Supplemental Materials

Shifts in facial impression structures across group boundaries

Youngki Hong[1] & Jonathan B. Freeman[1]

[1]Columbia University

Analysis S1. The baseline relationship between trustworthiness and dominance (Study 1)

To establish a baseline for the relationship between trustworthiness and dominance, upon which our expected overall negative relationships between trustworthiness and dominance for ingroup and outgroup were built, we calculated the correlation between Chicago Face Database (CFD) norming ratings of trustworthiness and dominance (Ma et al., 2015) for all 80 face identities used in Study 1 ($M_{trustworthy}$ = 3.41,$SD_{trustworthy}$ = .36, $Range_{trustworthy}$ = 1.76; $M_{dominant}$ = 2.78, $SD_{dominant}$ = .61, $Range_{dominant}$ = 2.54). We found a significantly negative correlation between trustworthiness and dominance, $r(78) = -.49$, 95% CI = [-.64, -.30], $p < .001$. Thus, an overall negative relationship between trustworthiness and dominance for both the ingroup and outgroup was expected. Our prediction was that it would be relatively more negative for the outgroup and relatively more positive (i.e., less negative) for the ingroup.

Analysis S2. Manipulation check (Study 1)

In Study 1, after completing face and stereotype ratings, each participant answered questions about their familiarity with Kandinsky and Klee, their agreement with the group assignment, and their collective identification with the assigned ingroup. The latter two served as manipulation checks. For all questions, participants responded on a 7-point scale (1 – strongly disagree, 7 – strongly agree).

For both Kandinsky and Klee, participants were asked their familiarity about the artist and the paintings: “I'm familiar with the artist Kandinsky”, “I'm familiar with Kandinsky’s paintings”, “I'm familiar with the artist Klee”, and “I'm familiar with Klee’s paintings”. The two Kandinsky and two Klee ratings were each added together, resulting in scores ranging from 2 to 14 with a lower score indicating less familiarity. A paired t-test revealed no significant difference in familiarity between Klee ($M = 4.87$, $SD = 3.58$) and Kandinsky ($M = 5.14$, $SD = 3.80$), $t(256) = 1.65$, $p = .10$, 95% CI [-.05, .60], Cohen's $d = .07$. In both cases, the most frequent response was "not at all familiar" (Klee – 124 people, Kandinsky – 120 people out of 257 people).

For agreement with group assignment, participants were asked: “How much do you agree that you have a preference for paintings by Kandinsky?” and “How much do you agree that you have a preference for paintings by Klee?”. Consistent with a successful manipulation, a paired t-test revealed significantly greater agreement with a preference for ingroup paintings ($M = 5.35$, $SD = 1.26$) compared to outgroup paintings ($M = 2.78$, $SD = 1.45$), $t(256) = 18.36$, $p < .001$, 95% CI [2.29, 2.84], Cohen's $d = 1.89$. The strength of agreement did not differ overall between the Klee ($M = 4.15$, $SD = 1.88$) and Kandinsky ($M = 3.98$, $SD = 1.86$) groups, $t(256) = 0.77$, $p = .44$, 95% CI [-.26, .58], Cohen's $d = .09$.

Participants also answered a three-item collective identification measure with the ingroup ("I value being a member of the Kandinsky/Klee group", "I am proud to be a member of the Kandinsky/Klee group", "Belonging to the Kandinsky/Klee group is an important part of my identity"), used in previous studies (Ashmore et al., 2004; Van Bavel et al., 2012). The three collective identification items were added together, thus ranging from 3 to 21. Although we did not collect outgroup collective identification as a comparison (however, we added this feature to Study 2; see Analysis S7), our participants showed a level of collective identification with their ingroup (M = 10.49, SD = 4.4) commensurate with previous studies (Van Bavel et al., 2012), suggesting a successful manipulation.

Analysis S3. Ingroup and outgroup differences in face ratings (Study 1)

In Study 1, each participant randomly rated five traits from a list of 14 traits on 15 ingroup and 15 outgroup faces. For ease of analysis, we calculated the average ratings across participants for each unique face identity. Since we used 80 distinct face identities, the resulting data consisted of 80 ingroup and 80 outgroup faces, with each face identity repeating twice (once for ingroup and once for outgroup) and having ratings for all 14 traits. We then used a repeated-measures multivariate analysis of variance (rMANOVA) to test the effects of group (ingroup, outgroup) on face trait ratings. A significant multivariate effect was found for group (Pillai's Trace = .45, $F = 3.90$, $df = (14, 66)$, $p < .001$). Next, we conducted univariate F-tests to examine the effects of group on each trait. The results showed that the ingroup faces were rated as more caring, confident, intelligent, responsible, and trustworthy than the outgroup faces, whereas the outgroup faces were rated as more aggressive, mean, threatening, unhappy, and weird. See Table S1 for full details.

Table S1. Study 1 face trait rating results.

Trait	Ingroup (SD)	Outgroup (SD)	F value	Cohen's d
Aggressive	3.67 (0.68)	3.78 (0.73)	4.34*	0.23
Attractive	3.67 (0.91)	3.61 (0.90)	1.47	0.14
Caring	4.04 (0.63)	3.86 (0.70)	12.23***	0.39
Confident	4.46 (0.75)	4.24 (0.65)	13.65***	0.41
Dominant	4.01 (0.68)	4.04 (0.69)	0.26	0.06
Emotionally stable	4.28 (0.64)	4.22 (0.69)	1.20	0.12
Intelligent	4.52 (0.60)	4.20 (0.64)	27.67***	0.59
Mean	3.69 (0.71)	3.84 (0.64)	6.30*	0.28
Responsible	4.44 (0.64)	4.31 (0.66)	7.64**	0.31
Sociable	4.17 (0.70)	4.12 (0.76)	1.19	0.12
Threatening	3.23 (0.77)	3.44 (0.75)	12.00***	0.39
Trustworthy	4.15 (0.56)	3.96 (0.59)	13.05***	0.40
Unhappy	3.96 (0.76)	4.07 (0.74)	5.07*	0.25
Weird	3.73 (0.73)	3.86 (0.70)	4.73*	0.24

Significance codes: *** <.001 ** <.01 * <.05

Analysis S4. Do the two clusters reflect trustworthiness and dominance? (Study 1)

For Analyses S4 and S5, we recruited 100 independent raters from Prolific to participate in an online study about making social judgments, who received monetary compensation. After excluding two participants due to failed attention checks, our final sample size was 98 raters (M_{age} = 40.42 years, SD_{age} = 13.37 years; 53 female, 44 male, 1 other; 79 White, 7 Asian, 6 Black, 3 multiracial, 1 Pacific Islander, and 2 other). In the first block, the participants rated each of the 14 traits on valence (e.g., "How positive is the personality trait *trustworthy*?") on a 7-point scale (1 – very negative, 7 – very positive). In the second block, participants rated the similarity of each of the 14 traits used in Study 1 in relation to trustworthiness and dominance trait concepts at the same time (e.g., "Is the personality trait caring more similar to trustworthy/untrustworthy or dominant/submissive?") on a 7-point scale (1 – more similar to dominant/submissive, 7 – more similar to trustworthy/untrustworthy). The positions for trustworthiness and dominance concepts were randomized across participants so that, for some participants, 1 meant more similar to the trustworthiness concept and 7 meant more similar to the dominance concept. Each participant's rating was scored so that a lower score indicated more similarity to the dominance concept, while a higher score indicated more similarity to the trustworthiness concept.

Next, we compared relation to trustworthiness and dominance concepts ratings between traits in the trustworthiness cluster and traits in the dominance cluster using a paired t-test. We found that the traits in the trustworthiness cluster were rated as significantly more similar to the trustworthiness concept and less similar to the dominance concept ($M = 5.08$, $SD = .71$) than traits in the dominance cluster ($M = 2.52$, $SD = .64$), $t(97) = 22.78$, $p < .001$, 95% CI [2.35, 2.79], Cohen's $d = 3.79$, validating the clusters found in Study 1 as trustworthiness and dominance clusters. One-sample t-tests show that both trustworthiness and dominance clusters were significantly different from the midpoint, $t(97) = 70.48$, $p < .001$, 95% CI [4.94, 5.23], Cohen's $d = 1.52$ and $t(97) = 38.91$, $p < .001$, 95% CI [2.39, 2.64], Cohen's $d = 2.32$, respectively. Individual traits also mostly confirmed the trustworthiness and dominance clusters found in Study 1, with the exception of the trait *confident*, which was rated as more similar to the dominance concept and the traits *weird* and *attractive*, which did not differ from the midpoint (Figure S1a).

Lastly, we also compared valence ratings between the two clusters using a paired t-test. We found that the traits in the trustworthiness cluster were rated as more positive ($M = 6.31$, $SD = .49$) than the traits in the dominance cluster ($M = 2.31$, $SD = .51$), $t(97) = 48.90$, $p = < .001$, 95% CI [3.83, 4.16], Cohen's $d = 8.06$, indicating that the clusters found in Study 1 partly overlap with valence (i.e., positive vs. negative trait clusters). One-sample t-tests show that both trustworthiness and dominance clusters were significantly different from the midpoint, $t(97) = 128.76$, $p < .001$, 95% CI [6.21, 6.41], Cohen's $d = 4.76$ and $t(97) = 45.29$, $p < .001$, 95% CI [2.21, 2.42], Cohen's $d = 3.33$, respectively. Individual traits also confirmed positive vs. negative divide with all traits being significantly different from the midpoint (Figure S1b).

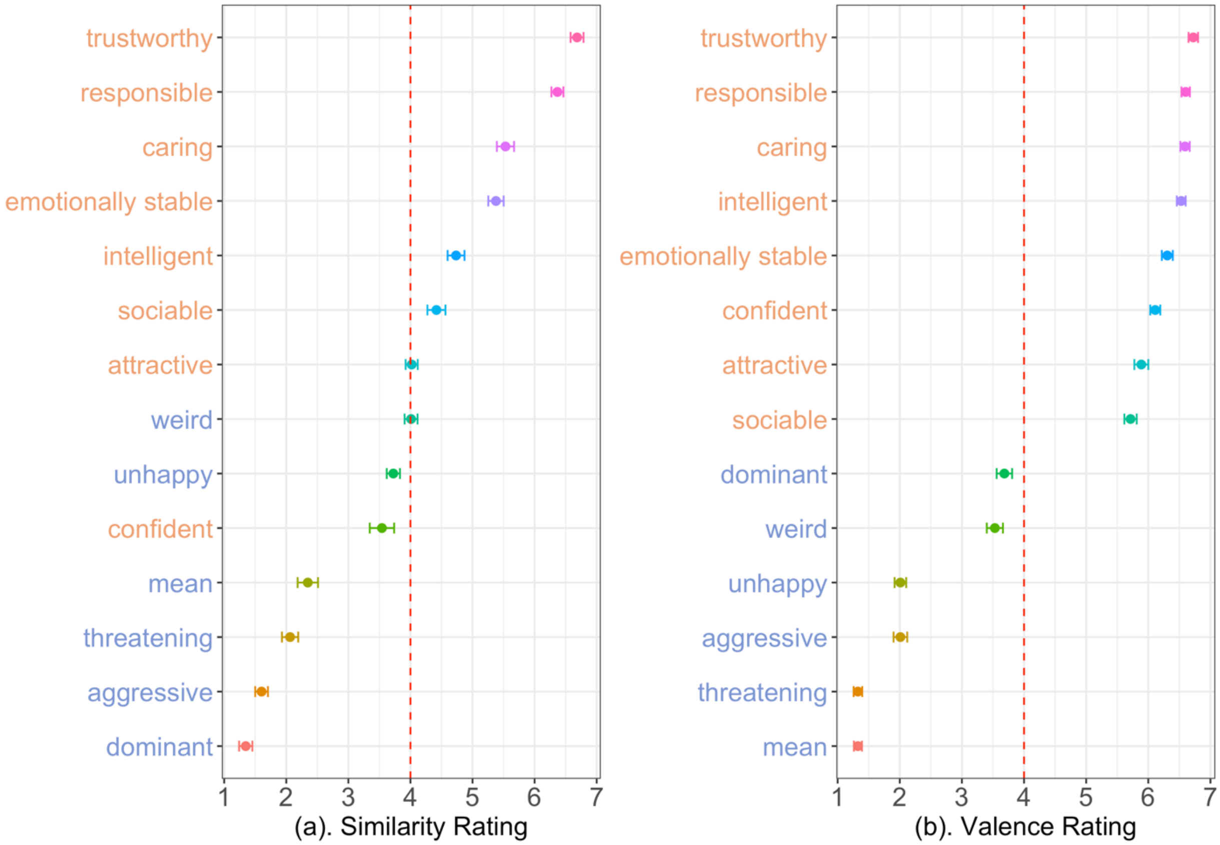

Figure S1. Individual trait ratings for (a) relation to the trustworthiness vs. dominance concept and (b) valence. The dotted red lines indicate the midpoint (4). Traits denoted in red were in the trustworthiness cluster and traits denoted in blue were in the dominance cluster found in Study 1.

Analysis S5. Can ingroup favoritism explain cross-cluster dissimilarity? (Study 1)

In this analysis, we aimed to alleviate the possibility that the more negative relationship between trustworthiness and dominance for the outgroup compared to ingroup could be explained by mere ingroup favoritism (i.e., that a more favorable perception of ingroup compared to outgroup could explain the differential relationships between trustworthiness and dominance across ingroup and outgroup). Specifically, we examined whether the overall favorable ratings of ingroup faces compared to outgroup faces would be related to overall more dissimilar cross-cluster ratings.

We computed ingroup favoritism scores based on how each participant from Study 1 rated ingroup faces relative to outgroup faces by subtracting the average outgroup ratings from the average ingroup ratings for each participant. Before doing so, however, we assigned different weights to different traits based on the valence ratings mentioned earlier. Specifically, we rescaled the valence ratings from Analysis S4 from a range of 1 to 7 to -3 to 3, where negative ratings indicated more negative traits and positive ratings indicated more positive traits. We then calculated the average ratings across participants. These means were used to weight the face ratings and compute ingroup favoritism scores, with traits that were either more positive or more negative contributing more to the scores than more neutral traits. By rescaling the ratings to -3 to 3, higher ingroup favoritism scores always indicated more favoritism for ingroup.

A one-sample t-test on the weighted face ratings data revealed that participants in Study 1 showed ingroup favoritism ($M = 0.28$, $SD = 1.19$) that was significantly greater than zero (no difference between ingroup and outgroup), $t(256) = 3.80$, 95% CI [0.14, .43], $p < .001$, Cohen's $d = 0.24$.

Next, we computed relative cross-cluster dissimilarity scores for each participant from Study 1 for both facial impression and stereotype association by subtracting the ingroup dissimilarity score from the outgroup dissimilarity score (refer to the Results section of Study 1 in the main text). For facial impression, we subtracted correlation values from 1 so that higher scores indicated more dissimilarity. We found that the Spearman rank-order correlation between ingroup favoritism and cross-cluster facial impression dissimilarity was not significant, $\rho(248) = .07$, 95% CI = [-.06, .20], $p = .25$. The Spearman rank-order correlation between ingroup favoritism and cross-cluster stereotype association dissimilarity was also not significant, $\rho(248) = .08$, 95% CI = [-.06, .21], $p = .26$ (Figure S2).

To corroborate this analysis without weighting, we again computed ingroup favoritism scores by simply changing the signs of the ratings based on valence. That is, positive traits, which have valence ratings above the midpoint of 0, were multiplied by 1, whereas negative traits, which have valence ratings below the midpoint of 0, were multiplied by -1. A one-sample t-test again confirmed that ingroup favoritism was significantly greater than zero ($M = 0.13$, $SD = .55$), $t(256) = 3.66$, 95% CI [.06, .19], $p < .001$, Cohen's $d = 0.23$. Furthermore, such non-weighted ingroup favoritism scores were again not related to cross-cluster dissimilarity for facial impressions, $\rho(248) = .05$, 95% CI = [-.07, .17], $p = .44$, or for stereotype associations, $\rho(248) = .05$, 95% CI = [-.07, .18], $p = .40$.

Together, these results indicate that individuals who showed more favoritism towards the ingroup compared to the outgroup were not more likely to rate traits in the trustworthy and dominant clusters as more dissimilar for the outgroup compared to ingroup. As ingroup favoritism showed no relationship with shifts between trustworthiness and dominance, this analysis suggests these are independent phenomena.

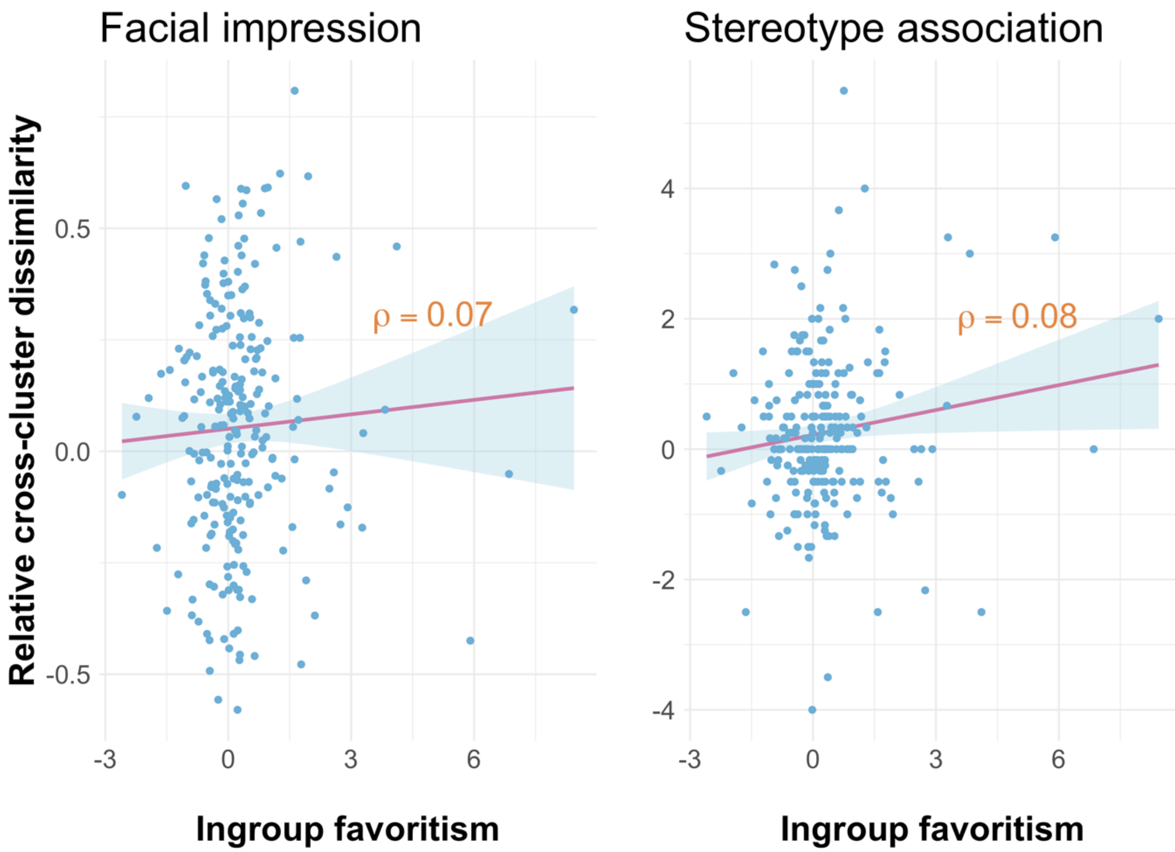

Figure S2. The relationships between ingroup favoritism and cross-cluster dissimilarities. Ingroup favoritism was not found to be related to the extent to which traits in the two clusters were rated as more dissimilar for outgroup compared to ingroup.

Analysis S6. The directionality of the shifts in person perception structures (Study 1 & 2)

In this analysis, we examined whether changes in perception of trustworthiness, dominance, or both drove the shifts in person perception structures across group boundaries found in our research. Due to the distinct designs of Study 1 and Study 2, we conducted separate analyses for each study.

Study 1. For Study 1, we focused on the correlation between ingroup and outgroup face ratings of traits in the trustworthiness cluster (referred to as the trustworthiness correlation) and compared it to the correlation between ingroup and outgroup face ratings of traits in the dominance cluster (referred to as the dominance correlation). By comparing these correlations, we could determine whether changes in trustworthiness or dominance ratings played a more significant role in the shifts observed in facial impression structures. Using aggregated data from Analysis S1, we found that both the trustworthiness correlation, $r(78) = .90$, 95% CI = [.85, .94], $p < .001$, and the dominance correlation, $r(78) = .90$, 95% CI = [.84, .93], $p < .001$, were highly significant. This indicates a high level of consensus in trustworthiness and dominance judgments across both ingroup and outgroup conditions. To assess if the two correlations significantly differed from each other, we employed Fisher's z transformation. The results indicated that they were not significantly different, with $z = .26$, $p = .40$, suggesting that the shifts in facial impression structures observed in Study 1 were not specifically driven by one trait over the other.

Study 2. In Study 2, we examined the bidirectional effects of changes in one trait on the perception of the other trait. Specifically, we manipulated trustworthiness and assessed its effect on dominance judgments, and vice versa. Our focus was on investigating whether the difference in the relationship between trustworthiness and dominance between ingroup and outgroup varied depending on the direction of manipulation (i.e., manipulated trustworthiness and measured dominance vs. manipulated dominance and measured trustworthiness). To simplify the analysis, we computed correlations between the manipulated dimension and the measured dimension for ingroup and outgroup faces separately for each participant. Next, we used a linear mixed-effects model to predict the correlation between the manipulated dimension and the measured dimension using manipulated trait (trustworthiness or dominance), faces' group membership (ingroup or outgroup), and their interaction. The model accounted for random intercepts and slopes for group membership across participants. The results revealed a significant main effect of the manipulated trait, suggesting that there was an overall more negative relationship between the two dimensions when trustworthiness was manipulated ($M = -.34$, $SD = .28$) compared to when dominance was manipulated ($M = .01$, $SE = .26$), $b = .33$, $\beta = .51$, $z = 15.68$, 95% CI [.29, .37], $p < .001$. We also found a significant main effect of group membership, indicating that the relationship between the two dimensions was more negative for outgroup ($M = -.18$, $SD = .32$) compared to the ingroup ($M = -.15$, $SD = .33$), $b = .06$, $\beta = .09$, $z = 2.82$, 95% CI [.02, .10], $p = .005$. However, the interaction between manipulated trait and group membership was not significant, with $b = .05$, $\beta = .07$, $z = 1.71$, 95% CI [-.01, .11], $p = .09$. This indicates that we did not find evidence that the difference in the relationship between trustworthiness and dominance across the ingroup and outgroup was driven by one specific trait over the other.

Next, we examined the directionality of the effects on stereotype association ratings. Since we presented two questions each for ingroup and outgroup (e.g., "How likely is a dominant *Klee* person to also be trustworthy?" and "How likely is a trustworthy *Klee* person to also be dominant?"), we used a linear mixed-effects model to predict stereotype association dissimilarity ratings (reverse-coded) using group membership (ingroup or outgroup), direction (trustworthy-also dominant or dominant-also trustworthy), and their interaction. The model allowed random intercepts and slopes for group membership across participants. The results indicated that the main effect of direction was not significant, $b = .04$, $\beta = .01$, $z = .49$, 95% CI [-.10, .17], $p = .62$. However, the main effect of group membership was significant, suggesting that trustworthiness and dominance were perceived as more dissimilar for outgroup ($M = 4.17$, $SD = 1.21$) compared to ingroup ($M = 3.70$, $SD = 1.21$), $b = .50$, $\beta = .20$, $z = 5.86$, 95% CI [.33, .67], $p < .001$. The interaction between group membership and direction was not significant, $b = .07$, $\beta = .02$, $z = .66$, 95% CI [-.13, .26], $p = .51$. Similar to the face ratings, these findings indicate that the difference in the relationship between trustworthiness and dominance across the ingroup and outgroup was not driven by one specific trait.

Analysis S7. Manipulation check (Study 2)

Following the tasks, Study 2 included the identical familiarity, agreement with group assignment, and collective identification measures as in Study 1 (see Analysis S2), with one exception. For the three-item collective identification with the ingroup measure used in Study 1, we added three corresponding items for collective identification with the outgroup, enabling a comparison of collective identification scores between ingroup and outgroup.

Unlike in Study 1, a paired t-test found that participants were somewhat more familiar with Kandinsky and his paintings ($M = 4.57$, $SD = 3.55$) than with Klee and his paintings ($M = 4.06$, $SD = 3.16$), $t(255) = 3.08$, $p = .002$, 95% CI [.18, .83], Cohen's $d = .15$. However, for both painters, the most frequent response was "not at all familiar" (Klee – 149 people, Kandinsky – 134 people out of 256 people).

Consistent with a successful manipulation, a paired t-test revealed that participants agreed significantly more with a preference for ingroup paintings ($M = 5.55$, $SD = 1.38$) compared to outgroup paintings ($M = 2.23$, $SD = 1.38$), $t(255) = 23.22$, $p < .001$, 95% CI [3.03, 3.59], Cohen's $d = 2.40$. There was no overall difference in the strength of agreement between the Klee ($M = 3.87$, $SD = 2.15$) and Kandinsky ($M = 3.91$, $SD = 2.17$) groups, $t(255) = .19$, $p = .85$, 95% CI [-.45, .54], Cohen's $d = .02$.

Consistent with a successful manipulation, a paired t-test demonstrated significantly greater collective identification with the ingroup ($M = 10.43$, $SD = 4.65$) compared to the outgroup ($M = 6.41$, $SD = 3.88$), $t(255) = 12.23$, $p < .001$, 95% CI [3.37, 4.67], Cohen's $d = .94$. There was no overall difference in strength of collective identification between the Klee ($M = 8.21$, $SD = 4.53$) and Kandinsky ($M = 8.64$, $SD = 4.91$) groups, $t(255) = .57$, $p = .29$, 95% CI [-.38, 1.25], Cohen's $d = .09$.

Table S2. Individual trait pair dissimilarity scores comparing the ingroup and outgroup. A higher score indicates greater dissimilarity between traits. Cross-cluster trait pairs are denoted in green, intra-trustworthy cluster trait pairs in red, and intra-dominant cluster trait pairs in blue. Trait pairs that are descriptively more dissimilar are shown in boldface.

Pair	FI (in)	FI (out)	SA (in)	SA (out)	Pair	FI (in)	FI (out)	SA (in)	SA (out)	Pair	FI (in)	FI (out)	SA (in)	SA (out)	Pair	FI (in)	FI (out)	SA (in)	SA (out)
agg att	1.00	**1.14**	7.29	**7.61**	emo mea	1.33	**1.35**	**8.85**	8.77	car emo	**0.57**	0.46	4.71	**5.20**	dom wei	**1.04**	1.01	**7.88**	7.00
agg car	1.39	**1.48**	9.21	**9.53**	emo thr	1.34	**1.36**	9.75	**9.94**	car int	0.72	**0.73**	4.55	**6.20**	mea thr	0.46	**0.50**	**5.86**	4.76
agg con	**0.91**	0.88	6.32	**6.65**	emo unh	**1.42**	1.37	**8.86**	8.72	car res	**0.58**	0.50	4.79	**5.36**	mea unh	0.70	**0.71**	**6.08**	4.92
agg emo	1.25	**1.32**	10.20	**10.44**	emo wei	1.31	**1.34**	**8.35**	8.19	car soc	0.52	0.52	4.52	**5.23**	mea wei	0.77	**0.78**	**7.73**	7.27
agg int	1.08	**1.16**	7.50	**8.08**	int mea	**1.17**	1.03	7.59	**8.35**	car tru	**0.64**	0.61	4.54	**5.42**	thr unh	**0.71**	0.57	**6.45**	5.45
agg res	1.28	**1.30**	7.83	**8.33**	int thr	**1.34**	1.29	**8.08**	7.76	con emo	**0.74**	0.59	4.85	**5.50**	thr wei	0.66	**0.69**	**6.88**	6.14
agg soc	**1.25**	1.24	8.28	**8.31**	int unh	1.31	**1.38**	**7.62**	7.50	con int	**0.73**	0.71	**5.88**	5.38	unh wei	0.86	**0.90**	6.90	**7.24**
agg tru	1.29	**1.31**	9.11	**9.26**	int wei	**1.26**	1.23	6.12	**6.60**	con res	0.67	**0.68**	4.81	**5.22**					
att dom	0.83	**0.99**	6.48	**6.84**	mea res	1.25	**1.30**	8.48	**8.89**	con soc	0.61	0.61	**5.00**	4.73					
att mea	**1.16**	1.15	8.09	8.09	mea soc	1.12	**1.25**	**8.91**	8.30	con tru	0.75	**0.87**	**5.43**	5.39					
att thr	1.25	**1.32**	8.35	8.35	mea tru	**1.41**	1.38	9.65	**9.70**	emo int	**0.64**	0.55	5.11	**6.00**					
att unh	1.09	**1.30**	8.50	**8.92**	res thr	**1.25**	1.19	9.76	**10.14**	emo res	0.64	**0.69**	5.19	**5.61**					
att wei	1.32	1.32	**7.45**	7.24	res unh	1.33	**1.40**	8.07	**8.38**	emo soc	**0.63**	0.59	5.27	**5.45**					
car dom	1.09	**1.15**	7.38	**8.16**	res wei	**1.18**	1.17	6.63	**7.47**	emo tru	**0.62**	0.56	5.15	**5.35**					
car mea	1.35	**1.40**	9.82	**10.21**	soc thr	1.35	**1.39**	8.64	**9.28**	int res	0.56	0.56	5.41	**5.73**					
car thr	1.43	**1.52**	10.00	**10.28**	soc unh	1.32	**1.43**	**9.72**	8.93	int soc	**0.74**	0.61	5.95	**6.77**					
car unh	1.40	**1.53**	7.97	**8.68**	soc wei	**1.40**	1.31	**8.04**	7.29	int tru	0.67	**0.72**	4.39	**6.26**					
car wei	**1.27**	1.24	6.76	**6.85**	thr tru	**1.46**	1.45	10.36	**10.47**	res soc	0.65	0.65	5.90	**6.17**					
con dom	0.63	**0.84**	5.22	**5.48**	tru unh	1.21	**1.23**	**8.55**	7.97	res tru	0.57	**0.59**	4.19	**4.58**					
con mea	0.92	**1.01**	6.88	**7.19**	tru wei	1.28	**1.29**	6.26	**6.94**	soc tru	0.60	**0.62**	5.18	**5.68**					
con thr	1.05	1.05	6.68	**7.18**	att car	**0.76**	0.57	6.08	**6.64**	agg dom	**0.64**	0.57	**5.41**	5.00					
con unh	1.26	**1.27**	8.89	**9.07**	att con	0.54	**0.58**	4.19	**4.74**	agg mea	**0.59**	0.56	**5.29**	4.97					
con wei	1.24	**1.25**	7.29	**7.93**	att emo	**0.61**	0.52	5.41	**5.96**	agg thr	**0.47**	0.41	**5.10**	4.03					
dom emo	**1.03**	1.01	6.83	**7.04**	att int	0.60	**0.64**	5.75	**6.80**	agg unh	**0.72**	0.65	**6.00**	5.11					
dom int	0.85	**0.98**	6.09	**7.14**	att res	**0.69**	0.68	5.79	**6.44**	agg wei	0.82	**0.91**	7.00	**7.16**					
dom res	0.97	**1.03**	5.61	**6.32**	att soc	0.56	0.56	4.89	**5.22**	dom mea	**0.76**	0.62	**6.07**	5.85					
dom soc	0.98	**1.01**	**7.26**	6.97	att tru	**0.65**	0.57	5.62	**5.90**	dom thr	**0.77**	0.65	**5.65**	4.79					
dom tru	1.03	**1.12**	6.80	**8.33**	car con	0.71	**0.74**	5.85	**6.77**	dom unh	**1.00**	0.89	**7.97**	6.89					

Legends: FI = facial impression (1-r), SA = stereotype association, in = ingroup, out = outgroup, agg = aggressive, att= attractive, car = caring, con = confident, dom =dominant, emo = emotionally stable, int =intelligent, mea = mean, res = responsible, soc = sociable, thr = threatening, tru = trustworthy, unh = unhappy, wei = weird

For cross-cluster trait pairs, 67 trait pairs were more dissimilar for outgroup than ingroup (32 FI, 35 SA) compared to 29 pairs that were more dissimilar for ingroup than outgroup (16 FI, 13 SA). On the other hand, for intra-cluster trait pairs, 43 trait pairs were more dissimilar for outgroup than ingroup (16 FI, 27 SA) and 43 pairs were more dissimilar for ingroup than outgroup (27 FI, 16 SA). A chi-square test of independence showed that the number of dissimilar trait pairs across ingroup and outgroup differed between intra vs. cross-cluster trait pairs, χ^2 (1) = 6.63, p = .01.

References

Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin, 130*, 80-114.

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.

Van Bavel, J.J., & Cunningham, W.A. (2012). A social identity approach to person memory: Group membership, collective identification, and social role shape attention and memory. *Personality and Social Psychology Bulletin, 38*, 1566-1578.