78)	3.58 (0.75)	3.42 (0.79)	3.69 (0.77)

Additional Dependent Measures

Feeling Thermometer

Alongside the seven traits on which participants in the image-generation experiment rated Robert (discussed in the main text), they also indicated their feelings about him by entering a number between 0 and 100. Higher numbers reflect warmer impressions.

A mixed ANOVA on the feeling thermometer revealed a three-way interaction, $F(1, 278) = 149.31$, $p < .001$, $\eta_p^2 = .35$.[2] We decomposed this interaction by conducting separate 2 (Time) × 2 (Time 2 information) mixed ANOVAs in the positive-induction and negative-induction conditions.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $F(1, 139) = 126.50$, $p < .001$, $\eta_p^2 = .48$. Learning about Robert's child molestation conviction prompted negative revision (Time 1: $M = 82.22$, $SD = 19.30$; Time 2: $M = 27.53$, $SD = 27.57$), $t(71) = 14.46$, $p < .001$, $d = 1.70$. Learning neutral information also prompted negative revision (Time 1: $M = 82.38$, $SD = 21.68$; Time 2: $M = 76.71$, $SD = 20.64$), $t(68) = 2.62$, $p = .011$, $d = 0.32$; however, this effect was smaller than that in the counter-attitudinal condition.

Evidence of revision also emerged in the negative-induction condition—Time × Time 2 information interaction, $F(1, 139) = 25.00$, $p < .001$, $\eta_p^2 = .15$. Learning about Robert's kidney donation prompted positive revision (Time 1: $M = 18.54$, $SD = 22.59$; Time 2: $M = 33.84$, $SD = 25.74$), $t(67) = -6.66$, $p < .001$, $d = -0.81$, but learning neutral information did not (Time 1: $M = 22.78$, $SD = 19.13$; Time 2: $M = 24.96$, $SD = 20.01$), $t(72) = -1.60$, $p = .113$, $d = -0.19$.

[2]Either due to technical errors in data collection or blank responses on the feeling thermometer measure, we excluded three participants' data, leaving a final sample of 282 participants for these analyses.

Race/Ethnicity Categorizations

In the image-generation experiment, participants also made race/ethnicity categorizations of Robert at Time 1 and Time 2. They selected from six response options the one that best matched their impression of Robert's race/ethnicity: Asian, Black, Latinx, Native American, White, or other. 'Other' responses account for approximately 3% of total race/ethnicity categorizations. Given that all but 2 of those responses were irrelevant, ambiguous, or left uncategorized, 'other' responses were excluded from the analyses below.

We conducted generalized linear model (GLM) analyses with logit link functions on the race/ethnicity categorizations. Time (1 vs. 2), Time 1 induction (positive vs. negative), Time 2 information (neutral vs. counter-attitudinal), and all interactions were included as predictors. To account for repeated-measures (Time) in the model, we also included random intercepts for each participant. Due to low frequencies in Latinx and Native American categorizations (frequencies <10 for both), we do not include analyses for these racial/ethnic categorizations.

Neither Asian nor Black categorizations produced significant three-way interactions (*B*s < -2.00, *p*s > .136; Table S8).[3] White categorizations, however, revealed a significant three-way interaction, $B = 0.33$, $SE = 0.13$, $z = 2.46$, $p = .014$. We decomposed this interaction by conducting separate 2 (Time) × 2 (Time 2) analyses in the two induction conditions.

Evidence of revision emerged in the positive-induction condition—Time × Time 2 information interaction, $B = -0.38$, $SE = 0.17$, $z = -2.26$, $p = .024$. Whereas the odds of White categorization were not revised after learning about Robert's child molestation conviction, $OR = 0.74$, $p = .503$, they did decrease after learning neutral information, $OR = 3.37$, $p = .012$. No

[3]Either due to technical errors in data collection, blank responses, or other categorizations, we excluded 31 participants' data, leaving a final sample of 254 participants for these analyses.

significant evidence of revision emerged for White categorizations in the negative-induction condition—Time × Time 2 information interaction, $B = 0.21$, $SE = 0.21$, $z = 0.99$, $p = .321$.

Table S8

Descriptive Statistics (Proportions) for Race/Ethnicity Categorizations (Image-Generation Experiment)

Positive-induction Condition

	Neutral		Counter-attitudinal	
Race/Ethnicity	Time 1	Time 2	Time 1	Time 2
Asian	0.02	0.00	0.05	0.02
Black	0.38	0.60	0.49	0.48
Latinx	0.11	0.09	0.08	0.06
Native American	0.02	0.02	0.00	0.02
White	0.48	0.29	0.38	0.43

Negative-induction Condition

	Neutral		Counter-attitudinal	
Race/Ethnicity	Time 1	Time 2	Time 1	Time 2
Asian	0.05	0.00	0.00	0.02
Black	0.23	0.36	0.35	0.46
Latinx	0.11	0.07	0.14	0.09
Native American	0.03	0.00	0.02	0.02
White	0.57	0.57	0.49	0.42

Race/Ethnicity Prototypicality Ratings

In image-assessment Experiment 1, participants also rated the group images of Robert on how prototypical they were for each of four races/ethnicities: Asian, Black, Latinx, and White (0 = *not at all prototypical,* 100 = *extremely prototypical*).

We conducted a 2 (Time: 1 vs. 2) × 2 (Time 1 induction: positive vs. negative) × 2 (Time 2 information: neutral vs. counter-attitudinal) ANOVA on the prototypicality ratings separately for each race/ethnicity. These analyses revealed a three-way interaction only for ratings of Black

prototypicality, $F(1, 154) = 19.52$, $p < .001$, $\eta_p^2 = .11$. Ratings of Asian, Latinx, and White prototypicality revealed no significant three-way interactions (Fs < 3.43, ps > .07, η_p^2s < .02). We decomposed the interaction for Black prototypicality by inspecting the underlying patterns separately in the positive-induction and negative-induction conditions (see Table S9).

Revision was evident in the positive-induction condition—Time × Time 2 information interaction—for ratings of Black prototypicality. Visualizations of Robert were less prototypically Black after learning about his child molestation conviction, but not after learning neutral information. This pattern of revision did not emerge in the negative-induction condition.

Table S9

Descriptive Statistics and Univariate ANOVA Results for Race/Ethnicity Prototypicality Ratings of Group Images (Image-Assessment Experiment 1)

Positive-induction Condition

	Neutral			Counter-attitudinal			
Race/ethnicity	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	*d*	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	*d*	Time
Asian	51.91 (32.42)	50.65 (32.49)	0.08	51.85 (32.06)	49.15 (32.93)	0.15	3.61
Black	67.17 (26.87)	67.55 (27.54)	0.03	68.97 (25.06)	56.23 (30.30)	0.43	21.24***
Latinx	58.45 (25.37)	59.03 (27.28)	0.04	59.27 (25.77)	60.75 (26.96)	0.07	1.08
White	56.09 (29.61)	56.34 (29.53)	0.01	58.32 (29.00)	67.23 (23.70)	0.29	11.58***

Negative-induction Condition

	Neutral			Counter-attitudinal			
Race/ethnicity	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	*d*	Time 1 *M* (*SD*)	Time 2 *M* (*SD*)	*d*	Time
Asian	47.89 (33.99)	48.20 (33.85)	0.02	47.33 (33.79)	49.77 (33.70)	0.16	2.80
Black	59.84 (28.60)	53.75 (31.68)	0.27	60.01 (26.75)	54.05 (32.03)	0.25	12.75***
Latinx	57.90 (26.98)	58.66 (27.48)	0.04	59.95 (26.78)	57.48 (29.76)	0.13	0.67
White	65.53 (25.36)	71.13 (23.67)	0.25	63.10 (25.67)	70.49 (23.04)	0.32	19.27***

Note. Info = Information. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$.