

Supplementary Information for

**Impression formation through social interaction:  
Effects of ethnicity in the Dutch context**

Study 1 .....	1
Method.....	1
Deviations from the preregistration.....	2
Participants .....	2
Social reinforcement learning task .....	2
Group identification.....	3
Results .....	3
Computational modeling .....	3
Gender effects.....	4
Study 2.....	4
Methods .....	4
Results .....	5
Computational modeling .....	5
Gender effects.....	5
Study 3.....	6
IAT block order .....	6
Results .....	7
Computational modeling .....	7
References .....	7

**Study 1**

**Method**

## Deviations from the preregistration

In the preregistration, we stated that we would test the effects of ethnicity on instrumentally learned preferences using multiple computational models. Although we did this, we report our main results using logistic regressions, as these provided more direct, conventional, and interpretable tests of our main hypothesis regarding the effect of race on instrumental learning.

## Participants

Data were collected from an additional 15 respondents who were ineligible (did not identify as *only* Dutch, or indicated that they were not born in the Netherlands), but could not be pre-screened.

## Social reinforcement learning task

*Stimuli.* The players that we selected from the Amsterdam Dynamic Facial Expression Set (van der Schalk et al. 2011) for the main learning task are in Table s1.

**Table s1**

*Model numbers of the faces that were selected for the learning task*

	<b>Moroccan players</b>	<b>White players</b>
<b>Female players</b>	F06	F01
	F08	F02
	F09	F03
	F10	F05
<b>Male players</b>	M05	M02
	M07	M03
	M09	M04
	M10	M12

Except for the F10, who had a Turkish background, all faces had a Moroccan ethnic background. We chose F10 nonetheless, since the image fit well with the other images. All

faces had an expression of joy.

### **Group identification**

Participants' identification with being Dutch was measured using an adapted group identification measure (Doosje et al., 1995). The measure consisted of four items (I identify with other Dutch people, I see myself as a Dutch person, I am glad to be a Dutch person, I feel strong ties with Dutch people), and participants answered on a scale of 1 (do not agree at all) to 7 (completely agree).

## **Results**

### **Computational modeling**

We investigated whether the model-derived estimates of the prior and learning rate parameters varied based on the participants' ethnicity. Learning rates did not differ for White ( $M = 0.145$ ,  $SD = 0.26$ ) compared to Moroccan players ( $M = 0.147$ ,  $SD = 0.25$ ),  $Z = -0.50$ ,  $p = 0.618$ , and the prior ( $M = -0.35$ ,  $SD = 17.91$ ) did not differ from zero,  $Z = -1.16$ ,  $p = 0.248$ . Despite the lack of average differences in these parameters, the fact that the prior + learning model is the best fit indicates that a substantial part of the participants based choices in the learning task based on group-based initial expectancies, and separate learning rates.

Neither the learning rate difference nor prior estimates differed as a function of EMS—a pattern that again reflects heterogeneity in these parameters across participants despite the best fit of behavior to the prior + learning model. Also, despite finding that high EMS participants showed a stronger pro-Moroccan choice preference the fit advantage of the best fitting prior + learning model was not better for these participants, indicated by the fact that EMS and AIC model fit were not correlated,  $t(63) = 1.39$ ,  $p = .168$ ,  $r = 0.17$ , 95% CI = [-0.07, 0.40].

## **Gender effects**

The gender of players was counterbalanced in this study, such that participants interacted with only male or only female partners. To test whether gender moderated the effect of ethnicity on choice, we conducted a logistic regression containing choice behavior as the outcome, relative reward rate and ethnicity as predictors and random slopes grouped within participants, and an Ethnicity x Player Gender interaction. This analysis produced a significant Ethnicity x Player Gender interaction,  $OR = 0.38$ ,  $95\% CI = [0.18, 0.81]$ ,  $p = .013$ : participants' preference for Moroccans over White partners was significant when players were male,  $\beta = 1.14$ ,  $t = 4.65$ ,  $p < .001$ , but there was no effect of ethnicity when players were female,  $\beta = 0.17$ ,  $t = 0.60$ ,  $p = 0.547$ .

## **Study 2**

### **Methods**

#### **Questions regarding the participant's decisions during the experiment**

In order to find out more about the salience of ethnicity in the participants' mind during the experiment, we asked them four questions. With the first question we inferred whether they had noticed ethnicity: Did you notice during the experiment that you were shown faces with different ethnicities? on which they could give a yes or no answer. If participants answered yes we gave them two follow-up questions: When did you notice this? (with possible answers: after the first few trials, during the second block, not until the end of the experiment, or other, which contained an open question), and Did you feel like this realization influenced your decision-making during the experiment? (with possible answers: yes, or no). Finally, we asked them How did this influence you? with an open answer.

## Results

### Computational modeling

As in Study 1, learning rates for White ( $M = 0.16$ ,  $SD = 0.25$ ) or Moroccan players ( $M = 0.19$ ,  $SD = 0.32$ ) did not differ significantly,  $Z = -0.06$ ,  $p = 0.954$ , but in contrast to Study 1, participants started with a prior that was significantly different from zero, such that they started with an initial choice preference for the Moroccan players ( $M = -7.62$ ,  $SD = 24.76$ ),  $Z = -2.98$ ,  $p = 0.003$ ).

Participants with a strong Moroccan choice bias were a substantially better fit to the prior + learning model, while participants lacking a pronounced Moroccan choice bias displayed a comparatively lesser degree of model fit (OR = 0.13, 95% CI = [0.08, 0.21],  $p < .001$ ), indicating that the computational model was able to successfully capture the observed behavioral outcomes.

These results combined, again suggest that participants acquired and maintained a group bias through a combination of group-based initial expectancies and the updating of separate representations for Moroccan and White players. The computational model was able to successfully capture the observed behavioral outcomes, as shown by the substantially better fit of participants with a strong Moroccan choice bias to the prior + learning model, while participants lacking a pronounced Moroccan choice bias displayed a comparatively lesser degree of model fit (OR = 0.13, 95% CI = [0.08, 0.21],  $p < .001$ ).

### Gender effects

In contrast to the findings in Study 1, there was no interaction effect between Ethnicity \* Gender, OR = 0.78, 95% CI = [0.31, 1.97],  $p = .599$ : participants preferred interactions with Moroccans over White partners for both genders.

## Study 3

### IAT block order

**Table 3**

*Sequence of blocks of the IAT: order 1.*

Block	No. of trials	Function	Items assigned to Left-key response	Items assigned to Right-key response
1	20	Practice	Moroccan faces	White faces
2	20	Practice	Pleasant words	Unpleasant words
3	20	Test	Pleasant + Moroc	Unpleasant + White
4	40	Test	Pleasant + Moroc	Unpleasant + White
5	20	Practice	White faces	Moroccan faces
6	20	Test	Pleasant + White	Unpleasant + Moroc
7	40	Test	Pleasant + White	Unpleasant + Moroc

*Note.* Half of the participants were assigned order 2 where blocks 1, 3, and 4 were switched with blocks 5, 6, and 7.

### IAT word list

*Positive words:*

Charmant, elegant, geslaagd, glorie, hoop, knap, lief, slim, vredig, wijsheid, begaafd, beloning, cadeau, creatief, feest, geluk, glimlach, grappig, handig, ideaal, klasse, leuk, opbeuren, paradijs, plezierig, respect, succes, troost, vriend, zonnig, aardig, briljant, geliefd, geschenk, gezellig, hemels, kameraad, luxe, perfect, vreugde

*Negative words:*

Agressief, dreigend, falen, giftig, hopeloos, leugenaar, noodlot, ramp, schade, vals, begaafd, beloning, cadeau, creatief, feest, geluk, glimlach, grappig, handig, ideaal, klasse, leuk, opbeuren, paradijs, plezierig, respect, succes, troost, vriend, zonnig, aardig, briljant, geliefd, geschenk, gezellig, hemels, kameraad, luxe, perfect, vreugde, argwanend, bedrog, berucht, dief, drama, fout, grauw, idioot, kwaal, leugen, monster, onrust, plaag, satan schande, slecht,

stress, verraad, vijandig, wraak, afkeer, beroerd, crisis, fataal, huivering, lelijk, paniek, razernij, smerig, woest

## Results

### Computational modeling

As in Study 1 and Study 2, there was no difference between the learning rates for White ( $M = 0.18$ ,  $SD = 0.26$ ) or Moroccan players ( $M = 0.24$ ,  $SD = 0.34$ ),  $Z = -0.73$ ,  $p = 0.46$ . As in Study 2, participants showed a prior significantly different from zero, such that they started with an initial choice preference for the Moroccan players ( $M = -5.54$ ,  $SD = 30.80$ ),  $Z = -2.34$ ,  $p = .019$ ). Participants with more pro-Moroccan choice preference, also were a better fit with the prior + learning model (AIC x Ethnicity effect on choice: OR = 2.99, 95% CI = [1.72, 5.18],  $p < .001$ ).

## References

Schalk, Job van der, Skyler T. Hawk, Agneta H. Fischer, and Bertjan Doosje. 2011. Moving Faces, Looking Places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, 11(4): 907–20. <https://doi.org/10.1037/a0023853>.