Supplemental Online Material

Stimuli

Race: White vs. Black Male Faces. In the categorization and the conformity task, photos of 20 black and 20 white male faces were used. These photos were selected on the basis of a pre-test in which participants saw 64 male faces (half of which of white-Caucasian) with neutral facial expression (downloaded from on-line databases, Minear & Park, 2004; Payne, 2001; Tarrés & Rama), after they were cropped for showing only the central features and equally resized (275 X 395 pixels). The 20 photos of white and the photos of 20 black faces used in the experiment were evaluated on average as equally pleasant (on a five-points scale, 1 = not al all; 5 = completely) (M_white = 2.32, SD = .44; M_black = 2.31, SD = .48) $t(20) = .10$, $p = .92$, although the firsts were judged (on a five-points scale, 1 = certainly Italian; 5 = certainly foreigner) as more likely Italian (M = 2.52, SD = .41) whereas the latters as foreigner (M = 4.50, SD = .60), $t(20) = 10.54$ $p < .001$.

Accent: native vs. nonnative Italian. In the categorization and in the conformity task, 10 voices speaking with a native Italian and 10 with a non-native Italian accent pronouncing two phrases each – were used. The selection process was as it follows. We contacted 19 Italian native and 14 non-native Italian speakers (all males, mainly from West Africa) and gained consented from them to be audio-registered while reading 6 short phrases. All the obtained audio-clips were then edited by Audacity® software (cut so as the phrase started immediately, noises were filtered and volume normalized) and then presented to 10 participants (5 female, age M = 21.50, SD = 1.08) in a pre-test conducted in one of our university lab. Participants listened to the audio-clips and - in a first task - indicated whether the speaker was Italian vs. Foreigner, by pressing D or K (counterbalanced across participants) on a QWERTY keyboard. Participants were asked to be accurate but also fast and therefore not to wait until the end of the phrase to respond. This task was implemented using Inquisit Milliseconds® software. Each speaker was pronouncing 6 different short phrases each. In a second task, participants rated voice' pleasantness while pronouncing a phrase with no relational

content ("Scusa, che ore sono?/Excuse me, what's time is it?"), (from 1 = not at all to 5 = completely). For each speaker, we selected two phrases (1: "Andiamo a bere un caffè/let's have some coffee"; 2. “Piacere di conoscerti/nice to meet you”) that could be used in a positive social interaction. Categorization and reaction time responses for these phrases were separately entered in a 2 (Accent: native vs. non-native) X 2 (Phrase: 1 vs. 2) within-subjects ANOVA. The (10) native and (10) non-native speakers were correctly recognized as native (M = 90.00%, SD = 10.27) or non-native Italian (M = 88.61%, SD = 9.44), p = .70 (all other effects p > .15) and with the same speed (M_nat = 1026.86 ms, SD = 152.56) (M_non-nat = 1035.80, SD = 166.79), p = .80 (all other effects p > .32). Moreover, the 10 native (M = 3.03, SD = 0.60) and 10 non-native speakers pleasantness (M = 2.62, SD = 0.58) did not differed from a medium point (3), t(9) = .16, p = .88 and t(9) = -2.08, p = .07 respectively. Finally, on an independent pre-test (N = 23, 12 females, average age = 31.57, SD = 3.40), the two phrases were judged as equally positive (on five-points scale, 1 = completely negative, 5 = completely positive), M = 5.30, SD = 1.15; M = 5.30, SD = .88, respectively, t(22) = .00, p = n.s..

Questionnaire Scale.

Language and race centrality in the conception of national identity. Based on Sloman and colleagues (1998) work, we assessed centrality with 5 items for language and 5 for race (i.e. How surprised would you be to encounter…; How easily can you imagine…; How representative is…; How similar to the typical Italian is… an Italian who does not speak Italian/is not white?; What percentage of Italians speak Italian/is white? 1= less than 10%; 7 = 100%).

Social identification. Participants rated their identification to the national ingroup (i.e. To what extent do you feel part of the Italian group? Postmes, Haslam, & Jans, 2012) and their self-investment in this group (3 items, to what extent …do you like to be Italian? …do you feel a bond with Italians? .. being Italian is an important part of your identity?; Leach, et al. 2008). An additional item directly tapped whether language played a central role in their own ingroup

identification (i.e. To what extent is speaking Italian an important part of your Italian identity?). Responses were registered on a 5-points scale (1 = not at all to 5 = completely).

The questionnaire included other questions among which a self-categorization pictorial item and an Inclusion of the other in the self pictorial scale on which participants rated the closeness to the first generation Italian they know the best. Given that the results for these questions were similar to those obtained for similar measures, these results were not reported.

Analysis including Study as factor: Results

Preliminary analyses were conducted including Study (1a vs.1b) as a between-participants factor in the Race X Accent ANOVA. For the outcome of the categorization, the Study factor was involved only in a marginal three way interaction, F (1, 100) = 3.75, p = .06, $\eta 2_p$ = .04. Importantly, the Race X Accent ANOVA conducted separately for Study 1 and Study 2 yielded a similar pattern of results. (Main effect of Accent: Study 1a - F (1, 41) = 506,49, p < .001, $\eta 2_p$ = .92, Study 1b - F (1, 58) = 410,81, p < .001, $\eta 2_p$ = .88; Main effect of Race: Study. 1a - F (1, 41) = 33.22, p < .001, $\eta 2_p$ = .44, Study 1b - F (1, 58) = 72.54, p < .001, $\eta 2_p$ = .56; Interaction effect Study 1a - F (1, 41) = 12.46, p = .001, $\eta 2_p$ = .23; Study 1b – F (1, 58) = 42.70, p < .001, $\eta 2_p$ = .42; in both studies all comparisons ps < .001). Means are shown in Table 1.

For the categorization process, we found a Study marginal main effect on initial RT, F (1, 75) = 3.38, p = .07, $\eta 2_p$ = .04, on RT, F (1, 75) = 18.26, p < .001, . $\eta 2_p$ = .20, and MD time, F (1, 75) = 21.72, p < .001, $\eta 2_p$ = .23. In general, participants were faster in starting the response movement in Study 1a (M = 225 ms, SD = 157) than 1b (M = 163 ms, SD = 138), but the reverse occurred in giving the final answer (1a: M = 1300 ms, SD = 216; 1b: M = 1529 ms, SD = 248). Also, the MD was reached earlier in Study 1a (M = 652 ms, SD = 127) than in Study 1b (M = 811 ms, SD = 166).

For the MD, the analyses yielded a significant main effect of the Study factor, F (1, 75) = 7.15, p = .009, $\eta 2_p$ = .09, Race X Accent interaction, F (1, 75) = 62.23, p < .001, $\eta 2_p$ = .45, and a three-way interaction, F (1, 75) = 7.29, p = .009, $\eta 2_p$ = .09. Importantly, in both studies, when analyzed separately, the Race X Accent interaction effect was significant (Exp. 1a: F (1, 34) = 54,52, p < .001, $\eta 2_p$ = .62; Exp. 1b: F (1, 41) = 14.14, p = .001, $\eta 2_p$ = .26). In Study 1a, all comparisons were significant, all ps < .001. In Study 1b, all comparison were significant, ps < . 007, with the exception of the comparison between White-Native and White-Nonnative that was marginally significant, p = .08 (see Table 1).

Table 1. Means and standard deviation (in brackets) for categorization as an ingroup member (Italian), and maximum deviation (MD) in function of the target accent and race in Study 1a and 1b, separately.

		Accent			
		Native		Nonnative	
	Race	Exp. 1a	Exp. 1b	Exp. 1a	Exp. 1b
Categorization	White	95% (7)	95% (11)	18% (19)	18% (17)
	Black	67% (31)	56% (36)	4% (7)	4% (8)
MD	White	.47 (.31)	.41 (.26)	.70 (.34)	.52 (.30)
	Black	.84 (.44)	.57 (.37)	.41 (.28)	.35 (.26)

Note. Means within the same column, and within the same line are different from each other for p <.05
Black-native and White-native targets: MD for response "Italian". Black-nonnative and White-nonnative targets: MD for response "foreigner".

RT and MD: ANOVA and correlations (N = 90)

For the RT and the MD responses, we focused the analysis on the response "Italian" for black-native accent and white-native accent trials, and for the response "Foreigner" for black-non native accent and white-non native accent trials. Twelve participants responded "Italian" (i.e., ingroup

member) to none of the 10 black-native accent trials; therefore they provided no data to be analyzed. Of the remaining 90 participants, 4 responded "Italian" just to 1 black-native accent trials, 6 participants 2 trials, and 3 participants 3 trials, the rest (N = 77) answered "Italian" to at least 4 or more black-native accent trials. Given that RT and MD calculated on one or few trails risks to provide an unreliable index for the variable of interest, we decide to focus our analysis only on participants who gave at least 40% (4 out of 10 trials) of answers according to the target's accent (i.e., the response "Italian" for Black-Native accent and White-Native accent trials and "Foreigner" for Black-Nonnative accent and White-Nonnative accent trials); the RT and MD indices were calculated on the responses to, at least, 4 trials. The choice of "4 trials" was a sort of compromise to not to consider in the analysis scores based on few trials and on the other hand not to exclude too many participants. In the main text we reported the results applying this restrictive criteria so to avoid the criticism that the results were influenced by an index based on few trials.

For completeness, we report here the analysis (ANOVA and correlation) conducted on 90 participants, that is all participants who provided at least 1 ingroup response to a black-native accent trial.

ANOVA

For the initial RT, we only found a main effect of Race, F (1, 89) = 9.29, p = .003, $\eta 2_p$ = .10, showing that the participants were slower in starting the response when they saw a Black (M = 190 ms, SD = 158) in comparison to a White face (M = 174, ms, SD = 145). All other ps >. 92. From the analysis of the final RTs, a significant interaction emerged, F (1, 89) = 113,96, p < .001, $\eta 2_p$ = .56, showing a longer RT in the categorization of targets with mixed cues compared to those with convergent (in terms of categorization) cues (all ps <. 001) (means are reported in Table 2). For the MD, the ANOVA yielded a significant main effect of Race, F (1, 89) = 6.40, p = .01, $\eta 2_p$ = .07, qualified by the significant interaction, F (1, 89) = 67.24, p < .001, $\eta 2_p$ = .43. All comparisons between the means (Tab.2) were significant, p < .001. Additional comparison between mixed cue

targets (i.e. Black- Native accent and White-Nonnative accent) put in evidence that the MD, although generally high, were higher when categorizing Black faces with a native accent than White faces with a non-native accent, $t(89) = 2.43$, $p = .02$. The same comparison for cue convergent trials (i.e. Black- Non Native accent and White-Native accent) showed that the MD were somehow lower for Black-faces with a nonnative accent than White-faces with a native accent, $t(89) = 1.86$, $p = .07$.

Analyzing the MD time we found an interaction effect, $F(1, 89) = 80.48$, $p < .001$, $\eta 2_p = .48$. Conflict between the activated categories was solved faster in the case of cue congruent than mixed cues trials, $p < .001$. Further comparison yield no statistical significant difference neither between mixed trials, $t(89) = .06$, $p = .96$, nor between congruent trials $t(89) = .63$, $p = .53$.

Table 2. Means and standard deviation (in brackets) for reaction time (RT), maximum deviation (MD) and maximum deviation time (MD Time) in function of the target accent and race.

		Accent	
		Native	Non-native
	Race		
RT	White	1341 ms (260)	1507 ms (310)
	Black	1484 ms (275)	1371 ms (311)
MD	White	.44 (.29)	.60 (.33)
	Black	.69 (.42)	.37 (.27)
MD time	White	688 ms (180)	787 ms (192)
	Black	786 ms (200)	679 ms (204)

Note. Means within the same column, and within the same line are different from each other.
Black-native and white-native targets: RT and MD for response "Italian". Black-non native and white-non native targets: RT and MD for response "Foreigner".

Correlations for the full sample are reported in Table 3. A larger MD—a greater parallel activation of in-group and out-group categories—for Black-Native accent targets was positively linked with the diagnosticity of ethnic appearance (i.e., being White) and the prejudice toward

immigrants and negatively related to the importance attributed to contact with first-generation Italians. Note that this last effect was in this sample only marginal significant.

To conclude, the analyses conducted on the full (N = 90) and the restricted sample (N = 77) yielded similar results except for the negative correlation between larger MD and importance attributed to the personal experiences with first-generation Italians that in the latter was significant and the first marginal significant.

Table 3. Bivariate correlations between MD and questionnaire measures (N = 85-90). MD was computed on the accent-based responses ("Italian" for black-native accent and white-native accent targets, and the RT for the response "Foreigner" for black-non native accent and white- non native accent targets).

	White-native accent	Black-native accent	White-non native accent	Black-non native accent
1. Social identification	.02	.02	-.06	-.07
2. Language role in social identification	-.07	-.14	-.11	-.05
3. Language centrality	.11	.02	.08	.08
4. Language diagnosticity	-.04	.08	.02	-.08
5.Language importance	.01	-.12	-.10	.003
6. Ethnic appearance centrality	.05	.07	.08	.07
7. Ethnic appearance diagnosticity	.12	.20*	.15	.09
8. Ethnic appearance importance	.02	.01	.09	.05
9. Prejudice	.27**	.23*	.21*	.02
10. Social dominance orientation	.09	.06	.21*	.20 ††
11. Frequency of contact	.003	-.08	.02	-.02
12. Importance of contact	-.07	-.20†	-.14	-.21 ††

† p = .07, †† p = .06, p =< .05, ** p =< .01, *** p < .001

References

Leach, C. W., Van Zomeren, M., Zebel, S., Vliek, M. L., Pennekamp, S. F., Doosje, B., W. Ouwerkerk, & Spears, R. (2008). Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*, 144 – 165. http://dx.doi.org/10.1037/0022-3514.95.1.144

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, and Computers*, *36*, 630-633. http://dx.doi.org/10.3758/BF03206543

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181. http://dx.doi.org/10.1037/0022-3514.81.2.181

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology*, *52*. 597-617. https://doi.org/10.1111/bjso.12006

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189-228. https://doi.org/10.1207/s15516709cog2202_2

Tarrés, F. & Rama A. *GTAV Face Database*. http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm