SUPPLEMENTARY MATERIAL

Interracial contact shapes racial bias in the learning of person-knowledge

Margaret Welte[1], Tzipporah P. Dang[1], Shaina Rosenblum[1], Jennifer T. Kubota[1,2], and Jasmin Cloutier[1]

[1] Department of Psychological and Brain Sciences, University of Delaware, 105 The Green, Newark, DE 19716, USA

[2] Department of Political Science and International Relations, University of Delaware, 18 Amstel Ave., Newark, DE 19716, USA

Supplementary Discussion 1. Preregistration Clarifications.

The main text for the current paper focuses on the confirmatory predictions and tasks used to assess those confirmatory predictions. There were also exploratory tasks and individual differences measures assessed during data collection. All procedures, including tasks and measures, are outlined in the online preregistration (https://osf.io/4k8ve/) published on December 21, 2020. This section serves to summarize the full study design and to clarify details of data analyses.

1.1. Study Design. This section details the full experimental procedure, including the timeline of participation and all data collected. This online study consisted of five separate days of participation completed over six possible days. The extra day served to promote participant retention by allowing participants to have one excused absence if necessary. On the first day directly before starting training, participants completed a pre-training evaluative priming task (EPT) and independent face recognition task. Participants proceeded to their first training session and continued to complete one training session per day for four more days. Each training session included one encoding phase and one test phase. After their fifth and final training session, participants completed a post-training EPT, independent face recognition task, and individual difference measures. Supplementary Table 1 lists the tasks each participant completed on the five days of participation, including the full list of individual difference measures.

Supplementary Table 1. Outline of Study Progression and Tasks. Bolded tasks are analyzed in the main text.

Day	Tasks
1	1. Exploratory pre-training EPT **2. Exploratory pre-training independent face recognition task** **3. Confirmatory training session 1**
2	**1. Confirmatory training session 2**
3	**1. Confirmatory training session 3**
4	**1. Confirmatory training session 4**
5	**1. Confirmatory training session 5** 2. Exploratory post-training EPT

	3. Exploratory post-training independent face recognition task 4. Individual differences measures (random order except where noted) a. Exploratory likeability ratings (always 1st) b. Exploratory race IAT (always 2nd) c. Demographics (always 3rd) **d. Confirmatory interracial contact (sequential order)** **i. Childhood contact** **ii. Current contact** e. Exploratory EMS/IMS f. Exploratory Modern Racism Scale (MRS) g. Exploratory Symbolic Racism Scale (SRS) h. Exploratory feeling thermometer (random order) i. White people ii. Black people i. Exploratory Perceived Stress Scale j. Exploratory fear of the Coronavirus

1.2. Preregistered Hypotheses. The preregistration contains all confirmatory and exploratory hypotheses for the full study design included in Supplementary Table 1. This section clarifies the hypotheses that are included in the current paper and the preregistered analyses used to test them.

This paper includes only the confirmatory hypotheses from the preregistration and one exploratory analysis of the independent face recognition task. To explicitly map the portions of the preregistration relevant to these confirmatory hypotheses onto the current manuscript, Supplementary Table 2 lists for each tested hypothesis its description both in the preregistration and in the manuscript. Also included are the preregistered analyses to test the hypothesis, any revisions to the preregistered analyses, and the location of the results. The same preregistered analyses were utilized to test all hypotheses, with one exception being that Confirmatory Hypothesis 3 refers specifically to face recognition rather than all three person memory tasks during training. Consequently, the details regarding the analyses and results are discussed with relation to all hypotheses in the table.

It should be noted that this table only discusses analyses for the person memory tasks during training (i.e., face recognition, valence attribution, and cued recall of person-knowledge) and does not include the independent face recognition task. The analysis of the independent face recognition task included a preregistered exploratory model for which we did not have a preregistered hypothesis. All statistical models run, including a confirmatory model for each person memory task during training and an exploratory model for the independent face recognition task, may be reviewed in the next section (1.3).

Supplementary Table 2. Summary of Preregistered Hypotheses and Related Analyses.

Preregistered Hypothesis	**Description of Hypothesis in Main Manuscript**	**Preregistered Analysis Plan**	**Revised Analysis Plan (if applicable)**	**Where to Find Results**
Preregistration p. 3: "Confirmatory Hypothesis 1: Perceivers will display different learning rates over time for differently valenced person-knowledge. Prediction 1: In line with the negativity bias, we will observe a main effect of statement valence such that there will be increased learning in the series of learned memory tasks (i.e., greater face recognition, valence attribution, and cued recall of person knowledge) for negative face-statement pairs compared to positive face-statement pairs."	Study Overview section of Methods in manuscript: "In line with the negativity bias…we predicted that perceivers would display increased learning across all three person memory indices for negative relative to positive face-statement pairs."	Preregistration p. 22-24 discusses 5 trial-level statistical models for the person memory tasks during training (i.e., face recognition, valence attribution, cued recall of person-knowledge), including contrast codes, Bonferroni correction, and random effects determination.	The transparent changes document uploaded to OSF clarified the confirmatory analysis plan for the person memory tasks during training, to include one aggregate-level model per task for a total of 3 confirmatory training models. The trial-level models specified in the preregistration are specified as exploratory. In this Supplementary Material document in the following section (Discussion 1.3), we clarify that the Bonferroni correction method must also be updated due to the use of aggregate-level models, which reduced the overall number of tests run on the training data from 5 to 3. Therefore, the adjusted alpha is .05/3=.017.	Omnibus results may be found in the Face Recognition, Valence Attribution, and Cued Recall of Person-Knowledge sections of the Results in the manuscript. Simple effects conducted to analyze significant interactions from the omnibus model may be found in this Supplementary Material document in Discussion 4. Descriptive statistics for the training tasks may be found in Discussion 5.
Preregistration p. 3: "Confirmatory Hypothesis 2: In line with the ORE, if same-race faces are more memorable, White perceivers will display increased learning for White targets. Prediction 2: We will observe a main effect of target race such that there will be increased learning in the series of learned memory tasks (i.e., greater face recognition, valence attribution, and cued recall of person knowledge) for White targets compared to Black targets."	Study Overview section of Methods: "For our confirmatory predictions, we hypothesized that self-identified White participants would display increased learning for perceived White compared to perceived Black faces. Importantly, we predicted that this race-based bias would not only occur for face recognition, as previously extensively documented by			

	literature on the CRD, but also for valence attribution and cued recall of person-knowledge.”			
Preregistration p. 3: “Confirmatory Hypothesis 3: Perceivers will display different learning rates over time depending on the presence of paired information. Prediction 3: We will observe a main effect of person-knowledge condition such that there will be greater face recognition in the series of learned memory tasks of faces paired with person-knowledge compared to faces that are only perceptually familiar.”	Study Overview section of Methods: “Furthermore, we expected perceivers to show a preferential learning rate in face recognition for faces paired with person-knowledge compared with faces that were only perceptually familiar.”			
Preregistration p. 3: “Confirmatory Hypothesis 4: If affectively incongruent information is more memorable, perceivers will display increased learning for negative-White and positive-Black targets. Prediction 4: Because of previous evidence showing greater individuation for violations of affective expectations…we expect a target race x statement valence interaction in which there is increased learning in the series of learned memory tasks (i.e., greater face recognition, valence attribution,	Study Overview section of Methods: “Specifically, based on findings of greater individuation for violations of affective expectations…we expected increased learning (i.e., greater face recognition, valence attribution, and cued recall of person knowledge) for face-statement pairs affectively incongruent with race-based prejudices (i.e., perceived White faces with negative person-knowledge and perceived Black faces			

and cued recall of person knowledge) for pairings that violate affective expectations (e.g., meek Black man) versus more expected affective pairings of race and person-knowledge (e.g., aggressive Black man).”	with positive person-knowledge).”			
Preregistration p. 3-4: “Confirmatory Hypothesis 5: If perceivers have greater lifetime interracial contact, they will display greater recognition and learning of outgroup face-statement pairs irrespective of valence compared to those with less lifetime interracial contact. Previous research on face processing and interracial contact predicts increasing face processing efficiency with increasing interracial contact…Additionally, people with greater interracial contact are particularly good at individuating given motivation to do so…Therefore, we predict a reduction in the preferential learning rate for ingroup faces for those with high compared to low lifetime interracial contact scores (extension of confirmatory hypothesis 2). In other words, those with greater lifetime interracial contact will show a smaller difference in learning rates in the series of learned memory tasks	Study Overview section of Methods: “Because interracial contact has been found to reduce cross-race recognition bias, to improve individuation, and to be associated with efficiency in face processing…we also predicted that perceivers with greater lifetime interracial contact would show reduced racial biases in face recognition, valence attribution, and cued recall of person-knowledge. As we anticipated a benefit of paired person-knowledge for all perceivers, we expected the impact of contact on racial bias in person memory to be similar across Information Type conditions. For example, perceivers with more interracial contact were expected to show similarly reduced differences in face recognition between perceived Black faces and			

(i.e., face recognition, valence attribution, and cued recall of person knowledge) between Black and White faces compared with those with lower lifetime interracial contact.”	perceived White faces regardless of whether faces were paired with person-knowledge at encoding.”			

1.3. Statistical Models. This section clarifies the specific models from the OSF preregistration that are reported in the main text as well as the Bonferroni correction method utilized to interpret the results.

Due to an oversight, the initial preregistration document (p. 22-23) listed only exploratory trial-level models for the training tasks (i.e., face recognition, valence attribution, cued recall of person-knowledge). A transparent changes document (https://osf.io/xj73d) was later added to the study OSF page on February 5, 2021 to report the mistakenly omitted confirmatory aggregate-level model to be run for each training task. This model structure was utilized for confirmatory analyses of the training tasks that are described in the main text of the current paper. Specifically, for our confirmatory analyses of the training tasks (see Supplementary Table 3), we utilized the fixed effects structure specified in the transparent changes document, but we still followed the random effects determination procedure described in the preregistration (p. 23-24) to model between-participants and slopes for all within-subject factors (i.e., random effects) and the correlations between these random effects. The exploratory trial-level training models from the initial preregistration have not yet been conducted.

The statistical model run on the independent face recognition task data (see Supplementary Table 3) is an exploratory model specified on p. 20-21 of the preregistration. However, we did not model random effects for between-stimulus variance as written in the preregistration. Rather, to maintain consistency with the training models and follow the random effects determination procedure described on p. 23-24 of the preregistration document, we modelled only between-participants variance in intercepts and slopes for all within-subject factors (i.e., random effects) and the correlations between these random effects.

In the main text, we used an alpha of .05 and additionally reported what remained significant after Bonferroni correcting for the training models (see Supplementary Table 3). The Bonferroni correction method for the training models (i.e., face recognition, valence attribution, cued recall of person-knowledge) described in the initial preregistration was based on the five exploratory trial-level models. We did not explicitly describe in the transparent changes document with the three confirmatory aggregate-level models that the Bonferroni corrected alpha would also need to be updated to reflect the changed number of comparisons. A Bonferroni corrected alpha of .05/3 = .017 was implemented to account for the three linear regression models run on the training data (i.e., face recognition, valence attribution, cued recall of person-knowledge). As reflected in the initial preregistration, there was no Bonferroni correction applied to independent face recognition task analysis.

Supplementary Table 3. Summary of Preregistered Statistical Models and Criteria for Statistical Significance.

Statistical Model	**Preregistration**	**Clarification to Preregistration**	**Significant Results (p<.05)**	**Significant after Bonferroni correction (p<.017)?**
Face Recognition Base model before random effects determination: model <- lmer(Accuracy.DV ~ Race_Contrast * Session_Contrast * Condition_Contrast * Contact + (1+ Race_Contrast * Session_Contrast * Condition_Contrast | subject), data = data)	Yes; confirmatory model with inclusion of interracial contact preregistered on p. 1 of transparent changes document (https://osf.io/xj73d)	The transparent changes document mistakenly omitted the details of determining the model's random effects structure that were included on p. 23-24 of the preregistration document (https://osf.io/4k8ve/). As a result, the model in the transparent changes document incorrectly suggests that we only	Main effect of Race: b=-0.117, SE=0.013, *95% confidence interval* ($CI_{95\%}$)=[-0.142, -0.093], t(68.000)=-9.256, p<.001	Yes
			Main effect of Session: b=0.062, SE=0.006, $CI_{95\%}$=[0.051, 0.073], t(68.000)=11.068, p<.001	Yes

		modelled between-participants variance in intercepts. However, using the fixed effects structure specified in the transparent changes document, we followed the random effects determination procedure described in the preregistration and modelled between-participants and slopes for all within-subject factors (i.e., random effects) and the correlations between these random effects.	Main effect of Information Type: b=0.126, SE=0.016, $CI_{95\%}$=[0.096, 0.157], t(68.000)=8.022, p<.001	Yes
			Interaction of Race x Session: b=0.024, SE=0.006, $CI_{95\%}$=[0.013, 0.035], t(68.000)=4.242, p<.001	Yes
			Interaction of Race x Information Type: b=0.042, SE=0.018, $CI_{95\%}$=[0.006, 0.077], t(68.000)=2.284, p=.026	No
			Interaction of Session x Information Type: b=-0.016, SE=0.005, $CI_{95\%}$=[-0.025, -0.006], t(68.000)=-3.245, p=.002	Yes
			Interaction of Information Type x Contact: b=-0.039, SE=0.016, $CI_{95\%}$=[-0.070, -0.008], t(68.000)=-2.443, p=.017	Equal to the Bonferroni correction with an adjusted alpha of 0.017
			Interaction of Race x Session x Information	Yes

			Type interaction: *b*=0.028, *SE*=0.008, $CI_{95\%}$=[0.012, 0.044], *t*(1608.000)=3.392, *p*<.001	
Valence Attribution Base model before random effects determination: model <- lmer(Accuracy.DV ~ Race_Contrast * Session_Contrast * Condition_Contrast * Contact + (1+ Race_Contrast * Session_Contrast * Condition_Contrast | subject), data = data)	Yes; confirmatory model with inclusion of interracial contact preregistered on p. 1 of transparent changes document (https://osf.io/xj73d)	The transparent changes document mistakenly omitted the details of determining the model's random effects structure that were included on p. 23-24 of the preregistration document (https://osf.io/4k8ve/). As a result, the model in the transparent changes document incorrectly suggests that we only modelled between-participants variance in intercepts. However, using the fixed effects structure specified in the transparent changes document, we followed the random effects determination procedure described in the preregistration and modelled between-participants and slopes for all within-subject factors (i.e., random effects) and the correlations between these random effects.	Main effect of Race: *b*=-.036, *SE*=.008, $CI_{95\%}$=[-0.052, -0.020], t(67.104)=-4.345, *p*<.001	Yes
			Main effect of Session: *b*=.040, *SE*=.004, $CI_{95\%}$=[0.032, 0.048], t(68.230)=9.520, *p*<.001	Yes
			Main effect of Information Type: *b*=-.042, *SE*=.018, $CI_{95\%}$=[-0.078, -0.006], t(67.941)=-2.264, *p*=.027	No
			Interaction of Race x Information Type: *b*=-.056, *SE*=.020, $CI_{95\%}$=[-0.096, -0.016], t(67.913)=-2.736, *p*=.008	Yes
			Interaction of Session x Information Type: *b*=-.021, *SE*=.009, $CI_{95\%}$=[-0.039, -	No

<table>
<tr><td></td><td></td><td></td><td>0.002], $t(67.913)=-2.192$, $p=.032$</td><td></td></tr>
<tr><td rowspan="5">Cued Recall of Person-Knowledge

Base model before random effects determination: model <- lmer(Accuracy.DV ~ Race_Contrast * Session_Contrast * Condition_Contrast * Contact + (1+ Race_Contrast * Session_Contrast * Condition_Contrast | subject), data = data)</td><td rowspan="5">Yes; confirmatory model with inclusion of interracial contact preregistered on p. 1 of transparent changes document (https://osf.io/xj73d)</td><td rowspan="5">The transparent changes document mistakenly omitted the details of determining the model's random effects structure that were included on p. 23-24 of the preregistration document (https://osf.io/4k8ve/). As a result, the model in the transparent changes document incorrectly suggests that we only modelled between-participants variance in intercepts. However, using the fixed effects structure specified in the transparent changes document, we followed the random effects determination procedure described in the preregistration and modelled between-participants variance in intercepts <u>and</u> slopes for all within-subject factors (i.e., random effects) and the correlations between these random effects.</td><td>Main effect of Race: $b=-0.059$, $SE=0.010$, $CI_{95\%}=[-0.078, -0.040]$, $t(68.175)=-6.095$, $p<.001$</td><td>Yes</td></tr>
<tr><td>Main effect of Session: $b=0.058$, $SE=0.007$, $CI_{95\%}=[0.044, 0.073]$, $t(67.921)=7.819$, $p<.001$</td><td>Yes</td></tr>
<tr><td>Interaction of Race x Session: $b=-0.011$, $SE=0.004$, $CI_{95\%}=[-0.017, -0.004]$, $t(69.446)=-3.000$, $p=.004$</td><td>Yes</td></tr>
<tr><td>Interaction of Race x Session x Contact: $b=0.010$, $SE=0.004$, $CI_{95\%}=[0.003, 0.016]$, $t(68.576)=2.700$, $p=.009$</td><td>Yes</td></tr>
<tr><td>Interaction of Race x Valence x Contact: $b=-0.033$, $SE=0.013$, $CI_{95\%}=[-0.059, -0.007]$, $t(68.847)=-2.454$, $p=0.017$</td><td>Equal to the Bonferroni correction with an adjusted alpha of 0.017</td></tr>
<tr><td>Independent Face Recognition</td><td>Yes; exploratory model with inclusion of interracial contact preregistered on p. 20-21 of</td><td>The model on p. 21 of the preregistration includes random effects for between-stimulus variance.</td><td>Main effect of Race: $b=-0.042$, $SE=0.009$, $CI_{95\%}=[-0.059, -$</td><td>N/A; no Bonferroni correction preregistered</td></tr>
</table>

Base model before random effects determination**:** FR.omnibus = lmer(Accuracy ~ Contrast.Race * Contrast.Session * Contrast.Familiarity * Contact + (1+ Contrast.Race * Contrast.Session * Contrast.Familiarity |subject), data = FR)	preregistration document (https://osf.io/4k8ve/)	However, to maintain consistency with the training models and follow the random effects determination procedure described on p. 23-24 of the preregistration document, we modelled only <u>between-participants</u> variance in intercepts and slopes for all within-subject factors (i.e., random effects) and the correlations between these random effects.	0.024], t(68.000)=-4.709, p<.001	
			Main effect of Session: b=-0.049, SE=0.017, $CI_{95\%}$=[-0.082, -0.015], t(68.000)=-2.846, p=.006	
			Main effect of Familiarity: b=-0.071, SE=0.020, $CI_{95\%}$=[-0.111, -0.031], t(68.000)=-3.472, p=< .001	
			Interaction of Session x Familiarity: b=-0.068, SE=0.030, $CI_{95\%}$=[-0.126, -0.010], t(68.000)=-2.291, p=.025	
			Interaction of Session x Contact: b=0.039, SE=0.017, $CI_{95\%}$=[0.005, 0.073], t(68.000)=2.265, p=.027	
			Interaction of Race x Session x Familiarity: b=-0.157, SE=0.049, $CI_{95\%}$=[-0.253, -0.061], t(68.000)=-3.204, p=.002	

Due to the scope of the current paper, we have not yet conducted any exploratory analyses other than the independent face recognition task model described above and reported in the main text. For transparency, the preregistered models that have not been analyzed are as follows:

1. Exploratory evaluative priming task models (preregistration p. 19)
2. Exploratory trial-level training (face recognition, valence attribution, cued recall of person-knowledge) models (p. 22, 23)
3. Exploratory training foil faces model (p. 23)

1.4. Data Exclusion. This section describes one correction and one clarification to the data exclusion criteria as defined in the preregistration.

The preregistered analysis plan lists the reaction time criteria used to exclude participant-specific trials for all tasks in the Data Exclusion section (preregistration p. 26). However, it was incorrectly reported that a unique set of reaction time criteria (p. 22) would be used to exclude participant-specific trials for the independent face recognition task. The two preregistered sets of exclusion criteria (p. 22, 26) utilize the same rules but vary in the exact reaction time cutoff (e.g., 3.5 SD instead of 3SD; 75 ms instead of 100ms). With no functional difference in the criteria, we utilized the trial-level reaction time exclusion criteria included in the Data Exclusion section (p. 26) for all analyses reported in the main text. Accordingly, we eliminated trials based on reaction time in the following order: 3SD away from the participant trial mean, 100ms or less, and 3SD away from the group mean.

The preregistered exclusion criteria (p. 26) did not specify how the participant mean reaction time and the group mean reaction time would be calculated. To account for the potential effects of learning, both the participant mean and group mean were calculated for each session of a task. In other words, trials in a task were only excluded if a participant's reaction time was 3SD away from their mean reaction time or the group mean reaction time for that session. For the number of trials excluded by task and session, refer to Supplementary Discussion 3.

Supplementary Discussion 2. Development of Stimuli.

2.1. Development of Facial Stimuli

To equate the faces to be encoded throughout training, we first curated a set of about 750 pictures of perceived Black and White male faces from various online face databases (Burton et al., 2010; DeBruine & Jones, 2017; Eberhardt et al., 2006; Kennedy et al., 2009; Ma et al., 2015; Mende-Siedlecki et al., 2020; Nordstrøm et al., 2004; Strohminger et al., 2016; Thomaz & Giraldi, 2010; Tottenham et al., 2009). All stimuli displayed direct eye gaze. Stimuli were cropped around the hair and from the neck up, placed on white backgrounds, and grey-scaled using Adobe Photoshop or Gimp. Images were also processed using the SHINE toolbox (Willenbockel et al., 2010) to remove low-level visual variation.

To obtain ratings for the modified photos, we collected data from human subjects recruited through Amazon Mechanical Turk (AMT). AMT participants were eligible to participate if they were at least 18 years old, located within the United States, and approved for at least 70% of the online studies they completed previously. Eligible participants provided informed consent via a Qualtrics survey, completed a thirty-minute rating task, and were compensated $2 for their participation. To ensure participants were paying attention throughout the rating task, we included 19 catch trials (~5% of total trials) composed of faces with a superimposed red number on the forehead, and participants were required to enter that number as their response. We eliminated participants who either (i) did not pass at least 75% of catch criteria (14/19 catch trials) or (ii) responded with the same key for all non-catch trials. We also excluded participant-specific trials based on reaction time (RT) in the following order: (i) 3SD away from the participant's individual mean RT, (ii) less than 100ms, and (iii) 3SD away from the group mean RT. Compliance with data exclusion criteria was monitored throughout data collection to ensure that each face had at least 30 ratings per dimension. Our final sample included 740 AMT participants.

For the rating task, the initial set of around 750 faces was split into two groups, each containing approximately equal numbers of perceived Black and White male faces. Each participant rated one of these two groups of faces (i.e., approximately 375 faces) on a single dimension to prevent fatigue and boredom. Participants provided ratings for all faces in a group on one of the following nine dimensions: distinctiveness ("How distinctive is this person's face?", from 1 (*not distinctive*) to 7 (*very distinctive*)); likability ("How likable does this person look?", from 1 (*not likable*) to 7 (*very likable*)); attractiveness ("How attractive is this person?" from 1 (*not attractive*) to 7 (*very attractive*)); trustworthiness ("How trustworthy does this person look?", from 1 (*not trustworthy*) to 7 (*very trustworthy*)); dominance ("How dominant does this person look?", from 1 (*not dominant*) to 7 (*very dominant*)); picture quality ("Rate the picture quality" from 1 (*poor quality*) to 7 (*high quality*)); perceived age ("How old does this person appear?", categorical, with 1 (*18-25*), 2 (*26-35*), 3 (*36-45*), 4 (*46-55*), 5 (*56-65*), 6 (*66-75*), 7 (*76+ years*)); perceived race ("What is his race?" categorical, with 1 (*Black/African American*), 2 (*Asian/Asian American*), 3 (*White/Euro-American*), 4 (*Latino/Hispanic American*), 5 (*Middle Eastern/Arab American*), 6 (*Native American*), 7 (*Other*)); and emotional expression ("What emotion is he expressing?", categorical with 1 (*angry*), 2 (*happy*), 3 (*neutral*), 4 (*sad*)).

Once we collected at least 30 data points per rating dimension for each face, we used the average ratings to eliminate faces before equating. For facial expression, we eliminated any faces that did not have at least 70% of raters identifying the face as having either a happy or neutral expression. In the final equated faces, the distribution of happy and neutral facial expressions did not differ by perceived race. For picture quality, we eliminated any faces that had a mean rating of less than 3. For perceived race, we eliminated any faces that did not have at least 80% consensus as being perceived to be either Black or White. After exclusions based on ratings, we ended up with 141 faces (78 perceived Black, 63 perceived White) to equate.

From these 141 faces, we equated four groups of 16 faces each (8 perceived Black, 8 perceived White faces) on all nine dimensions. The 8 perceived Black faces and 8 perceived White faces in each group did not significantly differ from (i) the other faces in the group within their perceived race nor (ii) the group as a whole on any of the nine measured dimensions (see Supplementary Table 4 below for summary statistics). Specifically, faces were equated to have a high rater consensus for facial expression and perceived race (>70% and >80% consensus respectively). Faces in a group were equated to be average in their ratings for the remaining seven dimensions.

Supplementary Table 4. Summary statistics for equating 4 groups of training faces.

		***F*-value**	***DF***	***p*-value**
Age	**Group**	0.100	3, 56	0.960
	Race	0.306	1, 56	0.583
	Group x Race	0.238	3, 56	0.870
Attractive	**Group**	0.481	3, 56	0.697
	Race	0.787	1, 56	0.379
	Group x Race	0.904	3, 56	0.445
Distinct	**Group**	0.332	3, 56	0.802
	Race	0.079	1, 56	0.780

	Group x Race	0.457	3, 56	0.714
Dominant	**Group**	1.137	3, 56	0.342
	Race	3.464	1, 56	0.068
	Group x Race	0.116	3, 56	0.950
Expression (rater consensus)	**Group**	0.735	3, 56	0.535
	Race	0.037	1, 56	0.847
	Group x Race	0.773	3, 56	0.514
Like	**Group**	0.299	3, 56	0.826
	Race	1.990	1, 56	0.164
	Group x Race	0.050	3, 56	0.985
Quality	**Group**	1.325	3, 56	0.275
	Race	0.526	1, 56	0.471
	Group x Race	0.446	3, 56	0.721
Race (rater consensus)	**Group**	0.722	3, 56	0.543
	Race	2.347	1, 56	0.131
	Group x Race	0.693	3, 56	0.560
Trust	**Group**	0.491	3, 56	0.690
	Race	0.968	1, 56	0.329
	Group x Race	0.266	3, 56	0.850

Note. Each face had at least 30 ratings for all nine dimensions. Differences in ratings on these dimensions were not statistically significant.

After equating, we had a final total of 64 faces split into four equated stimuli groups each consisting of 8 perceived Black and 8 perceived White male faces. Eight different orderings of these four stimuli groups were created using random selection without replacement to determine which group would be used for each of the following information type conditions during training: perceptually familiar only faces, faces paired with positive person-knowledge, and faces paired with negative person-knowledge. The leftover group in an ordering was used for novel faces in the post-training evaluative priming task (see Supplementary Discussion 1.1). The rated faces that were previously excluded throughout the equating process were utilized as novel distractor faces for the test phases of training. Novel distractor faces were equated only on perceived race (i.e., at least 80% rater consensus that the face was perceived to be Black or White) due to the large number of novel faces needed for five training sessions. Each novel distractor face was presented only once throughout training.

2.2. Development of Person-Knowledge Statements

To convey valenced person-knowledge to be encoded throughout training, we additionally developed a set of 48 positive and 48 negative statements from a database of sentences with available pilot data (Mende-Siedlecki & Havlicek, in prep). From this database, we identified a subset of statements that referred to individual male actors (e.g., "he") and described commonplace behaviors or characteristics. The database included ratings of gender stereotypicality [0 (*very stereotypically male*) to 8 (*very stereotypically female*)], race stereotypicality [0 (*very stereotypically Black*) to 8 (*very stereotypically White*)], valence [-4 (*very negative*) to +4 (*very positive*)], and arousal [0 (*not at all arousing*) to 8 (*extremely arousing*)] for each statement. Using these ratings, we eliminated statements that were perceived to be (i) more stereotypically female (i.e., average rating greater than 4 for gender stereotypicality) or (ii) relevant to racial stereotypes (i.e., average rating greater than 6 or less than 2 for race stereotypicality). We divided the remaining sentences based on valence by sorting average ratings between 0 and 1.5 into a positive statement group and average ratings between -1.5 and 0 into a negative group.

We ran Welch's two-sample t-tests to equate the remaining statements such that absolute valence (i.e., valence intensity), arousal, and race stereotypicality did not significantly differ by valence condition. Specifically, statements were equated to be average for valence intensity and arousal but low for race stereotypicality. The final equated group of 96 statements (48 positive, 48 negative) did not differ on absolute value of valence by valence condition, $M_{StereotypicalityPositiveStatements}$=2.121, $M_{StereotypicalityNegativeStatements}$=2.256, $t(29)$=-1.844, p=.161, $CI_{95\%}$=[-0.057, 0.326]; arousal by valence condition $M_{StereotypicalityPositiveStatements}$=3.510, $M_{StereotypicalityNegativeStatements}$=3.739, $t(18)$=1.489, p=.153, $CI_{95\%}$=[-0.094, 0.555]; or race stereotypicality by valence condition, $M_{StereotypicalityPositiveStatements}$=4.517, $M_{StereotypicalityNegativeStatements}$=4.155, $t(30)$=-1.844, p=.075, $CI_{95\%}$=[-0.762, 0.039].

After equating, statements were randomly divided into groups of three similarly-valenced statements. This resulted in 16 unique groups of three positive statements and 16 unique groups of three negative statements. The valenced statement groups were randomly paired with each of the 8 face orderings. As a result, the four equated face groups in an ordering were assigned to one of the following four conditions: faces paired with positive person-knowledge, faces paired with negative person-knowledge, perceptually familiar faces with no paired person-knowledge, and leftover faces used for the post-training evaluative priming task (see Supplementary Discussion 1.1). The faces within the groups assigned to the two person-knowledge conditions were randomly paired with the corresponding valenced person-knowledge statements. For example, each of the 16 faces in the group within an ordering assigned to the positive person-knowledge condition were randomly paired with one of the 16 groups consisting of three positive statements. Participants were randomly assigned to one of the 8 orderings of the four equated face groups such that the equated face group assigned to a condition remained constant within subjects across the study. In other words, participants assigned to an ordering encoded and were subsequently tested on the same faces paired with the same statements (or no statements for perceptually familiar only faces) in each training session.

Supplementary Discussion 3: Data Exclusion

This section reports the number of trials excluded for each task by session. Participant-specific trials were excluded based on reaction time according to the data exclusion criteria specified in the online preregistration and Supplementary Discussion 1.3. Trial-level exclusion criteria were implemented for face recognition and valence attribution during training as well as the independent face recognition task. Trial-level exclusion criteria were not applied for cued recall of person-knowledge during training. The following tables (Supplementary Table 5 and 6) report the average number of trials excluded per participant for each task.

Supplementary Table 5: Average number of trials excluded per participant for each session of training.

Session	Face Recognition	Valence Attribution
1	1.386	1.014
2	1.471	1.157
3	1.571	1.129
4	1.443	1.086
5	1.543	1.157

Supplementary Table 6: Average number of trials excluded per participant for the pre- and post-training independent face recognition tasks.

Session	Independent Face Recognition
Pre-training	2.157

Post-training	1.986

The following tables (Supplementary Table 7 and Table 8) further specify the data exclusion criteria used to exclude trials.

Supplementary Table 7: Average number of trials excluded per participant for each session of training by data exclusion criterion.

Session	**Face Recognition**			**Valence Attribution**		
	3 SD away from participant average RT	< 100 ms	3 SD away from group average RT	3 SD away from participant average RT	< 100 ms	3 SD away from group average RT
1	1.371	0	0.014	0.857	0	0.157
2	1.429	0	0.043	0.943	0	0.214
3	1.500	0	0.071	1.086	0	0.043
4	1.371	0	0.071	1.071	0	0.014
5	1.514	0	0.029	1.143	0	0.014

Supplementary Table 8: Average number of trials excluded per participant for the pre- and post-training independent face recognition tasks by data exclusion criterion.

Session	**Independent Face Recognition**		
	3 SD away from participant average RT	< 100 ms	3 SD away from group average RT
Pre-training	1.529	0	0.343
Post-training	1.571	0.286	0.414

Supplementary Discussion 4: Tables Decomposing Main Text Interactions

This section reports the simple differences and slopes decomposing significant interactions reported in the main text of the current paper. Final omnibus models, which were fitted using the procedure detailed by Bates and colleagues (2015), as well as the follow-up models reflected in this section may be found in the analysis script made available on OSF (https://osf.io/vrxj9).

4.1. Face Recognition. The omnibus model for face recognition during training used the following contrast codes: Race, Black (+0.5), White (-0.5); Session, session 1 (-2), session 2 (-1), session 3 (0), session 4 (+1), session 5 (+2); and Information Type, perceptually familiar only (-2/3), positive person-knowledge (+1/3), negative person-knowledge (+1/3). Follow-up analyses were run for the significant Information Type x Contact and Race x Session x Information Type interactions. All contrasts with at least one significant finding are detailed in the tables within this section. The following two paragraphs describe the additional contrast codes utilized in these models.

To break down the Information Type conditions, we compared perceptually familiar only (PF) faces, faces paired with positive person-knowledge (Pos PK), and faces paired with negative person-knowledge (Neg PK). For simple effects of Information Type, we utilized the following contrast codes: PF v. Pos PK, PF (-0.5), Pos PK (+0.5), Neg PK (0); PF v. Neg PK, PF (-0.5), Neg PK (+0.5), Pos PK (0); and Pos PK v. Neg PK, Pos PK (+0.5), Neg PK (-0.5), PF (0).

To identify patterns over the training period for the Race x Session x Information Type interaction, we specifically compared sessions 1, 3, and 5. For simple effects of Session, we utilized the following contrast codes: sessions 1 v. 3, 1 (-0.5), 3 (+0.5), all other sessions (0); sessions 1 v. 5, 1 (-0.5), 5 (+0.5), all other sessions (0); and sessions 3 v. 5, 3 (-0.5), 5 (+0.5), all other sessions (0).

Contrast Statistics for **Information Type x Contact** Interaction:

Supplementary Table 9. Simple effects of Information Type.

Analyses	Contact Level	*b*	*SE*	$CI_{95\%}$	*t*	*df*	*p*
PF v. Pos PK	Low	.189	.037	[0.117, 0.262]	5.122	68.000	**<.001***
	Average	.119	.016	[0.087, 0.152]	7.265	68.000	**<.001***
	High	.050	.037	[-0.023, 0.122]	1.339	68.000	.185
PF v. Neg PK	Low	.219	.036	[0.149, 0.288]	6.131	68.000	**<.001***
	Average	.133	.016	[0.102, 0.164]	8.414	68.000	**<.001***
	High	.048	.036	[-0.022, 0.118]	1.352	68.000	.181

Contrast Statistics for **Race x Session x Information Type** Interaction:

Supplementary Table 10. Simple effects of perceived Race.

Analyses	Session	Trial Type	*b*	*SE*	$CI_{95\%}$	*t*	*df*	*p*
Black v. White	1	PF	-.155	.021	[-0.196, -0.114]	-7.419	229.392	**<.001***
		Pos PK	-.174	.023	[-0.219, -0.129]	-7.591	2016.000	**<.001***
		Neg PK	-.165	.023	[-0.210, -0.121]	-7.316	2016.000	**<.001***
	3	PF	-.145	.016	[-0.176, -0.114]	-9.309	90.402	**<.001***
		Pos PK	-.114	.013	[-0.140, -0.088]	-8.595	2016.000	**<.001***
		Neg PK	-.093	.013	[-0.119, -0.068]	-7.143	2016.000	**<.001***
	5	PF	-.135	.020	[-0.175, -0.095]	-6.609	240.216	**<.001***
		Pos PK	-.054	.022	[-0.097, -0.010]	-2.405	456.313	**.017***
		Neg PK	-.021	.023	[-0.065, 0.023]	-0.932	2016.000	.351

Supplementary Table 11. Simple effects of Session.

Analyses	Info Type	Trial Type	*b*	*SE*	$CI_{95\%}$	*t*	*df*	*p*
Sessions 1 v. 3	PF	Black	.246	.029	[0.189, 0.303]	8.400	2016.000	**<.001***

		White	.242	.029	[0.185, 0.300]	8.278	2016.000	**<.001***
	Pos PK	Black	.281	.031	[0.220, 0.341]	9.071	2016.000	**<.001***
		White	.159	.031	[0.098, 0.220]	5.134	2016.000	**<.001***
	Neg PK	Black	.277	.031	[0.217, 0.337]	9.061	2016.000	**<.001***
		White	.157	.031	[0.097, 0.217]	5.139	2016.000	**<.001***
Sessions 1 v. 5	PF	Black	.314	.033	[0.249, 0.379]	9.442	143.836	**<.001***
		White	.304	.029	[0.248, 0.360]	10.666	180.118	**<.001***
	Pos PK	Black	.320	.030	[0.262, 0.379]	10.740	2016.000	**<.001***
		White	.186	.030	[0.128, 0.245]	6.247	2016.000	**<.001***
	Neg PK	Black	.323	.029	[0.265, 0.381]	10.980	2016.000	**<.001***
		White	.181	.029	[0.124, 0.239]	6.163	2016.000	**<.001***
Sessions 3 v. 5	PF	Black	.068	.032	[0.006, 0.131]	2.146	2016.000	**.032***
		White	.062	.032	[-0.001, 0.124]	1.941	2016.000	.052
	Pos PK	Black	.039	.033	[-0.026, 0.105]	1.181	2016.000	.238
		White	.027	.033	[-0.038, 0.093]	0.819	2016.000	.413
	Neg PK	Black	.046	.033	[-0.019, 0.111]	1.391	2016.000	.164
		White	.024	.033	[-0.041, 0.089]	0.732	2016.000	.464

Supplementary Table 12. Simple effects of Information Type.

Analyses	**Session**	**Trial Type**	***b***	***SE***	***CI$_{95\%}$***	***t***	***df***	***p***
PF v. Pos PK	1	Black	.142	.022	[0.098, 0.185]	6.409	2016.000	**<.001***
		White	.160	.022	[0.117, 0.204]	7.257	2016.000	**<.001***
	3	Black	.135	.013	[0.110, 0.160]	10.588	2016.000	**<.001***
		White	.104	.013	[0.079, 0.129]	8.138	2016.000	**<.001***

	5	Black	.129	.023	[0.084, 0.173]	5.627	203.471	**<.001***
		White	.047	.021	[0.006, 0.089]	2.246	203.108	**.026***
PF v. Neg PK	1	Black	.160	.022	[0.117, 0.202]	7.335	2016.000	**<.001***
		White	.170	.022	[0.127, 0.213]	7.803	2016.000	**<.001***
	3	Black	.159	.019	[0.123, .0.196]	8.581	76.715	**<.001***
		White	.108	.016	[0.077, 0.138]	6.855	81.885	**<.001***
	5	Black	.159	.023	[0.113, 0.204]	6.815	178.201	**<.001***
		White	.045	.021	[0.005, 0.085]	2.178	218.423	**.030***
Pos PK v. Neg PK	1	Black	-.018	.023	[-0.064, 0.027]	-0.783	2016.000	0.434
		White	-.010	.023	[-0.055, 0.036]	-0.415	2016.000	0.678
	3	Black	-.024	.013	[-0.050, 0.002]	-1.804	2016.000	0.071
		White	-.004	.013	[-0.030, 0.023]	-0.273	2016.000	0.785
	5	Black	-.030	.023	[-0.076, 0.015]	-1.300	2016.000	0.194
		White	.002	.023	[-0.043, 0.048]	0.099	2016.000	.921

Supplementary Table 13. Comparison of racial bias across sessions. Positive b *values indicate increased preferential accuracy for perceived White faces (coded as -0.5) at the earlier session (coded as -0.5).*

Analyses	**Trial Type**	***b***	***SE***	***CI95%***	***t***	***df***	***p***
Sessions 1 v. 3	PF	.082	.026	[0.032, 0.132]	3.205	2016.000	**.001***
	Pos PK	.082	.024	[0.034, 0.129]	3.385	2016.000	**<.001***
	Neg PK	.082	.024	[0.034, 0.130]	3.343	2016.000	**<.001***
Sessions 1 v. 5	PF	.095	.025	[0.047, 0.143]	3.872	2016.000	**<.001***
	Pos PK	.095	.023	[0.050, 0.141]	4.113	2016.000	**<.001***

	Neg PK	.095	.023	[0.049, 0.141]	4.057	2016.000	**<.001***
Sessions 3 v. 5	PF	.013	.027	[-0.040, 0.067]	0.491	2016.000	.624
	Pos PK	.013	.026	[-0.038, 0.065]	0.514	2016.000	.608
	Neg PK	.013	.026	[-0.038, 0.065]	0.508	2016.000	.611

4.2. Valence Attribution. The omnibus model for valence attribution during training used the following contrast codes: Race, Black (+0.5), White (-0.5); Session, session 1 (-2), session 2 (-1), session 3 (0), session 4 (+1), session 5 (+2); and Information Type, perceptually familiar only (-2/3), positive person-knowledge (+1/3), negative person-knowledge (+1/3). Follow-up analyses were run for the significant Race x Information Type and Session x Information Type interactions. All contrasts with at least one significant finding are detailed in the tables within this section. The following two paragraphs describe the additional contrast codes utilized in these models.

To break down the Information Type conditions, we compared perceptually familiar only (PF) faces, faces paired with positive person-knowledge (Pos PK), and faces paired with negative person-knowledge (Neg PK). For simple effects of Information Type, we utilized the following contrast codes: PF v. Pos PK, PF (-0.5), Pos PK (+0.5), Neg PK (0); PF v. Neg PK, PF (-0.5), Neg PK (+0.5), Pos PK (0); and Pos PK v. Neg PK, Pos PK (+0.5), Neg PK (-0.5), PF (0).

To identify patterns over the training period for the Session x Information Type interaction, we specifically compared sessions 1, 3, and 5. For simple effects of Session, we utilized the following contrast codes: sessions 1 v. 3, 1 (-0.5), 3 (+0.5), all other sessions (0); sessions 1 v. 5, 1 (-0.5), 5 (+0.5), all other sessions (0); and sessions 3 v. 5, 3 (-0.5), 5 (+0.5), all other sessions (0).

Contrast Statistics for **Race x Information Type** Interaction:

Supplementary Table 14. Simple effects of perceived Race.

Analyses	Trial Type	*b*	*SE*	*CI*$_{95\%}$	*t*	*df*	*p*
Black v. White	PF	.002	.015	[-0.027, 0.031]	0.156	2011.041	.876
	Pos PK	-.016	.015	[-0.045, 0.013]	-1.052	2011.008	.293
	Neg PK	-.093	.015	[-0.122, -0.064]	-6.335	2011.002	**<.001***

Supplementary Table 15. Simple effects of Information Type.

Analyses	Trial Type	*b*	*SE*	*CI*$_{95\%}$	*t*	*df*	*p*
Pos PK v. PF	Black	-.055	.015	[-0.084, -0.026]	-3.695	2011.021	**<.001***
	White	-.037	.015	[-0.066, -0.008]	-2.495	2011.028	**.013***
PF v. Neg PK	Black	-.085	.015	[-0.113, -0.056]	-5.759	2011.014	**<.001***
	White	.011	.015	[-0.018, 0.040]	0.745	2011.028	.456
Pos PK v. Neg PK	Black	.030	.015	[0.001, 0.059]	2.033	2011.008	**.042***
	White	-.048	.015	[-0.077, -0.019]	-3.228	2011.002	**.001***

Contrast Statistics for **Session x Information Type** Interaction:

Supplementary Table 16. Simple effects of Session.

Analyses	Trial Type	*b*	*SE*	*CI*$_{95\%}$	*t*	*df*	*p*
Sessions 1 v. 3	PF	.249	.023	[0.203, 0.295]	10.688	2011.048	**<.001***
	Pos PK	.143	.023	[0.097, 0.188]	6.115	2011.017	**<.001***
	Neg PK	.162	.023	[0.116, 0.207]	6.971	2011.002	**<.001***
Sessions 1 v. 5	PF	.237	.023	[0.191, 0.283]	10.134	2011.048	**<.001***
	Pos PK	.143	.023	[0.097, 0.189]	6.089	2011.017	**<.001***

	Neg PK	.149	.023	[0.104, 0.195]	6.405	2011.002	**<.001***
Sessions 3 v. 5	PF	-.012	.024	[-0.059, 0.036]	-0.481	2011.002	.631
	Pos PK	.000	.024	[-0.048, 0.048]	-0.001	2011.002	.999
	Neg PK	-.013	.024	[-0.060, 0.035]	-0.517	2011.002	.605

Supplementary Table 17. Simple effects of Information Type.

Analyses	**Session**	***b***	***SE***	***CI*** $_{95\%}$	***t***	***df***	***p***
Pos PK v. PF	1	-.003	.018	[-0.039, 0.032]	-0.184	2011.040	.854
	3	-.046	.010	[-0.066, -0.025]	-4.377	2011.024	**<.001***
	5	-.088	.018	[-0.124, -0.053]	-4.875	2011.006	**<.001***
PF v. Neg PK	1	.003	.018	[-0.032, 0.039]	0.175	2011.029	.861
	3	-.037	.010	[-0.057, -0.016]	-3.545	2011.020	**<.001***
	5	-.077	.018	[-0.112, -0.042]	-4.272	2011.005	**<.001***
Pos PK v. Neg PK	1	-.006	.018	[-0.042, 0.029]	-0.354	2011.011	.723
	3	-.009	.010	[-0.029, 0.012]	-0.843	2011.005	.399
	5	-.011	.018	[-0.047, 0.024]	-0.619	2011.003	.536

4.3. Cued Recall of Person-Knowledge. The omnibus model for cued recall of person-knowledge during training used the following contrast codes: Race, Black (+0.5), White (-0.5); Session, session 1 (-2), session 2 (-1), session 3 (0), session 4 (+1), session 5 (+2); and Valence, negative (-0.5), positive (+0.5). Follow-up analyses were run for the significant Race x Session x Contact and Race x Valence x Contact interactions. All contrasts with at least one significant finding are detailed in the tables within this section.

To identify patterns over the training period for the Race x Session x Contact interaction, we specifically compared sessions 1, 3, and 5. For simple effects of Session, we utilized the following contrast codes: sessions 1 v. 3, 1 (-0.5), 3 (+0.5), all other sessions (0); sessions 1 v. 5, 1 (-0.5), 5 (+0.5), all other sessions (0); and sessions 3 v. 5, 3 (-0.5), 5 (+0.5), all other sessions (0).

Contrast Statistics for **Race x Session x Contact** Interaction:

Supplementary Table 18. Simple effects of perceived Race.

Analyses	**Session**	**Contact Level**	***b***	***SE***	***CI$_{95\%}$***	***t***	***df***	***p***
Black v. White	1	Low	-.011	.024	[-0.058, 0.037]	-0.442	103.507	.659
		Average	-.038	.011	[-0.059, -0.017]	-3.510	105.833	**<.001***
		High	-.065	.025	[-0.114, -0.017]	-2.663	108.674	**.009***
	3	Low	-.070	.022	[-0.113, -0.028]	-3.225	67.413	**.002***
		Average	-.059	.010	[-0.078, -0.040]	-6.095	68.175	**<.001***
		High	-.048	.022	[-0.091, -0.005]	-2.196	69.016	**.031***
	5	Low	-.123	.027	[-0.176, -0.069]	-4.477	1271.099	**<.001***
		Average	-.082	.012	[-0.106, -0.057]	-6.616	1271.188	**<.001***
		High	-.041	.028	[-0.096, 0.014]	-1.452	1271.466	.147

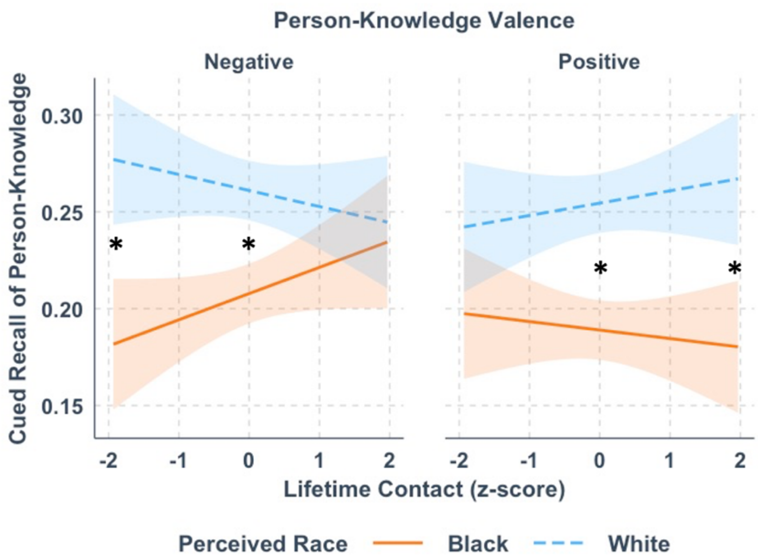
Supplementary Table 19. Simple effects of Session.

Analyses	**Contact Level**	**Trial Type**	***b***	***SE***	***CI$_{95\%}$***	***t***	***df***	***p***
Sessions 1 v. 3	Low	Black	.113	.041	[0.033, 0.193]	2.769	1271.014	**.006***
		White	.190	.040	[0.111, 0.270]	4.697	1271.054	**<.001***
	Average	Black	.119	.018	[0.084, 0.155]	6.532	1271.109	**<.001***
		White	.162	.018	[0.126, 0.198]	8.913	1271.261	**<.001***

	High	Black	.126	.042	[0.045, 0.207]	3.033	1271.137	**.002***
		White	.134	.041	[0.053, 0.214]	3.257	1271.244	**.001***
Sessions 1 v. 5	Low	Black	.145	.037	[0.073, 0.217]	3.923	1271.048	**<.001***
		White	.245	.037	[0.173, 0.317]	6.630	1271.034	**<.001***
	Average	Black	.214	.017	[0.182, 0.247]	12.877	1271.169	**<.001***
		White	.261	.017	[0.229, 0.294]	15.700	1271.156	**<.001***
	High	Black	.284	.038	[0.210, 0.368]	7.540	1271.280	**<.001***
		White	.278	.038	[0.203, 0.353]	7.293	1271.260	**<.001***
Sessions 3 v. 5	Low	Black	.031	.041	[-0.050, 0.113]	0.752	1271.009	.452
		White	.061	.042	[-0.020, 0.143]	1.473	1271.145	.141
	Average	Black	.093	.019	[0.056, 0.129]	4.982	1271.080	**<.001***
		White	.096	.019	[0.060, 0.133]	5.147	1271.293	**<.001***
	High	Black	.154	.042	[0.072, 0.237]	3.665	1271.234	**<.001***
		White	.132	.042	[0.049, 0.215]	3.107	1271.689	**.002***

Figure and Contrast Statistics for **Race x Valence x Contact** Interaction:

Supplementary Figure 1. High-contact participants were similarly accurate at recalling negative person-knowledge for perceived Black and White faces. Low-contact participants were similarly accurate at recalling positive person-knowledge for perceived Black and White faces. The shaded region surrounding each linear regression line represents the 95% confidence interval for the regression coefficient.



Supplementary Table 20. Simple effects of perceived Race.

Analyses	Contact Level	Trial Type	*b*	*SE*	*CI*$_{95\%}$	*t*	*df*	*p*
Black v. White	Low	Pos PK	-.043	.022	[-0.086, 0.001]	-1.912	1271.038	.056
		Neg PK	-.096	.022	[-0.140, -0.052]	-4.270	1271.037	**<.001***
	Average	Pos PK	-.065	.010	[-0.085, -0.046]	-6.551	1271.098	**<.001***
		Neg PK	-.056	.010	[-0.076, -0.037]	-5.571	1271.136	**<.001***
	High	Pos PK	-.088	.023	[-0.132, -0.044]	-3.908	1271.218	**<.001***
		Neg PK	-.017	.023	[-0.062, 0.028]	-0.726	1271.110	.468

Supplementary Table 21. Simple effects of Valence.

Analyses	Contact Level	Trial Type	*b*	*SE*	*CI*$_{95\%}$	*t*	*df*	*p*
Positive v. Negative	Low	Black	.018	.022	[-0.026, 0.062]	0.823	1271.238	.410
		White	-.036	.020	[-0.075, 0.004]	-1.777	97.557	.079
	Average	Black	-.017	.010	[-0.037, 0.003]	-1.661	1271.357	.097
		White	-.006	.009	[-0.024, 0.011]	-0.687	99.539	.494

	High	Black	-.052	.023	[-0.097, -0.007]	-2.273	1271.877	**.023***
		White	.023	.020	[-0.017, 0.063]	1.150	101.400	.253

4.4. Independent Face Recognition. The omnibus model for the independent face recognition task used the following contrast codes: Race, Black (+0.5), White (-0.5); Familiarity, novel (-0.5), familiar (+0.5); and Session, pre-training (-0.5), post-training (+0.5). Follow-up analyses were run for the significant Session x Contact and Race x Session x Familiarity interactions. All contrasts with at least one significant finding are detailed in the tables within this section.

Contrast Statistics for **Session x Contact** Interaction:

Supplementary Table 22. Simple effects of Session.

Analyses	**Contact Level**	***b***	***SE***	***CI₉₅%***	***t***	***df***	***p***
Pre v. Post	Low	-.127	.038	[-0.202, -0.051]	-3.294	68.000	**.002***
	Average	-.049	.017	[-0.082, -0.015]	-2.846	68.000	**.006***
	High	.029	.038	[-0.046, 0.105]	0.763	68.000	.448

Supplementary Table 23. Simple effects of Contact.

Analyses	***b***	***SE***	***CI₉₅%***	***t***	***df***	***p***
Pre	-.031	.014	[-0.059, -0.003]	-2.179	119.645	**.031***
Post	.008	.014	[-0.020, 0.036]	0.561	119.645	.576

Contrast Statistics for **Race x Session x Familiarity** Interaction:

Supplementary Table 24. Simple effects of perceived Race.

Analyses	**Session**	**Trial Type**	***b***	***SE***	***CI₉₅%***	***t***	***df***	***p***
Black v. White	Pre	Novel	-.115	.029	[-0.171, -0.059]	-4.017	483.000	**<.001***
		Familiar	.000	.029	[-0.056, 0.056]	0.003	483.000	.998

	Post	Novel	-.005	.029	[-0.061, 0.051]	-0.172	483.000	.863
		Familiar	-.047	.029	[-0.103, 0.010]	-1.625	483.000	.105

Supplementary Table 25. Simple effects of Session.

Analyses	**Familiarity**	**Trial Type**	***b***	***SE***	***CI*** $_{95\%}$	***t***	***df***	***p***
Pre v. Post	Novel	Black	.040	.029	[-0.016, 0.097]	1.411	483.000	.159
		White	-.070	.029	[-0.126, -0.014]	-2.435	483.000	**.015***
	Familiar	Black	-.106	.029	[-0.162, -0.050]	-3.701	483.000	**<.001***
		White	-.059	.029	[-0.116, -0.003]	-2.074	483.000	**.039***

Supplementary Table 26. Simple effects of Familiarity.

Analyses	**Session**	**Trial Type**	***b***	***SE***	***CI*** $_{95\%}$	***t***	***df***	***p***
Novel v. Familiar	Pre	Black	.020	.029	[-0.036, 0.077]	0.715	483.000	.475
		White	-.095	.029	[-0.151, -0.039]	-3.305	483.000	**.001***
	Post	Black	-.126	.029	[-0.182, -0.070]	-4.397	483.000	**<.001***
		White	-.084	.029	[-0.141, -0.028]	-2.944	483.000	**.003***

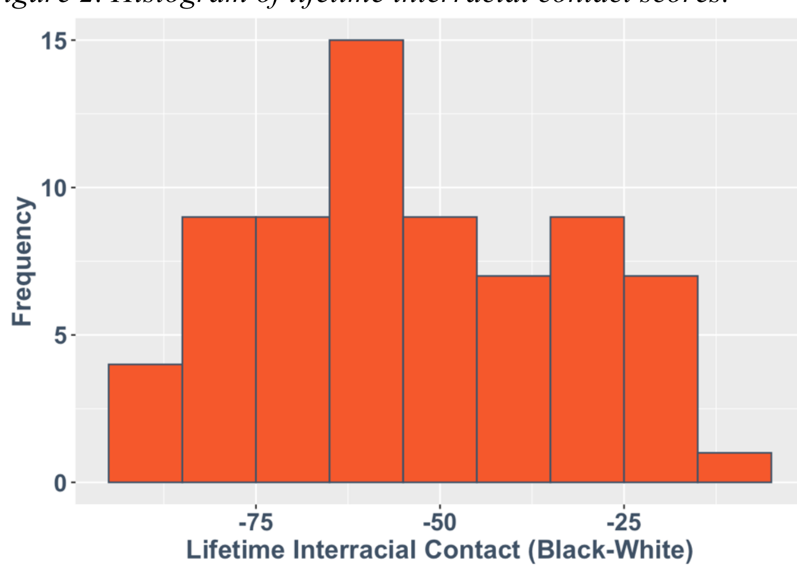
Supplementary Discussion 5: Descriptive Statistics

This section reports the descriptive statistics for the interracial contact questionnaire and all memory tasks discussed in the main text of the current paper.

5.1. Lifetime Interracial Contact. As described in the main text, a lifetime interracial contact index was calculated from participant responses on the contact questionnaire. This lifetime interracial contact index is an average of childhood contact and current contact, each calculated as a difference score of the participant's average Black contact for that time period minus their average White contact for that time period. Consequently, negative values in the lifetime interracial contact index indicate comparatively more White contact relative to Black contact. Although the lifetime interracial contact index was z-scored for analysis, here we provide descriptive statistics for the raw lifetime interracial contact scores to facilitate ease in interpretation.

The sample mean of lifetime interracial contact was -53.345 with a standard deviation of 20.889. Supported also by the histogram below, this reflects that participants had more contact with White individuals relative to Black individuals throughout the lifetime.

Supplementary Figure 2. Histogram of lifetime interracial contact scores.



5.2. Face Recognition.

Supplementary Table 27. Descriptive statistics for face recognition accuracy during training.

Session	Perceived Race	Information Type	Mean Accuracy	Standard Deviation
1	Black	Perceptually Familiar	0.446	0.346
1	Black	Negative Person-knowledge	0.574	0.327
1	Black	Positive Person-knowledge	0.547	0.305
1	White	Perceptually Familiar	0.570	0.297
1	White	Negative Person-knowledge	0.750	0.256
1	White	Positive Person-knowledge	0.749	0.240
2	Black	Perceptually Familiar	0.566	0.302
2	Black	Negative Person-knowledge	0.764	0.258
2	Black	Positive Person-knowledge	0.755	0.248
2	White	Perceptually Familiar	0.765	0.237
2	White	Negative Person-knowledge	0.902	0.192
2	White	Positive Person-knowledge	0.881	0.194
3	Black	Perceptually Familiar	0.692	0.302
3	Black	Negative Person-knowledge	0.851	0.217
3	Black	Positive Person-knowledge	0.827	0.232
3	White	Perceptually Familiar	0.813	0.247
3	White	Negative Person-knowledge	0.907	0.186
3	White	Positive Person-knowledge	0.908	0.195
4	Black	Perceptually Familiar	0.688	0.327
4	Black	Negative Person-knowledge	0.864	0.244
4	Black	Positive Person-knowledge	0.832	0.249
4	White	Perceptually Familiar	0.856	0.213
4	White	Negative Person-knowledge	0.926	0.170
4	White	Positive Person-knowledge	0.924	0.184
5	Black	Perceptually Familiar	0.761	0.247
5	Black	Negative Person-knowledge	0.897	0.202
5	Black	Positive Person-knowledge	0.867	0.201
5	White	Perceptually Familiar	0.874	0.223
5	White	Negative Person-knowledge	0.931	0.180
5	White	Positive Person-knowledge	0.935	0.161

5.2. Valence Attribution.

Supplementary Table 28. Descriptive statistics for valence attribution accuracy during training.

Session	Perceived Race	Information Type	Mean Accuracy	Standard Deviation
1	Black	Perceptually Familiar	0.609	0.353
1	Black	Negative Person-knowledge	0.553	0.298
1	Black	Positive Person-knowledge	0.655	0.229

1	White	Perceptually Familiar	0.613	0.354
1	White	Negative Person-knowledge	0.710	0.286
1	White	Positive Person-knowledge	0.638	0.275
2	Black	Perceptually Familiar	0.786	0.313
2	Black	Negative Person-knowledge	0.697	0.306
2	Black	Positive Person-knowledge	0.716	0.239
2	White	Perceptually Familiar	0.774	0.311
2	White	Negative Person-knowledge	0.819	0.232
2	White	Positive Person-knowledge	0.740	0.258
3	Black	Perceptually Familiar	0.860	0.246
3	Black	Negative Person-knowledge	0.764	0.261
3	Black	Positive Person-knowledge	0.760	0.238
3	White	Perceptually Familiar	0.862	0.260
3	White	Negative Person-knowledge	0.823	0.266
3	White	Positive Person-knowledge	0.815	0.247
4	Black	Perceptually Familiar	0.849	0.294
4	Black	Negative Person-knowledge	0.777	0.291
4	Black	Positive Person-knowledge	0.759	0.280
4	White	Perceptually Familiar	0.840	0.277
4	White	Negative Person-knowledge	0.815	0.265
4	White	Positive Person-knowledge	0.775	0.273
5	Black	Perceptually Familiar	0.843	0.264
5	Black	Negative Person-knowledge	0.736	0.293
5	Black	Positive Person-knowledge	0.789	0.246
5	White	Perceptually Familiar	0.855	0.267
5	White	Negative Person-knowledge	0.826	0.264
5	White	Positive Person-knowledge	0.786	0.266

5.3. Cued Recall of Person-Knowledge.

Supplementary Table 29. Descriptive statistics for cued recall accuracy during training.

Session	Perceived Race	Information Type	Mean Accuracy	Standard Deviation
1	Black	Negative	0.099	0.123
1	Black	Positive	0.071	0.101
1	White	Negative	0.130	0.133
1	White	Positive	0.107	0.128
2	Black	Negative	0.178	0.167
2	Black	Positive	0.144	0.140
2	White	Negative	0.198	0.186
2	White	Positive	0.212	0.187
3	Black	Negative	0.211	0.187
3	Black	Positive	0.195	0.174
3	White	Negative	0.296	0.240

3	White	Positive	0.267	0.232
4	Black	Negative	0.288	0.270
4	Black	Positive	0.268	0.258
4	White	Negative	0.332	0.294
4	White	Positive	0.327	0.285
5	Black	Negative	0.294	0.293
5	Black	Positive	0.292	0.294
5	White	Negative	0.378	0.305
5	White	Positive	0.379	0.330

5.3. Independent Face Recognition.

Supplementary Table 30. Descriptive statistics for face recognition accuracy on the independent face recognition task.

Session	Perceived Race	Familiarity	Mean Accuracy	Standard Deviation
Pre	Black	Familiar	0.650	0.198
Pre	Black	Novel	0.629	0.210
Pre	White	Familiar	0.650	0.157
Pre	White	Novel	0.744	0.209
Post	Black	Familiar	0.544	0.170
Post	Black	Novel	0.670	0.189
Post	White	Familiar	0.590	0.171
Post	White	Novel	0.675	0.198

Supplementary Discussion 6: Exploratory Follow-Up Analysis for Independent Face Recognition Task

This section reports an exploratory analysis of false alarms (novel faces incorrectly identified as familiar) in the independent face recognition task. This analysis was conducted at request during the peer review process to explore potential convergence between false alarms and the observed decline in overall accuracy from pre- to post-training on the independent face recognition task among low- and average- contact participants.

Following the pre-registered accuracy analysis for this task (hits/total hits possible), false alarms were calculated as the number of false alarms divided by the total number of novel faces in a condition. This false alarms DV was fitted as a function of perceived race, task session, and lifetime interracial contact. The omnibus model used the following contrast codes: Race, Black (+0.5), White (-0.5); and Session, pre-training (-0.5), post-training (+0.5). The same procedure from the models reported in the main text was performed to determine random effects structures and decompose significant interactions.

Results revealed significant main effects of Race, b= 0.059, SE= 0.005, *95% confidence interval* ($CI_{95\%}$)=[0.048,0.069], t(2164.000)=10.680, p<.001; and Session, b= 0.013, SE= 0.005, $CI_{95\%}$=[0.002,0.024], t(2164.000)= 2.397, p= 0.017. These were qualified by a significant two-way interaction of Race x Session, b= -0.108, SE= 0.011, $CI_{95\%}$=[-0.130,-0.087], t(2164.000)=-9.865, p<.001. There was also a significant interaction of Session x Contact, b= -0.061, SE=0.006, $CI_{95\%}$=[-0.072,-0.050], t(2164.000)=-11.048, p<.001.

We analyzed simple effects for the significant interaction of Race x Session (see Supplementary Table 31 and 32). Participants demonstrated significantly more false alarms for perceived Black faces compared to perceived White faces only at the pre-training session. There was no significant difference based on perceived race at the post-training session. From pre- to post-training, participants showed significantly decreased false alarms for perceived Black faces and significantly increased false alarms for perceived White faces.

Contrast Statistics for **Race x Session x Familiarity** Interaction:

Supplementary Table 31. Simple effects of perceived Race.

Analyses	Session	b	SE	$CI_{95\%}$	t	df	p
Black v. White	Pre	.113	.008	[0.098, 0.128]	14.527	2164.000	**<.001***
	Post	.004	.008	[-0.011, 0.020]	0.577	2164.000	.564

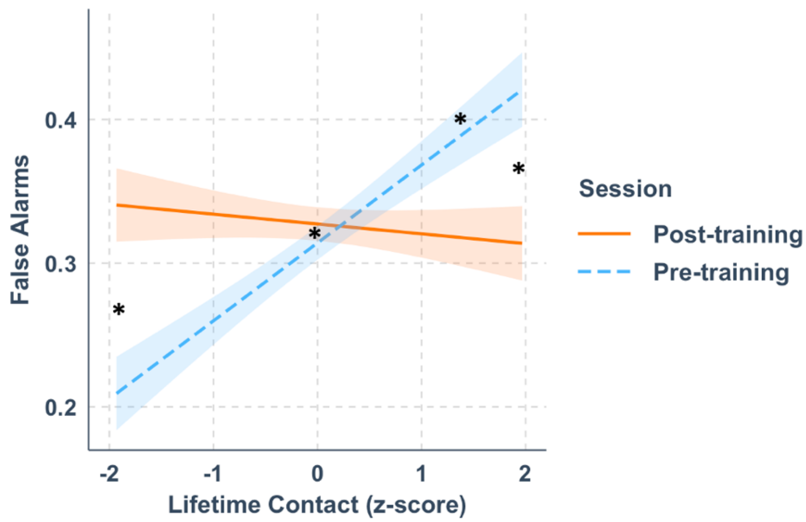
Supplementary Table 32. Simple effects of Session.

Analyses	Perceived Race	b	SE	$CI_{95\%}$	t	df	p
Pre v. Post	Black	-.041	.008	[-0.056, -0.026]	-5.280	2164.000	**<.001***
	White	.067	.008	[0.052, 0.083]	8.670	2164.000	**<.001***

When decomposing the significant interaction of Session x Contact, for simple effects of contact, which is a z-scored continuous variable, we analyzed simple differences at the following scores: 0 (the sample average for contact), +2 (two standard deviations above the sample average), and -2 (two standard deviations below the sample average). The simple differences were analyzed to describe the interaction.

Follow-up analyses for the Session x Contact interaction (see Supplementary Figure 3 and Table 33 and 34 below) reflect that both low- and average-contact participants demonstrated significantly increased false alarms at the post-training session compared to the pre-training session, while high-contact participants demonstrated significantly decreased false alarms from pre- to post-training. Additionally, the contact slope was significant only at the pre-training session, meaning that before training, false alarms increased as contact increased. In other words, even though lower contact participants started off at the pre-training session demonstrating less false alarms, they increased their false alarms post-training as higher contact participants simultaneously reduced their false alarms.

Supplementary Figure 3. From pre- to post-training, participants with less contact experienced an increase in false alarms and participants with high contact decreased their false alarms.



Contrast Statistics for **Session x Contact** Interaction:

Supplementary Table 33. Simple effects of Session.

Analyses	**Contact Level**	***b***	***SE***	***CI_{95%}***	***t***	***df***	***p***
Pre v. Post	Low	.135	.012	[0.111, 0.160]	10.962	2164.000	**<.001***
	Average	.013	.005	[0.002, 0.024]	2.397	2164.000	**.017***
	High	-.109	.012	[-0.133, -0.085]	-8.830	2164.000	**<.001***

Supplementary Table 34. Simple effects of Contact.

Analyses	***b***	***SE***	***CI_{95%}***	***t***	***df***	***p***
Pre	.054	.018	[0.018, 0.090]	2.972	71.229	**.004***
Post	-.007	.018	[-0.043, 0.029]	-0.375	71.229	.709

Supplementary References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

DeBruine, L., & Jones, B. (2017). *Face Research Lab London Set* (p. 88208545 Bytes) [dataset]. figshare. https://doi.org/10.6084/M9.FIGSHARE.5047666.V2

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Kennedy, K. M., Hope, K., & Raz, N. (2009). Life Span Adult Faces: Norms for age, familiarity, memorability, mood, and picture quality. *Experimental Aging Research*, *35*(2), 268–275. https://doi.org/10.1080/03610730902720638

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Mende-Siedlecki, P., & Havlicek, L. (in prep). *The Delaware behavior database: A set of social behaviors and corresponding norming data.* https://osf.io/wg8by/

Mende-Siedlecki, P., Qu-Lee, J., Lin, J., Drain, A., & Goharzad, A. (2020). The Delaware Pain Database: A set of painful expressions and corresponding norming data. *PAIN Reports*, *5*(6), e853. https://doi.org/10.1097/PR9.0000000000000853

Nordstrøm, M. M., Larsen, M., Sierakowski, J., & Stegmann, M. B. (2004). The IMM Face Database—An annotated dataset of 240 face images. *Technical University of Denmark, DTU Informatics, Building 321*. http://www.imm.dtu.dk/pubdb/p.php?3160

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, *48*(3), 1197–1204. https://doi.org/10.3758/s13428-015-0641-9

Thomaz, C. E., & Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, *28*(6), 902–913. https://doi.org/10.1016/j.imavis.2009.11.005

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671