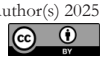
Group Processes & Intergroup Relations
GPIR

Article

Impression formation through social interaction: The effect of ethnicity in the Dutch context

Group Processes & Intergroup Relations
1–25



DOI: 10.1177/13684302241305054
journals.sagepub.com/home/gpi


Iris J. Traast, Bertjan Doosje and David M. Amodio

Abstract

Research conducted in the United States shows that White Americans form more positive impressions of White than Black interaction partners through instrumental learning. We asked whether this pattern generalizes to the cultural context of the Netherlands, which differs in norms for expressing intergroup bias. In three preregistered studies (*N*s = 66, 83, 80), White Dutch participants played a money-sharing game, based on a reward reinforcement task, with White and Moroccan partners. Although players shared at different rates, average sharing rates for White and Moroccan players were equated. Unexpectedly, and despite anti-Moroccan explicit and implicit attitudes, participants displayed a pro-Moroccan choice preference across studies. Nevertheless, computational modeling indicated the same learning effects of ethnicity as in past research: ethnicity biased initial reward expectations, and these were updated via group-specific learning rates. We discuss potential explanations for this unexpected pattern and broader implications for cross-cultural research on intergroup social cognition.

Keywords

ethnicity, impression, interaction, learning, prejudice

Paper received 14 March 2024; revised version accepted 19 November 2024.

The impressions we form of others are often influenced by their race or ethnicity. Decades of research conducted in the United States find that White Americans tend to form more positive impression of White than Black individuals, even when attributes other than race are held constant (Dovidio et al., 2010; Richeson & Sommers, 2016). This pattern of bias was found in a recent study of interaction-based impression formation (Traast et al., 2024), in which White Americans formed more positive impressions of White interaction partners than Black interaction partners despite identical feedback. In the present research, we asked whether the findings of Traast et al. (2024) would generalize to a cultural context

University of Amsterdam, the Netherlands

Corresponding authors:
Iris J. Traast, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129, REC G, Amsterdam, 1001 NK, the Netherlands.
Email: iris.traast@gmail.com

David M. Amodio, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129, REC G, Amsterdam, 1001 NK, the Netherlands.
Email: d.m.amodio@uva.nl

with a different history and dynamic of intergroup relations: the Netherlands. By examining the effects of White and Moroccan ethnicity in the Dutch context, we sought to determine whether the effect of race/ethnicity on interaction-based impression formation replicates beyond the U.S. context or whether aspects of this process are culture-specific.

Race Effects on Social Instrumental Learning and Impression Formation

Racial prejudice and stereotypes often affect how people learn about, interact with, and form impressions of others (Allport, 1954; Fiske, 1988; Kawakami et al., 2017; Shelton & Richeson, 2006). Among White Americans, this bias may be expressed through the avoidance of interactions with Black Americans (Amodio & Devine, 2006; Dovidio et al., 1997), unfriendly nonverbal behaviors toward Black interaction partners (Dovidio et al., 2002; Fazio et al., 1995; McConnell & Leibold, 2001), and negative judgements of performance based on race (Biernat et al., 2009, 2010; Gaertner & Dovidio, 2000).

Recently, race was shown to also affect how people form impressions of group members through repeated direct interaction, through the process of instrumental learning (Traast et al., 2024). Instrumental learning is a form of reward reinforcement in which an agent learns the reward value of approaching an object (or person) through choice and feedback: choices that result in positive feedback are repeated, whereas those resulting in negative feedback are avoided (Sutton & Barto, 1998). Instrumental learning is an action-based form of learning, supported by dopaminergic activity in the striatum and represented in terms of reward value (O'Doherty et al., 2004), and it has been proposed to support the process of forming social preferences through direct interaction (Hackel et al., 2015, 2020). This form of learning contrasts with the kind of semantic inference previously examined in studies of trait-based impression formation, which tends to be expressed most directly in conceptual judgments and verbal behavior (Amodio, 2019).

An instrumental learning account of impression formation is useful because it provides a theoretical basis for interaction-based social learning as well as a model for how this impression is updated and expressed. According to learning theory (Sutton & Barto, 1998), reward associations are updated incrementally in response to feedback as a function of a *prediction error* (the difference between expected and actual reward feedback on a choice) and a *learning rate* (the degree to which the expected value is updated in response to a prediction error). This theory permits the formalization of specific patterns of learning that can be tested using a computational modeling approach, in which the fit of alternative models to task-based behavioral data is compared (Hackel & Amodio, 2018; Lockwood & Klein-Flügge, 2021).

Traast et al. (2024) used an instrumental learning approach to investigate the effect of race on interaction-based impression formation. In their experiments, White American participants interacted with four Black and four White players in a reinforcement learning task, presented as a money-sharing game. On each trial, participants viewed two players—one Black and one White—and chose to interact with the player expected to share a point (later converted to cash). Although individual players varied in their sharing rate, the average sharing rate was identical between the Black and White groups of players. Nevertheless, participants formed stronger reward associations with White compared with Black players, as indicated by their choice preferences. This effect was moderated by participants' racial attitudes, such that it emerged only for participants with relatively high anti-Black explicit prejudice and low internal motivation to respond without prejudice (Plant & Devine, 1998). These results provided an initial of a race effect on instrumentally learned impressions.

To determine the cognitive mechanisms through which this effect of race occurred, Traast et al. (2024) tested a computational model of race-based instrumental learning. Their model, adapted from a model of stereotype biased learning Schultner, Stillerman, et al. (2024), proposed that race (a) biased White participants' initial expectancies of a

player's feedback behavior before an interaction, modeled as a prior, and then (b) led participants to update reward representations for White and Black players with separate updating rules, modeled as learning rates. In comparisons with alternative models, this hypothesized prior + learning model provided the best fit to behavioral data, revealing that race can influence the process of learning in addition to biasing initial expectancies.

Generalization Beyond the U.S. Context: Ethnic Prejudice in the Netherlands

Although the results of Traast et al. (2024) comport well with previously observed effects of race on impression formation and intergroup behavior (e.g., Devine, 1989; Dovidio et al., 2002; Macrae & Bodenhausen, 2000), it remains unclear whether these effects generalize beyond the U.S. context. Indeed, expressions of racial bias in the US reflect the unique history and current dynamics of race relations. Here, we compared the U.S. and Dutch contexts and considered their implications for effects of ethnicity on impression formation in the Netherlands.

In Dutch society, prejudice based on race is prevalent (Essed & Hoving, 2014; Verkuyten & Thijs, 2002), with White European ancestry being the racial background for native and majority Dutch individuals (Essed & Trienekens, 2008; Mok & Mok, 1999). However, political and public discourse focuses predominantly on concepts of ethnicity or national identity rather than race (Essed & Hoving, 2014; Essed & Trienekens, 2008). In the Netherlands, there are multiple ethnic minority groups due to migration from former colonies such as Surinam, Indonesia, and the Dutch Antilles, as well as labor migration from Morocco and Turkey and refugee migration from countries like Syria, Ethiopia, Eritrea, Bosnia, Iran, Iraq, Somalia, and Rwanda, among others.

Of these minority groups, the Moroccan Dutch population is generally subjected to the most discrimination (Andriessen et al., 2012; Hagendoorn & Pepels, 2017) and may thus be most comparable to Black Americans in the United States. Similar to Black Americans (Bleich et al., 2019; Bowleg et al., 2020; Wingfield & Chavez, 2020), Moroccan Dutch individuals face discrimination on the job market (Andriessen et al., 2012), in health care (Lamkaddem et al., 2012), and during encounters with law enforcement (Bonnet & Caillault, 2015). Dutch stereotypes portray Moroccan individuals as aggressive and violent, similar to Black American stereotypes (Bleich et al., 2019; de Jong, 2007; Hagendoorn, 2017; Kleider-Offutt et al., 2017). These stereotypes are reflected in studies in White Dutch people's responses to Moroccans, such as increased social distance (Dotsch & Wigboldus, 2008) and readiness to perceive facial expressions of anger (Bijlstra et al., 2014).

Despite these parallels, there are several cultural differences. Whereas White Americans include Black Americans in a superordinate American identity (Hehman et al., 2012), ethnic minorities in the Netherlands are often excluded from Dutch identity (i.e., viewed as non-Dutch immigrants), even when born in the Netherlands and native speakers of Dutch (da Silva et al., 2022). Moroccans in the Netherlands are also distinguished from White Dutch by their Islamic faith (De Graaf et al., 2011), whereas White and Black Americans are both predominantly Christian (Kramer et al., 2022).

Finally, social norms regarding intergroup interactions differ between the US and the Netherlands. In the US, there exist strong norms against the expression of prejudice toward Black people and other racial minorities (Crandall et al., 2002). Adherence to this norm has been measured in terms of external motivation to respond without prejudice (Plant & Devine, 1998). External motivation to respond without prejudice is distinct from internal motivation (i.e., based on one's personal beliefs) and, unlike internal motivation, the expression of external motivation is stronger in public situations (Plant & Devine, 1998; Plant et al., 2003). In contrast to the US, norms prohibiting expressions of prejudice are weaker in the Netherlands, where a premium is placed on directness and individual expression (Rottier et al., 2011), and thus the role of external motivation may differ between countries. These cultural

norm differences may be particularly relevant to ethnic prejudice in interaction-based impression formation.

Research Overview

In three preregistered experiments, we investigated the effect of ethnicity on social instrumental learning. These studies were conducted in a Dutch context, where the dominant ethnic majority group is White Dutch, and the primary minority group target of prejudice is Moroccan Dutch (i.e., Moroccan Dutch nationals or immigrants). Therefore, in these studies, White Dutch participants completed the social-interactive task used in Traast et al. (2024), here with White and Moroccan players, and formed impressions based on players' reward feedback.

Following Traast et al. (2024), we hypothesized that ethnicity would modulate impression formation, such that White Dutch participants would form more positive instrumental reward representations for White than for Moroccan players, despite identical reward feedback from each group. We further expected that the effect of ethnicity on learning would be moderated by participants' internal motivation to respond without prejudice as well as their explicit prejudice attitudes, such that this effect would be greater for participants with lower internal motivation and stronger prejudice.

In addition, we hypothesized that the effect of ethnicity on impression formation would stem from two mechanisms: (a) different initial reward expectancies for the ethnic groups (group-based prior), such that participants would begin the task expecting more frequent rewards from White than from Moroccan players, and (b) separate updating rules for White and Moroccan players (group-based learning rates), such that participants would maintain separate representations of White and Moroccan players and update them at separate rates in response to reward feedback. This hypothesis was investigated using computational model fitting, following prior work (Schultner, Stillerman, et al., 2024; Stillerman et al., 2022; Traast et al., 2024).

Deviations from preregistrations are explicitly described in the text and Supplemental Materials. It is notable that our preregistrations for Studies 1 and 2 did not include the central regression analyses; they only included the computational analyses of the main hypothesized effects. However, these regression analyses are based on theory-derived hypotheses and directly replicate analyses reported by Traast et al. (2024), and they were included in the preregistration for Study 3.

Study 1

In Study 1, we conducted an initial test of our hypothesis that player ethnicity—White Dutch and Moroccan Dutch—would moderate the formation of choice preferences through direct sociointeractive learning. We also explored whether this effect would be moderated by internal or external motivation to respond without prejudice and explained by the computational learning mechanisms reported by Traast et al. (2024).

Method

Participants. Participants were 74 self-identified White Dutch psychology university students who completed the study in person in the laboratory. At the end of the experiment, participants indicated their ethnicity, prompted by the question "Please select all categories that you feel apply to you," and chose among the following: *Dutch*, *Moroccan*, *Turkish*, *Antillean*, *Surinamese or different group*, (an open-ended response). They next indicated whether they were born in the Netherlands. The unique selection of "Dutch" was interpreted as White Dutch, given the White European background of Dutch people and the usage of this term in the Netherlands. Participants indicated their gender as either *female*, *male*, *other*, or *choose not to respond*.

Following exclusions based on preregistered criteria (https://aspredicted.org/9pt8i.pdf) for below-chance learning (under 50% choice accuracy; two participants) or extremely fast reaction times (median RT $<$ 500 ms; six participants), the final

sample for analysis included 66 participants (47 female-identified, 12 male-identified, and six who did not indicate gender; M_{age} = 19.80, SD_{age} = 1.95). The preregistered stopping goal was N = 100 Dutch participants, following previous studies using a similar task design (Stillerman et al., 2025). However, due to the COVID-19 outbreak and ensuing lockdown, we were forced to end in-person data collection at 74 White Dutch participants. At this point, we decided to proceed with data analysis, in conjunction with the planning of additional preregistered online replications (Studies 2 and 3). Participants received one research credit plus a performance-based bonus ranging from €1.00 to €2.00.

Procedure. In-person data collection occurred during February and March 2020. Upon arrival to the laboratory, participants provided informed consent and then received instructions regarding the tasks. Participants completed the main learning task, followed by a set of posttask questionnaires. The task and questionnaires were administered on a laboratory computer via the open-source framework psiTurk (Version 3.3.0; Eargle et al., 2020; Gureckis et al., 2016). All data were analyzed using R Statistical Software (R Core Team: Version 4.3.1, 2023).

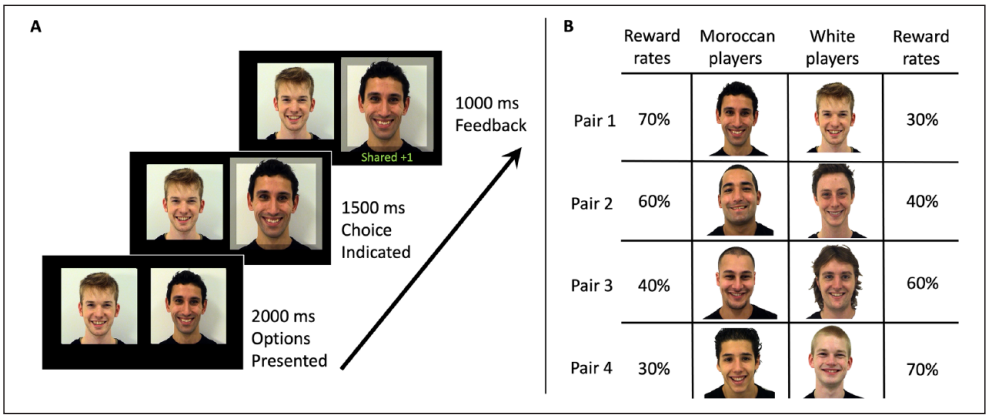
Task and measures

Social reinforcement learning task. Participants engaged in an interactive money-sharing task based on a probabilistic reward reinforcement paradigm (Frank et al., 2004) and adapted for the study of social instrumental learning (Hackel et al., 2015, 2022; Stillerman et al., 2025; Traast et al., 2024). Participants were informed that they would participate in a point-sharing game with eight other players, with the aim of choosing players most likely to share in order to accumulate the maximal points for themselves (converted to a cash bonus at the study's conclusion). Other players were presented as real participants who completed the task previously and whose sharing responses for each trial were taken from this prior study. In actuality, players were fictional and shared according to predetermined fixed reward rates (Figure 1B).

The eight players represented members of two groups, four with a White appearance and four with a Moroccan appearance. In a pretest, White Dutch participants (N = 50; 33 men, 17 women; M_{age} = 31.58, SD_{age} = 12.54) viewed all 16 stimulus faces in randomized order and classified each as Dutch or Moroccan. Mean accuracy was high (M_{Dutch} = 98.67%; $M_{Moroccan}$ = 91.33%), suggesting ethnicity was easily discerned. The gender of players was counterbalanced between participants such that a participant interacted with either all male or all female players. Faces representing players were selected from the Amsterdam Dynamic Facial Expression Set (ADFES; van der Schalk et al., 2011; see Supplemental Material for model numbers). All faces displayed smiles, consistent with the cover story that players were past participants who posed for a picture in their session.

The learning task included two phases: a training phase and a test phase. The training phase comprised two blocks of 80 trials. On each trial of the training phase, participants viewed a pair of faces—always one Moroccan and one White player—and chose which they would like to interact with based on their expectation of who was more likely to share. Following each choice, participants received immediate feedback on whether the chosen player shared 1 or 0 points (Figure 1A). Participants knew that only one player would share per trial. If no response was given within 2.5 s, the trial ended without reward feedback, and a "too slow" message was displayed before proceeding to the next trial. During the training phase, participants chose among four fixed pairs of faces (Figure 1b). The respective reward rates of Moroccan and White players in each pair differed (70/30, 60/40, 40/60, or 30/70) such that in some pairs, the Moroccan player shared more often, whereas in other pairs, the White player shared more often. Critically, although the reward rates of players within each group varied, the average reward rate between White and Moroccan groups was equated at 50%. The assignment of face stimuli to reward rate, face gender, and trial order was randomized across participants, and the presentation side for faces in each pair was randomized across trials.

Figure 1. Trial sequence and player reward rates.



Note. Panel A: sample trial sequence of the training phase. Participants viewed two player faces, chose one to interact with (player on the right in the current trial), and then received feedback ("Shared: +1" or "Shared: 0"). Panel B: reward rates for player pairs during the training phase. Player images were randomized, and gender was counterbalanced.

Next, participants completed the test phase, which assessed what was learned. To this end, no feedback was given. Participants were presented with all possible White–Moroccan player pairings to assess fine-grained preferences generalizing beyond the pairs presented in the training phase. Participants were again instructed to choose the player who was more likely to share, and although no feedback was given, points for correct responses were added to participants' final monetary bonus. Reward learning was indicated by the degree to which participants selected players according to their reward rate during training. Critically, because reward rates were equated between groups, any group-based choice preference during the test phase would represent a group preference.

Perceived reward rates. Following task completion, participants viewed the faces of each player one at a time, in random order, and rated each by typing in a number ranging from 0 to 100% on "What percentage of the time did this player share with you?"

Feeling thermometers. To assess explicit prejudice (Verkuyten & Thijs, 2010), participants reported their warmth toward five major immigrant groups in the Netherlands, as well as White Dutch people, on a scale from 0 (*very cold*) to 100 (*very warm*) degrees. These included Dutch people with no migration background (i.e., White Dutch), a Moroccan migration background, a Turkish migration background, an Antillean migration background, a Surinamese migration background, and a Western migration background (e.g., European or North American). Explicit prejudice was assessed near the end of the session to avoid making our interest in ethnic attitudes salient during the main task.

External and internal motivation to respond without prejudice scales. Internal and external motivations to respond without prejudice toward Dutch Moroccan people were measured using a translated and adapted version of Plant and Devine's (1998) scales (see Supplemental Material; see also Derous et al., 2009; Jargon & Thijs, 2021). The internal motivation scale (IMS) assesses one's personal motivation for responding without prejudice (e.g., "I am personally motivated by my beliefs to be nonprejudiced toward Moroccan Dutch people"), whereas the external motivation scale (EMS) assesses one's motivation due to real or perceived normative pressure (e.g., "I try to hide any negative thoughts about Moroccan Dutch people in order to avoid negative reactions from others"). Participants rated their agreement with each item on a 9-point scale (1 = *strongly disagree*, 9 = *strongly agree*). Separate IMS and EMS scores for each participant were computed following Plant and Devine (1998).

Table 1. Means, standard deviations, and intercorrelations for key variables: Study 1.

Variable	1	2	3	4	5
1. Ethnic difference in choice preference					
2. IMS	.07				
3. EMS	.25*	−.13			
4. Ethnic difference in perceived reward	.85**	−.02	.24		
5. Explicit prejudice	−.04	−.62**	.11	.00	
M	0.56	7.82	4.32	5.33	10.66
SD	0.13	0.90	1.27	15.00	14.69

Note. Ethnic difference in choice preference = proportion of Moroccan over White player choices in test phase, from 0 (choosing only White players) to 1 (choosing only Moroccan players). IMS = Internal Motivation Scale (range: 5.2–9.0; α = .68). EMS = External Motivation Scale (range: 1.3–7.5; α .62). Ethnic difference in perceived reward = perceived reward rate for Moroccan minus White players (scored −100 to 100). Explicit prejudice = feeling thermometer difference score for White Dutch minus Moroccan Dutch; higher scores represent more positive attitudes for Whites over Moroccans.
*$p < .05$. **$p < .01$.

Additional posttask measures, not analyzed or reported here, are described in the Supplemental Material.

Results

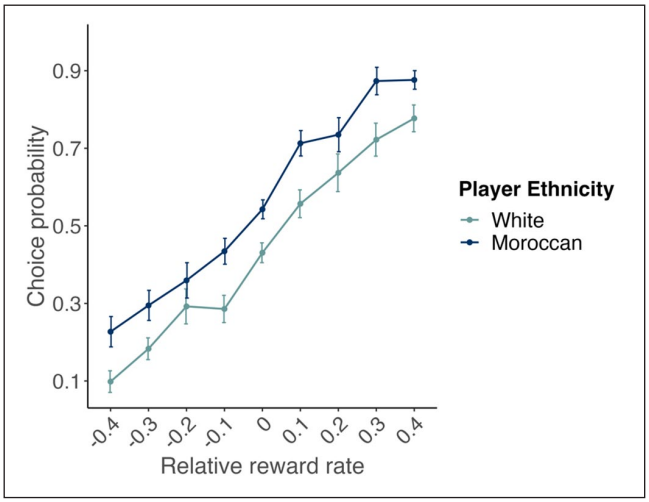
Descriptive statistics and intercorrelations for key variables are presented in Table 1. Preregistration for this study included only the computational modeling analyses; all other analyses were not preregistered but follow directly from analyses reported in Traast et al. (2024).

Explicit prejudice. Prior to testing our primary hypotheses, we examined whether participants showed explicit prejudice consistent with patterns of discrimination in the Netherlands. Indeed, participants reported more positive feeling thermometer ratings of White Dutch ($M = 74.00$, $SD = 14.06$) than Moroccans ($M = 63.34$, $SD = 14.57$), $t(64) = 5.85$, $p < .001$, Cohen's $d = 0.74$, 95% CI [0.46, 1.03]. Relative to other social groups (Turks: $M = 65.03$, $SD = 14.62$; Antilleans: $M = 66.92$, $SD = 13.97$; Surinamese: $M = 69.20$, $SD = 14.07$; Westerners: $M = 70.92$, $SD = 13.02$), participants' attitudes were numerically most positive toward White Dutch people and least positive toward Moroccans. Given our question concerning anti-Moroccan prejudice, an explicit prejudice score was computed as the difference between White Dutch and Moroccan Dutch ratings, with higher scores indicating a more pro-White/anti-Moroccan attitude.

Effects of ethnicity on instrumental learning. Our primary hypothesis was that ethnicity would moderate reward learning such that learned reward associations would be stronger for White than for Moroccan players, despite equated reward rate between groups. This prediction was tested using a generalized linear mixed model with scaled relative reward rate and ethnicity as predictors, with random slopes grouped within participants, and test phase choice behavior as the outcome (R "lme4" package Version 1.1-26; Bates et al., 2015).

This analysis produced the expected main effect of relative reward, $OR = 3.79$, 95% CI [2.99, 4.79], $p < .001$, such that participants learned to prefer players with higher actual reward rates. This analysis also produced an effect of ethnicity on choice behavior, $OR = 2.13$, 95% CI [1.44, 3.13], $p < .001$. However, the direction of this effect was opposite to our prediction: participants exhibited more positive reward associations with Moroccan than with White players (Figure 2). The direction of this ethnicity effect was especially surprising given participants' anti-Moroccan explicit attitudes, in addition to antiminority patterns observed previously (Stillerman et al., 2022; Traast et al., 2024). In what follows, we describe additional

Figure 2. Effects of ethnicity and reward on choice.



Note. Effects of ethnicity and relative reward rate of player on choice during test phase in Study 1, showing a preference for choosing high-rewarding players and for choosing Moroccan players over White ones across relative reward rates. Relative reward rate (difference between training-phase reward rates of a choice pair) is displayed on the x-axis, and choice probability (probability of choosing a player) is displayed on the y-axis. Error bars represent standard errors.

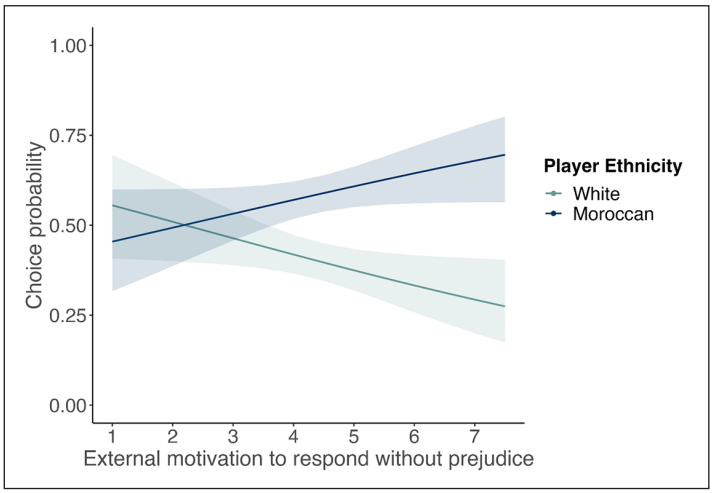
analyses conducted to help explain this unexpected result.

Individual difference effects on choice preference. We first explored whether internal motivation moderated the effect of ethnicity on instrumental learning. Previous research by Traast et al. (2024) found that choice preference for White compared with Black players was moderated by participants' internal motivation, such that participants with lower IMS scores expressed greater anti-Black bias in their choice preferences. Based on this finding, we would expect anti-Moroccan choice preferences among low-IMS participants, but pro-Moroccan preferences among high-IMS participants, relative to White preferences. We tested this using a general linear mixed model (GLMM) with scaled relative reward rate, ethnicity, IMS, and an Ethnicity × IMS interaction as predictors, with choice behavior as the outcome. The Ethnicity × IMS interaction was not significant, $OR = 1.17$, 95% CI [0.76, 1.80], $p = .481$, indicating that the pro-Moroccan choice preference could not be explained by internal motivation.

We then tested whether explicit prejudice moderated the ethnicity effect. Traast et al. (2024) found that participants' choice preference for White players was moderated by explicit prejudice. We therefore ran a GLMM with relative reward rate, ethnicity, explicit prejudice, and an Ethnicity × Explicit Prejudice interaction as predictors, and choice behavior as the outcome. The Ethnicity × Explicit Prejudice interaction was not significant, $OR = 0.99$, 95% CI [0.97, 1.02], $p = .602$, and thus the pro-Moroccan choice effect also could not be explained by participants' explicit prejudiced attitudes.

Finally, we speculated that participants may have shown a preference for Moroccan players to avoid the appearance of prejudice, reflecting external motivation. A GLMM testing main effects of reward rate, ethnicity, and EMS, as well as an interaction between Ethnicity and EMS produced a significant interaction effect between Ethnicity and EMS, $OR = 1.40$, 95% CI [1.04, 1.90], $p = .027$, such that participants with higher EMS scores showed a preference for Moroccan players ($\beta = 1.15$, $t = 4.17$, $p < .001$), whereas

Figure 3. EMS × Ethnicity interaction effect on choice.



Note. Ethnicity × EMS interaction effect on choice in Study 1, showing a stronger effect of ethnicity on choice preference among participants with higher external motivation. EMS is displayed on the x-axis, and choice probability (probability of player being chosen) on the y-axis. Shaded areas represent 95% confidence interval.

those with low EMS showed no ethnic choice preference ($\beta = 0.29$, $t = 1.06$, $p = .287$; Figure 3). This interaction remained significant when IMS and the IMS × Ethnicity effects were included as covariates, $OR = 1.44$, 95% CI [1.06, 1.96], $p = .018$.

If the unexpected pro-Moroccan choice preference were due to external motivation, then we would expect this preference to be already evident at the beginning of the training phase. Thus, to examine this possibility, we tested whether a pro-Moroccan preference was evident in the first 50 trials of training—the initial stage of learning during which participants have minimal experience with each choice pair, but with enough trials to obtain reliable estimates, and during which a motivated (i.e., strategic) preference for Moroccan players should be detectable (Rösler et al., 2024; Traast et al., 2024). We found that, indeed, participants did show a preference for Moroccan players in this earliest stage of learning, $OR = 1.37$, 95% CI [1.10, 1.72], $p = .005$. However, this initial Moroccan preference was not moderated by EMS (EMS × Ethnicity interaction: $OR = 1.05$, 95% CI [0.88, 1.26], $p = .566$), suggesting that this initial pro-Moroccan choice preference could not explain the external motivation effect observed in test-phase choice preferences.

Ethnicity effects on perceived reward rates. Next, we examined whether participants subjectively perceived a difference in the reward rates of Moroccan and White players. Self-reported perceived reward rates were submitted to a multilevel linear regression model with ethnicity and actual player reward rate as predictors. This analysis showed that, in addition to tracking players' actual reward rates, $\beta = 0.89$, 95% CI [0.76, 1.02], $p < .001$, participants estimated higher reward rates from Moroccan than from White players, despite equated feedback from each group, $\beta = 2.67$, 95% CI [0.85, 4.48], $p = .004$.

Finally, we asked whether participants' behavioral preference for Moroccan players simply reflected their subjective (mis)perception that Moroccan players were more rewarding. This was not the case: although the perception of player rewards was associated with behavioral choice

preferences (Perceived Reward Difference × Ethnicity: $OR = 1.09$, 95% CI [1.07, 1.11], $p < .001$), the effect of ethnicity on choice behavior remained significant when perceived rewards were covaried (i.e., when including the Perceived Reward Difference × Ethnicity interaction, the main effect of ethnicity remained significant, $OR = 1.33$, 95% CI [1.05, 1.69], $p = .018$).

Computational modeling results. Despite the surprising finding of pro-Moroccan choice preferences, we tested whether the effect of ethnicity on learning occurred through similar mechanisms as in prior research. Traast et al. (2024; see also Schultner et al., 2024; Stillerman et al., 2022) examined these mechanisms by fitting trial-by-trial behavioral data to alternative computational models of reinforcement learning. They found that choice behavior was best predicted by a *prior + learning model*, which included initial group-based expectancies (prior) and separate group representations for updating (learning rates).

In Traast et al. (2024), reward representations for the different groups were updated using the Rescorla–Wagner learning rule:

$$Q_{i,\,White}^{t+1} = Q_{i,\,White}^{t} + a_{White}\left(R^{t} - Q_{i,\,White}^{t}\right)$$

$$Q_{i,\,Moroccan}^{t+1} = Q_{i,\,Moroccan}^{t} + a_{Moroccan}\left(R^{t} - Q_{i,\,Moroccan}^{t}\right)$$

Priors were modeled as:

$Q_{White}^{t=0} = P$, and $Q_{Moroccan}^{t=0} = -P$

To evaluate the fit of this hypothesized model, we compared the fit of our behavioral data to this hypothesized prior + learning model and three alternatives:

(a) An *unbiased model*, which contains no prior and a single learning rate applied across player ethnicities. In this model, ethnicity does not influence expectations or learning.
(b) A *group-based prior model*, which contains a group-based prior; in this model, participants begin with different initial reward representations for White and Moroccan players but update them according to a single learning rate. This model aligns with classic stereotyping models in which stereotypes shape initial expectations but are replaced with individuated learning over time (Rothbart, 1981).
(c) A *group-based learning model*, which contains no prior but separate learning rates for White and Moroccan players; in this model, participants begin with no group-based expectancies but form group preferences according to separate updating rules.

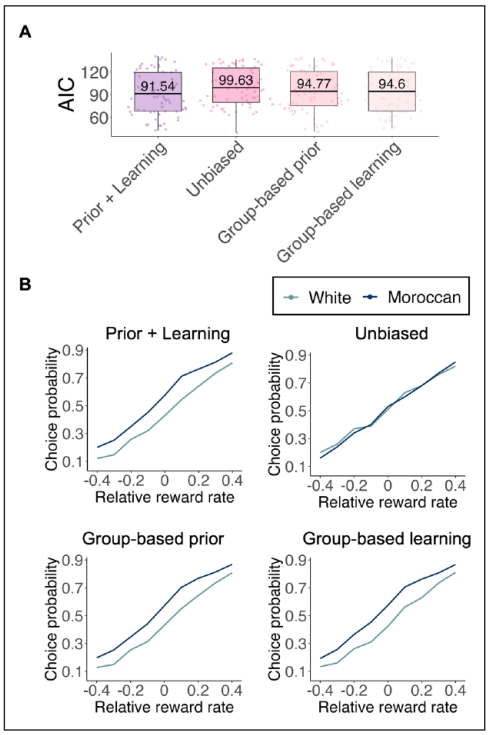
The Akaike information criterion (AIC; Cavanaugh & Neath, 2019; Sakamoto et al., 1986) was used to identify the best fitting model. Model comparisons revealed that the hypothesized prior + learning model fit the data best, explaining the most variation with the fewest parameters, as indicated by the lowest average AIC (Figure 4A). The difference in AIC between the prior + learning model (AIC = 91.54) and competing models (unbiased model: AIC = 99.63, ΔAIC = 8.09; group-based prior model: AIC = 94.77, ΔAIC = 3.23; group-based learning model: AIC = 94.60, ΔAIC = 3.06) suggests that participants did, in fact, form and sustain a group bias through the combination of adopting initial reward expectancies based on group identity and then updating these associations using separate learning rules for Moroccan and White players (Stillerman et al., 2022; Traast et al., 2024).

Discussion

In Study 1, we tested the effect of ethnicity on sociointeractive impression formation in the Dutch cultural context. Unexpectedly, we found that White Dutch participants displayed a behavioral preference for Moroccan players, indicating that they formed stronger reward representations of the Moroccan players compared with the White players. Participants also self-reported that Moroccan players shared more frequently than White players, despite no actual difference, although this subjective perception did not fully explain their behavioral preferences.

What might explain this unexpected pattern? One possibility is that task paradigm—in which

Figure 4. Computational model comparisons and simulated data.



Note. Panel A: Model comparisons between the hypothesized prior + learning model and the unbiased model, group-based prior model, and group-based learning model in Study 1. Panel B: Model-based simulations for each model in Study 1. AIC = Akaike information criterion.

participants repeatedly view and choose between White and Moroccan faces—made our interest in ethnic prejudice too obvious. Indeed, exploratory analyses suggest that participants' pro-Moroccan choice preferences were driven by external motivation to respond without prejudice. External motivation is pronounced in public contexts, where one's behavior may be evaluated by others (Plant & Devine, 1998; Plant et al., 2003), and it is possible that participants experienced the interactive, lab-based task as a public context, despite its confidentiality. However, it is also notable that, on average, participants reported negative attitudes toward Moroccans relative to White Dutch, suggesting that any concerns about appearing prejudiced did not extend to this very direct assessment of prejudice.

Another possible cause of the unexpected effect was the use of smiling faces to depict players, which could have mitigated prejudice by creating a more prosocial and humanizing interaction context (Raissi & Steele, 2021). However, although smiling expressions could have reduced anti-Moroccan prejudice, they are unlikely to have produced negative attitudes toward White players.

Despite observing an unexpected pro-Moroccan choice bias, computational modeling indicated that player ethnicity affected participants' behavior by inducing initial group-based reward expectancies and separate learning rates for White and Moroccan players—the same set of processes observed previously for group membership effects on social-interactive impression formation (e.g., Traast et al., 2024).

Study 2

Study 2 retested our original hypotheses—that participants would form choice preferences for White over Moroccan players—while controlling for factors that may have produced the unexpected result of Study 1. These included using images of players with neutral facial expressions and, to reduce external motivation, the experiment was conducted online. Study 2 refers to the same preregistered hypotheses and analyses used for Study 1.

Method

Participants. Participants were 100 self-identified White Dutch psychology university students, recruited via a university test portal. Participants indicated their ethnicity and gender as in Study 1. As preregistered, data collection stopped at 100 self-identified Dutch participants, with the goal of obtaining valid data from at least 80 participants. Following the same preregistered criteria as in Study 1, exclusions for below-chance learning (under 50% choice accuracy; eight participants) or extremely fast reaction times (median RT $<$ 500 ms; nine participants) resulted in a final sample of 83 participants (58 female-identified, 24 male-identified, one other-identified; M_{age} = 20.46, SD_{age} = 3.18). Participants received one research credit plus a €1.00 to €2.00 performance-based bonus.

Procedure. Data collection occurred from June to September 2020. The task and questionnaires were hosted via psiTurk (Gureckis et al., 2016). Posttask questionnaires were the same as in Study 1, except for four exploratory items not discussed here (see Supplemental Material).

Tasks and measures

Probabilistic reinforcement learning task. The probabilistic reinforcement learning task was the same as in Study 1, but players displayed neutral face expressions (from the same models used in Study 1).

Posttask measures. As in Study 1, participants completed estimates of player reward rates, feeling thermometer ratings for major Dutch ethnic groups, and the IMS and EMS.

Results

Given our aim to replicate Study 1, the same analysis plan was used for Study 2. Descriptives and intercorrelations are shown in Table 2. As in Study 1, only the computational modeling analyses were preregistered; all other analyses were not preregistered but follow those reported by Traast et al. (2024).

Explicit prejudice. As in Study 1, participants reported more positive attitudes toward White Dutch ($M = 76.02$, $SD = 12.41$) than toward Moroccans ($M = 65.84$, $SD = 14.81$), $t(82) = 7.30$, $p < .001$, Cohen's $d = 0.74$, 95% CI [0.51, 0.96]. Again, attitudes were numerically most positive towards White Dutch and least positive towards Moroccans relative to other groups (Turks: $M = 67.71$, $SD = 14.51$; Antilleans: $M = 70.48$, $SD = 14.07$; Surinamese: $M = 73.25$, $SD = 13.51$; Westerners: $M = 73.86$, $SD = 12.96$).

Effects of ethnicity on instrumental learning. Using the same regression model as in Study 1, we expected to find the originally predicted pattern of a pro-White choice preference. However, results again showed the opposite effect: participants displayed a choice preference for Moroccan over White players, $OR = 2.20$, 95% CI [1.37, 3.53], $p < .001$ (Figure 5), in addition to an effect of relative reward, $OR = 2.73$, 95% CI [2.21, 3.37], $p < .001$. As in Study 1, a separate analysis showed no Ethnicity × Relative Reward interaction, $OR = 1.09$, 95% CI [0.96, 1.22], $p = .187$. These results replicated the unexpected finding of Study 1.

As in Study 1, an analysis of choice preferences during the first 50 trials of training revealed an initial preference for Moroccan over White players, $OR = 1.25$, 95% CI [1.00, 1.57], $p = .055$, indicating a choice bias was already present prior to learning.

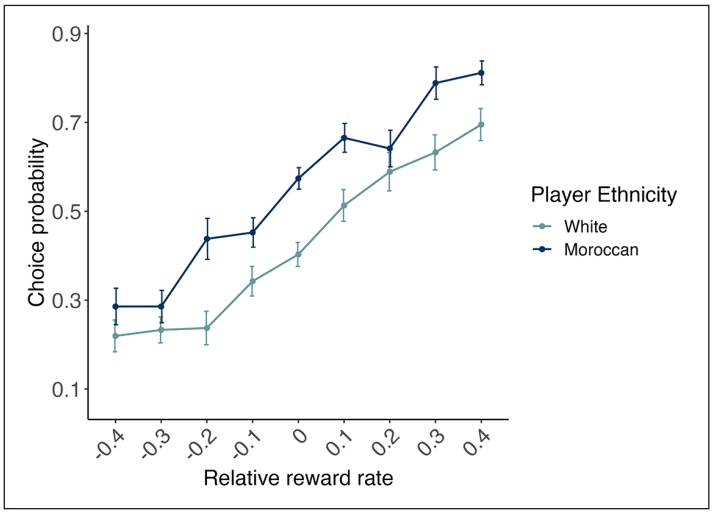
Table 2. Means, standard deviations, and correlations of key variables: Study 2.

Variable	1	2	3	4	5
1. Ethnic difference in choice preference					
2. IMS	.10				
3. EMS	.01	.08			
4. Ethnic difference in perceived reward	.82**	.08	.02		
5. Explicit prejudice	–.06	–.36**	.02	–.00	
M	0.56	7.55	4.80	3.97	10.18
SD	0.17	1.22	1.58	17.72	12.70

Note. Ethnic difference in choice preference = proportion of Moroccan over White player choices in test phase, from 0 (choosing only White players) to 1 (choosing only Moroccan players). IMS = Internal Motivation Scale (range: 3.8–9.0; $\alpha = .77$). EMS = External Motivation Scale (range: 1.0–8.8; $\alpha = .68$). Ethnic difference in perceived reward = perceived reward rate for Moroccan minus White players (scored −100 to 100). Explicit prejudice = feeling thermometer difference score for White Dutch minus Moroccan Dutch; higher scores represent more positive attitudes for Whites over Moroccans (range: −10 to 50).

**$p < .010$.

Figure 5. Effects of ethnicity and reward on choice.



Note. Effects of ethnicity of player and relative reward on choice during the test phase in Study 2, showing a preference for choosing high-rewarding players, and for choosing Moroccan players over White ones across relative reward rates. Relative reward rate (difference between training-phase reward rates of a choice pair) is displayed on the x-axis, and choice probability (probability of choosing a player) is displayed on the y-axis. Error bars represent standard errors.

Individual differences in ethnicity effects. Because Study 2 was conducted online, we expected EMS effects to be reduced or eliminated. Consistent with this reasoning, EMS no longer moderated the effect of ethnicity on choice preference: in a GLMM containing ethnicity and relative reward rate as fixed effects and random effects, respectively, EMS as fixed effect, and an Ethnicity × EMS interaction, this interaction was not significant, $OR = 0.94$, 95% CI [0.69, 1.27], $p = .693$. In a separate GLMM examining IMS effects, the IMS × Ethnicity interaction was also nonsignificant, $OR = 1.25$, 95% CI [0.85, 1.83], $p = .260$.

Ethnicity effects on perceived reward rates. As in Study 1, participants' subjective perceptions of reward rates were predicted by actual reward rates, $\beta = 0.03$, 95% CI [0.03, 0.04], $p < .001$, as well as ethnicity, $\beta = 0.17$, 95% CI [0.01, 0.33], $p = .034$, such that participants perceived more frequent rewards from Moroccan than White players (despite equated reward rates between groups).

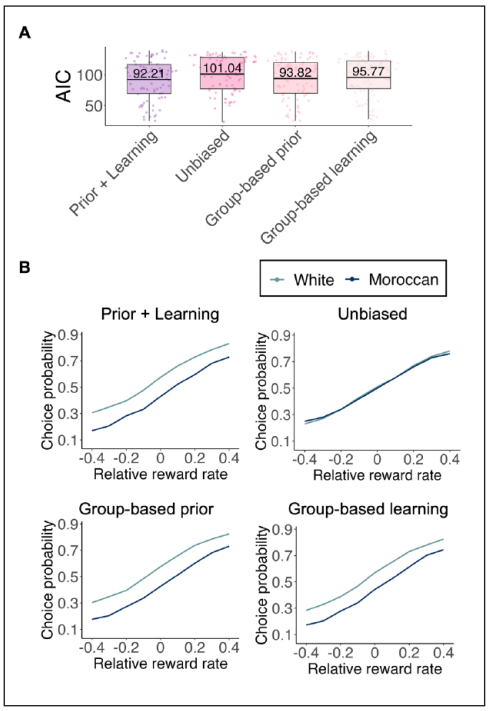
Furthermore, although perceived reward rates were associated with participants' choice preferences (Perceived Reward Difference × Ethnicity: $OR = 1.10$, 95% CI [1.08, 1.12], $p < .001$), perceived reward did not fully explain participants' pro-Moroccan choice behavior (when the Perceived Reward Difference × Ethnicity interaction was included in this regression, the main effect of ethnicity remained significant, $OR = 1.51$, 95% CI [1.13, 2.01], $p = .005$).

Computational modeling. As in Study 1, computational model comparison indicated that the prior + learning model fit best to the choice behavior data, suggesting that ethnicity influenced choice preferences through initial group-based expectancies and then updating according to separate learning rates (Figure 6).

Discussion

Study 2 replicated the unexpected main finding of Study 1: White Dutch participants showed a choice preference for Moroccan over White players, despite equal reward feedback from members of each group. Also replicating Study 1, participants perceived Moroccan players as sharing more frequently, but this subjective

Figure 6. Computational model comparisons and simulated data.



Note. Panel A: Model comparisons between the hypothesized prior + learning model and the unbiased model, group-based prior model, and group-based learning model in Study 2. Panel B: Model-based simulations for each model in Study 2. AIC = Akaike information criterion.

perception did not fully explain the behavioral choice preference. Moreover, these effects emerged despite participants' anti-Moroccan explicit prejudice, as measured by feeling thermometers.

In contrast to Study 1, however, pro-Moroccan choice preferences were no longer associated with external motivation, perhaps due to the online context. Thus, while Study 2 replicated the unexpected findings of Study 1, it continued to leave us without a clear explanation.

Study 3

Study 3 provided another replication while investigating additional reasons for the observed pro-Moroccan choice preference. First, we included a pretask measure of White versus Moroccan implicit attitudes to determine whether participants' pro-Moroccan behavior reflected their implicit rather than explicit attitudes. We also measured posttask implicit attitudes to test whether they were reduced following engagement in the interactive task. Moreover, because implicit attitudes were assessed using the face images of players in the task, we could test whether the pro-Moroccan preferences in the learning task reflected attitudes toward specific individuals as opposed to the broader ethnic group. Second, we included a posttask questionnaire to probe response strategies that could lead to pro-Moroccan task behavior. Finally, we updated our hypothesis to predict the pro-Moroccan effect observed in Studies 1 and 2 and included this new hypothesis in the Study 3 preregistration.

Method

Participants. Participants were 100 self-identified White Dutch psychology university students, recruited via a university test portal. Participants indicated their ethnicity and gender as in Study 1. As preregistered (https://aspredicted.org/2b7g-2z24.pdf), data collection stopped once we obtained valid data from 80 self-identified Dutch participants (62 female-identified, 18 male-identified; M_{age} = 20.50, SD_{age} = 2.29), following exclusions for below-chance learning (six participants) or extremely fast reaction times (14 participants). Participants received one research credit plus a performance-based bonus of €1.00 to 2.00.

Procedure. The study was conducted online from April to June 2021. The procedure was identical to that of Study 2, with the addition of pre- and posttask implicit attitude measures and posttask questionnaires regarding response strategies.

Tasks and measures

Implicit Association Test. Participants' implicit attitudes towards Moroccans versus White Dutch were measured with an Implicit Association Test (IAT; Greenwald et al., 1998, 2003). IATs were

completed immediately before (pretask IAT) and after (posttask IAT) the interactive learning task. Pre- and posttask IATs included 80 Dutch evaluative words unrelated to ethnic stereotypes for evaluative classifications (40 pleasant, 40 unpleasant; see Supplemental Material for word list; van Ravenzwaaij et al., 2011). Four Moroccan and four White faces were included for ethnicity classifications; these were the same faces used to represent players in the task (thus matching on gender). IATs included seven blocks (Greenwald et al., 2003); the order of critical blocks, in which faces and words were classified using the same keys was counterbalanced across participants. Each participant always completed pre- and posttask IATs in the same block order so that their scores on each IAT would be comparable.

Using natural log-transformed reaction times for correct responses, *D* scores were computed for each participant as in Amodio and Devine (2006): compatible block RTs were subtracted from incompatible block RTs and divided by the pooled *SD* separately for practice and test blocks. These resulting scores were then averaged for the final *D* score. Change in implicit attitude was scored as posttask *D* minus pretask *D*.

Probabilistic reinforcement learning task. The probabilistic reinforcement learning task was the same as in Study 2.

Posttask measures. Following task completion, participants indicated their perceived reward estimates for each player and completed feeling thermometers, as in Studies 1 and 2. Next, they completed new questionnaire items assessing response strategies during the task. Participants were asked "When you made a choice for one player or the other, how much was your choice influenced by the following consideration:" Participants then rated each of the following on a 6-point scale (0 = *no influence at all*, 5 = *a very strong influence*): (a) "the ethnicity of the player" (followed by "I predominantly chose Moroccan players" or "I predominantly chose White players"); (b) "I did not want to come across as prejudiced"; (c) "I wanted to choose players of my own ethnicity"; (d) " the appearance of the players, unrelated to their ethnicity"; (e) "whether a player shared money with me in the first few trials that I chose them."

Results

In Study 3, preregistered analyses included the main regression analysis of choice preferences and tests of IAT effects (indicated in text); all other analyses were not preregistered but follow directly from Studies 1 and 2 and Traast et al. (2024).

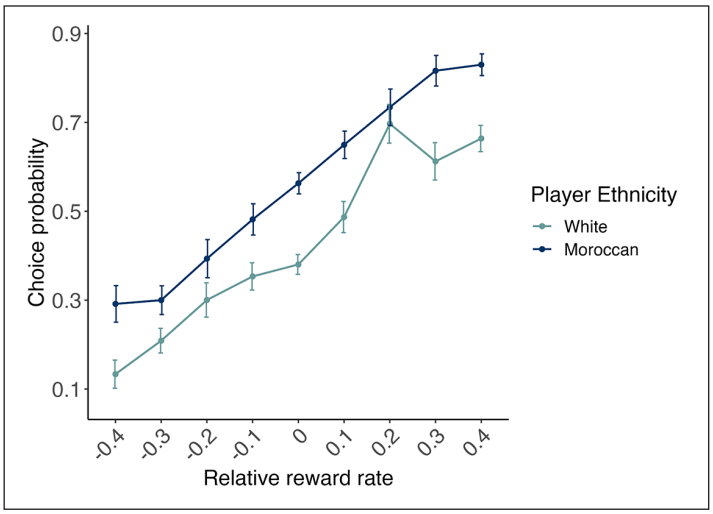
Explicit prejudice. As in Studies 1 and 2, participants reported more positive attitudes toward White Dutch ($M = 78.31$, $SD = 13.17$) than Moroccans ($M = 68.30$, $SD = 16.70$), $t(79) = 5.94$, $p < .001$ (preregistered), Cohen's $d = 0.66$, 95% CI [0.42, 0.90], and attitudes were numerically most positive towards White Dutch and least positive towards Moroccans relative to other groups (Turks: $M = 69.84$, $SD = 16.45$; Antilleans: $M = 73.28$, $SD = 16.51$; Surinamese: $M = 75.59$, $SD = 16.17$; Westerners: $M = 75.71$, $SD = 15.35$).

Effects of ethnicity on instrumental learning. The GLMM described in Studies 1 and 2 produced a significant ethnicity effect (preregistered), indicating a preference for Moroccan over White players, $OR = 2.40$, 95% CI [1.65, 3.50], $p < .001$ (Figure 7), in addition to an effect of actual reward, $OR = 2.64$, 95% CI [2.24, 3.12], $p < .001$. The ethnicity effect was not moderated by reward rate (Ethnicity × Reward Rate: $OR = 1.59$, 95% CI [0.94, 2.27], $p = .087$). These results replicated those of Studies 1 and 2.

To investigate initial choice preference, we again examined preferences during the first 50 trials of training. Again, participants' preference for Moroccan players was already evident in these early trials, $OR = 1.71$, 95% CI [1.44, 2.04], $p < .001$, suggesting that participants were already more likely to choose Moroccan players prior to learning.

Implicit attitude effects. Participants exhibited an implicit preference for White over Moroccan faces on both the pretask IAT ($M = 0.27$, $SD = 0.37$), $t(79) = 6.65$, $p < .001$, and the posttask

Figure 7. Effects of ethnicity and reward on choice.



Note. Effects of ethnicity of player and relative reward rate on choice during the test phase in Study 3, showing a preference for choosing high-rewarding players, and for choosing Moroccan players over White players across relative reward rates. Relative reward rate (difference between training-phase reward rates of a choice pair) is displayed on the x-axis, and choice probability (probability of choosing a player) is displayed on the y-axis. Error bars represent standard errors.

Table 3. Means, standard deviations, and correlations for key variables: Study 3.

Variable	1	2	3	4	5
1. Ethnic difference in choice preference					
2. Pretask IAT	–.03				
3. Posttask IAT	–.02	.32**			
4. Ethnic difference in perceived reward	.77**	.09	.04		
5. Explicit prejudice	–.15	.03	–.06	–.07	
M	0.57	0.27	0.19	8.82	10.01
SD	0.14	0.37	0.30	18.12	15.09

Note. Ethnic difference in choice preference = proportion of Moroccan over White player choices in test phase, from 0 (choosing only White players) to 1 (choosing only Moroccan players). Pretask and posttask implicit association tests (IATs) = *d* score from −1 (relative preference for Moroccan) to +1 (relative preference for White). Ethnic difference in perceived reward = perceived reward rate for Moroccan minus White players (scored −100 to 100). Explicit prejudice = feeling thermometer difference score for White Dutch minus Moroccan Dutch; higher scores represent more positive attitudes for Whites over Moroccans (range: −10 to 50).
**$p < .010$.

IAT ($M = 0.19$, $SD = 0.30$), $t(79) = 5.59$, $p < .001$ (preregistered). This pattern was consistent with participants' explicit attitudes toward White and Moroccan people but contrasted with their behavioral preference for Moroccan players in the task. Neither pretask nor posttask IAT scores correlated with pro-Moroccan choice preference (see Table 3), and thus participants' Moroccan choice bias did not reflect implicit attitudes, nor their attitudes toward individuals as opposed to the broader groups.

Next, we tested whether White participants' implicit attitudes towards Moroccan faces became more positive following their pro-Moroccan choices in the learning task (preregistered). Although a numerical reduction in IAT score was

observed ($M = -0.08$), the difference between pre- and posttask scores did not differ significantly, $t(79) = 1.89$, $p = .062$.

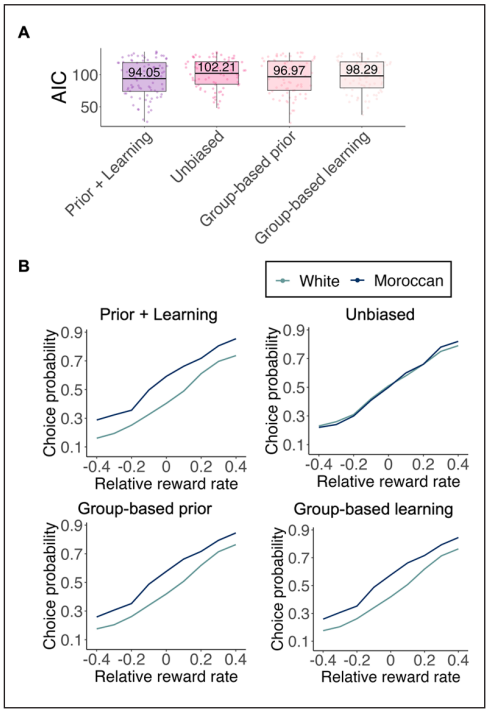
To test whether individual differences in IAT score change related to participants' task preferences, we conducted a GLMM testing main effects of reward rate, ethnicity, and implicit attitude difference, as well as an Ethnicity × Implicit Attitude Change interaction. This model did not produce a significant Ethnicity × Implicit Attitude Change interaction, $OR = 1.03$, 95% CI [0.40, 2.68], $p = .944$, and thus implicit attitude change was not associated with task behavior.

Ethnicity effects on perceived reward rates. Participants' subjective perception of reward rates was again predicted by players' actual reward rates ($\beta = 0.86$, 95% CI [0.72, 0.99], $p < .001$), as well as their ethnicity, ($\beta = 4.41$, 95% CI [2.50, 6.33], $p < .001$), with higher perceived reward rates for the Moroccan than for the White players. This perception of higher rewards from Moroccan players was associated with choice behavior (Perceived Reward Difference × Ethnicity: $OR = 1.08$, 95% CI [1.06, 1.09], $p < .001$), as in Studies 1 and 2. However, unlike the prior studies, inclusion of the Perceived Reward Difference × Ethnicity interaction reduced the main effect of ethnicity to nonsignificance, $OR = 1.28$, 95% CI [0.97, 1.68], $p = .080$, indicating that participants' behavioral choice preferences were not independent from subjective reward perceptions.

Computational modeling. Computational model comparison again indicated that the prior + learning model fit best to the choice behavior data (Figure 8), indicating 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. This result replicated Studies 1 and 2.

Posttask questionnaire. Means and correlations for posttask strategy items are shown in Table 4. Here, we describe responses to each item.

Figure 8. Computational model comparisons and data simulations.



Note. Panel A: Model comparisons between the hypothesized prior + learning model and the unbiased model, group-based prior model, and group-based learning model in Study 3. Panel B: Model-based simulations for each model in Study 3. AIC = Akaike information criterion.

Ethnicity of player. Participants indicated that, on average, a player's ethnicity weakly influenced their decisions ($M = 1.48$, $SD = 1.12$). However, when forced to indicate whether they predominantly chose Moroccan players or White players (Table 4: Variable 3), their answer tended to reflect their task choices, $OR = 0.23$, 95% CI [0.11, 0.49], $p < .001$; participants who indicated they predominantly chose Moroccan players showed a choice preference for Moroccan players during the task ($\beta = 1.61$, $t = 6.55$, $p < .001$), whereas participants who indicated they predominantly chose White players showed no ethnicity effect ($\beta = 0.16$, $t = 0.56$, $p = .574$).

Table 4. Correlations of posttask debriefing items with ethnic difference in choice preference.

Variable	1	2	3	4	5	6	7
1. Ethnic difference in choice preference							
2. Ethnicity	–.06						
3. Perceived group choice	.45**	.09					
4. Seeming nonprejudiced	–.12	.21	–.25				
5. Own group preference	–.25*	.44**	–.21	.26*			
6. Appearance	.10	.17	–.03	.19	.20		
7. Initial reward	–.01	.11	–.06	.06	.07	.16	
M	0.57	1.48	1.41	2.73	0.57	3.03	4.05
SD	0.14	1.12	0.50	1.44	0.89	1.35	1.09

Note. Ethnic difference in choice preference = proportion of Moroccan over White player choices in test phase, from 0 (choosing only White players) to 1 (choosing only Moroccan players). Ethnicity = "the ethnicity of the player" (range: 1–4). Perceived group choice = "I predominantly chose White(1)/Moroccan(2) players." Seeming nonprejudiced = "I did not want to come across as prejudiced" (range: 0–5). Own group preference = "I wanted to choose players of my own ethnicity" (range: 0–3). Appearance = "the appearance of the players, unrelated to their ethnicity" (range: 0–5). Initial reward = "whether a player shared money with me in the first few trials that I chose them" (range: 0–5).
*$p < .050$. **$p < .010$.

Avoiding appearance of prejudice. On average, participants indicated a moderate influence of wanting to avoid the appearance of prejudice in their decisions ($M = 3.03$, $SD = 1.35$). This item was not associated with an ethnicity bias in task choices, conceptually replicating the null effect of EMS in Study 2.

Desire to interact with players of own ethnicity. Participants generally did not report a desire to interact with players of their own ethnicity ($M = 0.57$, $SD = 0.89$). However, this item related to choice behavior, $OR = 0.64$, 95% CI [0.43, 0.97], $p = .037$, such that White participants with lower desire for own-ethnicity interaction showed a pro-Moroccan choice preference ($\beta = 1.25$, $t = 4.74$, $p < .001$), whereas this effect was nonsignificant for participants with higher desire for own-ethnicity interaction ($\beta = 0.48$, $t = 1.81$, $p = .070$). However, this variable was highly negatively skewed, suggesting that this effect was driven by a small number of participants reporting very strong ingroup preference.

Reciprocating player sharing. The most highly endorsed reason for choosing a player was that the player shared with the participant during initial trials ($M = 4.05$, $SD = 1.09$), an explanation that did not reference ethnicity and was unrelated to ethnic preference in choice behavior.

Discussion

Study 3 replicated several findings from Studies 1 and 2. First, White participants again showed a choice preference for Moroccan players over White players and also perceived Moroccan players as sharing more frequently, despite actually receiving equal feedback from both groups. Second, this pro-Moroccan choice bias was already observed early in training behavior, and computational modeling indicated that it reflected an existing pro-Moroccan prior combined with separate learning rates for each group. Third, despite their pro-Moroccan task preferences, participants, on average, reported anti-Moroccan explicit prejudice.

Study 3 additionally assessed implicit attitudes and examined whether repeated interactions with Moroccan players would reduce anti-Moroccan implicit bias. Participants showed anti-Moroccan implicit attitudes before and after the task. They also showed a slight reduction in implicit anti-Moroccan bias following the task. However, this change in implicit attitudes was not associated with choice behavior, and thus we could not conclude that this change was related to participants' engagement with Moroccan and White players in the task. An alternative explanation—that the reduction in IAT scores reflects a practice effect

(Thomas et al., 2007; Vaughn et al., 2011)—is thus more plausible.

IAT scores allowed us to address another possible explanation: that despite participants' explicit prejudice toward Moroccans as a group, they might prefer individual Moroccans in direct interactions. However, scores on the IAT, which assessed responses toward the specific players in the task, showed an anti-Moroccan bias, contradicting this explanation.

Finally, Study 3 probed potential reasons for participants' task behavior, such as whether a pro-Moroccan choice preference originated from a deliberate strategy. However, although many participants reported intentions to respond without bias or to deliberately choose outgroup members, these intentions were not associated with behavioral preferences and thus were not informative.

General Discussion

We examined the effect of ethnicity on impression formation through social interaction in a Dutch context, in an effort to generalize findings previously observed in the U.S. context. Previously, White American participants formed stronger preferences for White than for Black partners through repeated interaction, despite equivalent reward feedback from each racial group (Traast et al., 2024). Thus, we expected White Dutch participants to form stronger preferences for White than for Moroccan interaction partners in a similar interactive task. Unexpectedly, we observed a behavioral preference for Moroccan players over White Dutch players in all three experiments. This unexpected pattern was robust to several design changes across studies—including a switch from using smiling to neutral faces and a switch from in-lab to online data collection, and a cash incentive for accurate choices—and emerged despite participants' self-reported anti-Moroccan prejudice.

This puzzling pattern led us to hypothesize that pro-Moroccan behavior during the interactive task reflected an external motivation to respond without prejudice. However, while an effect of EMS was observed in Study 1, it was not observed in Study 2, based on EMS scores, or Study 3, based on a task-specific measure of external motivation. We further expected that a move to online data collection would reduce external motivation and reveal the originally hypothesized pro-White Dutch choice preference, matching participants' explicit and implicit anti-Moroccan prejudice. But this pattern was not observed—participants continued to show a pro-Moroccan choice preference.

It remains possible that participants intentionally chose Moroccan players for reasons other than external motivation. Participants reported the perception of higher sharing rates from Moroccan than from White players, despite equated rates, which correlated with their choice preferences. Thus, it is possible that participants truly believed Moroccan players had higher sharing rates.

It is also notable that this research was conducted prior to and during the COVID-19 pandemic. COVID-19 had major social and economic impacts in the Netherlands, as it did worldwide, which included increases in prejudice and discrimination toward ethnic minority group members. However, the impact of COVID-19 does not appear to explain our effects. First, we observed across studies a pattern of pro-Moroccan choice preference, which is opposite to the increases in prejudice associated with COVID-19. Furthermore, the main findings of Studies 2 and 3, which were conducted during the COVID-19 lockdown, were virtually identical to those observed in Study 1, which was conducted prior to COVID-19. Therefore, although COVID-19 may have influenced aspects of our research, it does not appear to explain the unexpected pattern of pro-Moroccan preferences.

Despite the unexpected pattern of ethnic preference, computational modeling results indicated that ethnicity influenced impressions through the same mechanisms as observed in past work. Replicating Traast et al. (2024), player impressions were based on a combination of initial expectancies (modeled as priors)—in this case, an expectancy that Moroccan players were more likely to share than White players, corroborating the behavioral preference in early training

trials—and maintaining separate representations of Moroccan and White player reward associations, as indicated by separate learning rates for each group. This pattern indicates that while the pro-Moroccan choice preference was in part due to preexisting expectations, it developed further through the process of learning across that task. While this pattern does not explain participants' pro-Moroccan expectancy, this expectancy may explain why participants perceived higher sharing rates from Moroccan players.

Potential Explanations From a Cultural Perspective

Further consideration of Dutch and American cultural differences may shed light on our unexpected findings. First, we considered the possibility that the nature of intergroup threat differs between contexts. Group threat theory states that both perceived economic threat (Quillian, 1995) and group size (Schlueter & Scheepers, 2010) contribute to perceived threat of a minority outgroup. In the Netherlands, people with a Moroccan background make up only approximately 2.5% of the total Dutch population (Centraal Bureau voor de Statistiek, 2024a; Centraal Bureau voor de Statistiek, 2024b) and are not typically considered an economic threat to the White Dutch majority (Andriessen et al., 2012; Hagendoorn & Pepels, 2017; Ramos et al., 2021). By contrast, in the US, Black Americans comprise approximately 14% of the U.S. population (U.S. Census Bureau, 2024), and there is widespread belief among White Americans that Black Americans and other minorities threaten their jobs (Perkins et al., 2022). Thus, Moroccans may be viewed as less threatening to White Dutch people than Black Americans are to White Americans. However, this cultural difference cannot alone explain our results: lower intergroup threat in the Netherlands would predict lower anti-Moroccan bias but not a reversal to pro-Moroccan bias.

Differences in Moroccan Dutch and Black American stereotypes may also help explain our findings. Moroccans in the Netherlands are often perceived as more generous and warmer than White Dutch, who, by contrast, are stereotyped as greedy and stingy (Van Ginkel, 1996). These stereotypes could have produced a relative pro-Moroccan preference in the context of our money-sharing tasks, and they may explain why participants explicitly reported higher sharing rates from Moroccan than from White players.

Finally, it is possible that cultural differences exist in the nature and expression of external motivation in the US and the Netherlands. Whereas strong norms prohibit the expression of prejudice toward Black people in the US (Plant & Devine, 1998), such norms are relatively weaker in the Netherlands where there is a premium on directness and individual expression (Rottier et al., 2011). Although we measured external motivation using an adapted version of Plant and Devine's (1998) scale, and observed mean EMS scores comparable to those found in U.S. samples, this measure might not sufficiently capture the effect of such norms in the Dutch culture. More research on cultural variation in external motivation may be needed to understand its effects in non-U.S. contexts.

Contributions to Theory on Intergroup Impression Formation

Despite our unexpected findings, this research contributes several advances to knowledge about sociointeractive impression formation and its underlying learning mechanisms. First, it demonstrated an effect of ethnicity on the formation of group member impressions through social interaction, and it replicated a computational model of race on impression formation through repeated interaction (Traast et al., 2024). Second, it provides a first test of the interactive impression formation process in a non-U.S. context, raising new questions regarding cross-cultural generalization. And third, in attempting to explain unexpected findings, this research identified and addressed multiple response strategies that may influence instrumental social learning. Ultimately, this research illuminates a previously unidentified gap in our understanding of sociaointeractive impression formation processes across cultures

and highlights the need for additional research on this topic.

More broadly, our findings add to the growing body of research suggesting that effects observed in the American context may not always emerge identically in different cultural contexts. Because the present research concerned responses to ethnic ingroup and outgroup members, which relate to the unique intergroup relations and histories in the Netherlands, as opposed to those in the US, cultural variability in this context was not completely unexpected. Indeed, the potential for such variability is what inspired this research. Nevertheless, our findings underscore the need to replicate research across cultures to better understand how basic psychological processes are engaged or expressed as a function of cultural context.

Data Availability

Preregistration of study design, hypotheses, and analyses can be found at https://aspredicted.org/3dt6-jdr7.pdf (Study 1 and Study 2) and https://aspredicted.org/2b7g-2z24.pdf (Study 3). The datasets for the three studies and the Supplemental Material are available on the Open Science Framework (OSF) repository (https://osf.io/2tn54/?view_only=d489d37fcfc949a7bb4f19cca62d76fc). All data were analyzed using R Statistical Software (Version 4.3.1). This study complies with Transparency and Openness Promotion (TOP) Level 2 guidelines.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was funded by the Netherlands Organisation for Scientific Research (VICI 016.185.058) awarded to David M. Amodio.

ORCID iD

David M. Amodio https://orcid.org/0000-0001-7746-0150

Supplemental Material

Supplemental material for this article is available online.

References

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley Publishing Company.

Amodio, D. M. (2019). Social cognition 2.0: An interactive memory systems account. *Trends in Cognitive Sciences*, *23*(1), 21–33. https://doi.org/10.1016/j.tics.2018.10.002

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652–661. https://doi.org/10.1037/0022-3514.91.4.652

Andriessen, I., Nievers, E., Dagevos, J., & Faulk, L. (2012). Ethnic discrimination in the Dutch labor market: Its relationship with job characteristics and multiple group membership. *Work and Occupations*, *39*(3), 237–269. https://doi.org/10.1177/0730888412444783

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using eigen and S4. R package version 1.1-10, URL http://CRAN.R-project.org/package= lme4.

Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality & Social Psychology Bulletin*, *36*(7), 855–868. https://doi.org/10.1177/0146167210369483

Biernat, M., Sesko, A. K., & Amo, R. B. (2009). Compensatory stereotyping in interracial encounters. *Group Processes & Intergroup Relations*, *12*(5), 551–563. https://doi.org/10.1177/1368430209337469

Bijlstra, G., Holland, R. W., Dotsch, R., Hugenberg, K., & Wigboldus, D. H. J. (2014). Stereotype associations and emotion recognition. *Personality & Social Psychology Bulletin*, *40*(5), 567–577. https://doi.org/10.1177/0146167213520458

Bleich, S. N., Findling, M. G., Casey, L. S., Blendon, R. J., Benson, J. M., SteelFisher, G. K., Sayde, J. M., & Miller, C. (2019). Discrimination in the United States: Experiences of Black Americans. *Health Services Research*, *54*(Suppl. 2), 1399–1408. https://doi.org/10.1111/1475-6773.13220

Bonnet, F., & Caillault, C. (2015). The invader, the enemy within and they-who-must-not-be-named: How police talk about minorities in Italy, the Netherlands and France. *Ethnic and Racial Studies*, *38*(7), 1185–1201. https://doi.org/10.1080/01419870.2014.970566

Bowleg, L., del Río-González, A. M., Mbaba, M., Boone, C. A., & Holt, S. L. (2020). Negative police encounters and police avoidance as pathways to depressive symptoms among US Black men, 2015–2016. *American Journal of Public Health*, *110*(S1), S160–S166. https://doi.org/10.2105/AJPH.2019.305460

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(3), Article e1460. https://doi.org/10.1002/wics.1460

Centraal Bureau voor de Statistiek. (2024a). *Bevolkingsontwikkeling: maand en jaar.* [Population Development; Month and Year] [Dataset] https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83474NED/table?ts=1735052074659

Centraal Bureau voor de Statistiek. (2024b). Hoeveel inwoners hebben een herkomst buiten Nederland. [How many residents have a background outside the Netherlands] [Dataset] https://www.cbs.nl/nl-nl/dossier/dossier-asiel-migratie-en-integratie/hoeveel-inwoners-hebben-een-herkomst-buiten-nederland

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. hhttps://doi.org/10.1037/0022-3514.82.3.359

da Silva, C., de Jong, J., Feddes, A. R., Doosje, B., & Gruev-Vintila, A. (2022). Where are you really from? Understanding misrecognition from the experiences of French and Dutch Muslim women students. *Journal of Social and Political Psychology*, *10*(1), 201–217. https://doi.org/10.5964/jspp.9395

De Graaf, P. M., Kalmijn, M., & Kraaykamp, G. L. M. (2011). *Sociaal-culturele verschillen tussen Turken, Marokkanen en autochtonen: Eerste resultaten van de Nederlandse LevensLoop Studie (NELLS).* [Socio-cultural differences between Turks, Moroccans, and natives: Initial results from the Netherlands Longitudinal Life Course Study (NELLS)]. [Dossier] https://www.cbs.nl/nl-nl/achtergrond/2011/51/sociaal-culturele-verschillen-tussen-turken-marokkanen-en-autochtonen-eerste-resultaten-van-de-nederlandse-levensloop-studie--nells--

de Jong, J. D. A. (2007). *Kapot moeilijk: een etnografisch onderzoek naar opvallend delinquent groepsgedrag van "Marokkaanse" jongens* [Incredibly Difficult: An Ethnographic Study of Notable Delinquent Group Behavior Among "Moroccan" Boys] [Doctoral thesis, University of Groningen]. University of Groningen Research Output. https://research.rug.nl/en/publications/kapot-moeilijk-een-etnografisch-onderzoek-naar-opvallend-delinque

Derous, E., Nguyen, H. H., & Ryan, A. M. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, *22*(4), 297–320. https://doi.org/10.1080/08959280903120261

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dotsch, R., & Wigboldus, D. H. J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, *44*(4), 1194–1198. https://doi.org/10.1016/j.jesp.2008.03.003

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick & V. M. Esses (Eds.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 3–28). Sage. https://doi.org/10.4135/9781446200919.n1

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. https://psycnet.apa.org/journals/psp/82/1/62/

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. https://doi.org/10.1006/jesp.1997.1331

Eargle, D., Gureckis, T., Rich, A. S., McDonnell, J., & Martin, J. B. (2020). *psiTurk: An open platform for science on Amazon Mechanical Turk* (Version 2.3.11) [Computer software]. https://github.com/NYUCCL/psiTurk.

Essed, P., & Hoving, I. (2014). Innocence, smug ignorance, resentment: An introduction to Dutch racism. In P. Essed & I. Hoving (Eds.), *Dutch racism* (pp. 9–29). Brill. https://brill.com/downloadpdf/book/edcoll/9789401210096/B9789401210096-s002.pdf

Essed, P., & Trienekens, S. (2008). "Who wants to feel White?" Race, Dutch culture and contested identities. *Ethnic and Racial Studies*, *31*(1), 52–72. https://doi.org/10.1080/01419870701538885

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. https://doi.org/10.1037/0022-3514.69.6.1013

Fiske, S. T. (1988). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Ed.), *The handbook of social psychology* (pp. 357–411). McGraw-Hill.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940–1943. https://doi.org/10.1126/science.1102941

Gaertner, S. L., & Dovidio, J. F. (2000). The aversive form of racism. In C. Stangor (Ed.), *Stereotypes and prejudice: Essential readings* (pp. 289–304). Psychology Press. https://psycnet.apa.org/fulltext/2000-16592-016.pdf

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). PsiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. https://doi.org/10.3758/s13428-015-0642-8

Hackel, L. M., & Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. *Current Opinion in Psychology*, *24*, 92–97. https://doi.org/10.1016/j.copsyc.2018.09.001

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: Dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*(9), 1233–1235. https://doi.org/10.1038/nn.4080

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology*, *88*, Article 103948. https://doi.org/10.1016/j.jesp.2019.103948

Hackel, L. M., Mende-Siedlecki, P., Loken, S., & Amodio, D. M. (2022). Context-dependent learning in social interaction: Trait impressions support flexible social choices. *Journal of Personality and Social Psychology*, *123*(4), 655–675. https://doi.org/10.1037/pspa0000296

Hagendoorn, L. (2017). Stereotypes of ethnic minorities in the Netherlands. In K. Phalet & A. Orkeny (Eds.), *Ethnic minorities and inter-ethnic relations in context*. Routledge.

Hagendoorn, L., & Pepels, J. (2017). Why the Dutch maintain more social distance from some ethnic minorities than others: A model explaining the ethnic hierarchy. In W. Vollebergh, J. Veenman & L. Hagendoorn (Eds.), *Integrating immigrants in the Netherlands* (pp. 57–78). Routledge.

Hehman, E., Gaertner, S. L., Dovidio, J. F., Mania, E. W., Guerra, R., Wilson, D. C., & Friel, B. M. (2012). Group status drives majority and minority integration preferences. *Psychological Science*, *23*(1), 46–52. https://doi.org/10.1177/0956797611423547

Jargon, M., & Thijs, J. (2021). Antiprejudice norms and ethnic attitudes in preadolescents: A matter of stimulating the "right reasons." *Group Processes & Intergroup Relations*, *24*(3), 468–487. https://doi.org/10.1177/1368430220902535

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. *Advances in Experimental Social Psychology*, *55*, 1–80. https://doi.org/10.1016/bs.aesp.2016.10.001

Kleider-Offutt, H. M., Bond, A. D., & Hegerty, S. E. A. (2017). Black stereotypical features: When a face type can get you in trouble. *Current Directions in Psychological Science*, *26*(1), 28–33. https://doi.org/10.1177/0963721416667916

Kramer, S., Hackett, C., & Beveridge, K. (2022, September 13th). *Modeling the future of religion in America*. Pew Research Center. https://adam2.org/headlines/support_docs/US-Religious-Projections_FOR-PRODUCTION-9.13.22.pdf

Lamkaddem, M., Essink-Bot, M.-L., Devillé, W., Foets, M., & Stronks, K. (2012). Perceived discrimination outside health care settings and health care utilization of Turkish and Moroccan GP patients in the Netherlands. *European Journal of Public Health*, *22*(4), 473–478. https://doi.org/10.1093/eurpub/ckr113

Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and

behaviour—a reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, *16*(8), 761–771. https://doi.org/10.1093/scan/nsaa040

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*(1), 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442. https://doi.org/10.1006/jesp.2000.1470

Mok, I., & Mok, R. J. M. (1999). *In de ban van het ras: Aardrijkskunde tussen wetenschap en samenleving [Under the Spell of Race: Geography Between Science and Society]. 1876–1992*. ASCA Press.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454. https://doi.org/10.1126/science.1094285

Perkins, K. M., Toskos Dils, A., & Flusberg, S. J. (2022). The perceived threat of demographic shifts depends on how you think the economy works. *Group Processes & Intergroup Relations*, *25*(1), 227–246. https://doi.org/10.1177/1368430220951621

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832. https://doi.org/10.1037/0022-3514.75.3.811

Plant, E. A., Devine, P. G., & Brazy, P. C. (2003). The bogus pipeline and motivations to respond without prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes & Intergroup Relations*, *6*(2), 187–200. https://doi.org/10.1177/1368430203006002004

Quillian, L. (1995). Prejudice as a response to perceived group threat: Population composition and anti-immigrant and racial prejudice in Europe. *American Sociological Review*, *60*(4), 586–611. https://doi.org/10.2307/2096296

Raissi, A., & Steele, J. R. (2021). Does emotional expression moderate implicit racial bias? Examining bias following smiling and angry primes. *Social Cognition*, *39*(5), 570–590. https://doi.org/10.1521/soco.2021.39.5.570

Ramos, M., Thijssen, L., & Coenders, M. (2021). Labour market discrimination against Moroccan minorities in the Netherlands and Spain: A cross-national and cross-regional comparison. *Journal of Ethnic and Migration Studies*, *47*(6), 1261–1284. https://doi.org/10.1080/1369183X.2019.1622824

R Core Team. Version 4.3.1. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, *67*, 439–463. https://doi.org/10.1146/annurev-psych-010213-115115

Rösler, I. K., Kerber, I., & Amodio, D. M. (2024). Effects of moral stereotypes on the formation and persistence of group preferences. PsyArXiv. https://doi.org/10.31234/osf.io/vkgrx

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–181). Erlbaum.

Rottier, B., Ripmeester, N., & Bush, A. (2011). Separated by a common translation? How the British and the Dutch communicate. *Pediatric Pulmonology*, *46*(4), 409–411. https://doi.org/10.1002/ppul.21380

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Springer Netherlands.

Schlueter, E., & Scheepers, P. (2010). The relationship between outgroup size and anti-outgroup attitudes: A theoretical synthesis and empirical test of group threat- and intergroup contact theory. *Social Science Research*, *39*(2), 285–295. https://doi.org/10.1016/j.ssresearch.2009.07.006

Schultner, D. T., Lindström, B. R., Cikara, M., & Amodio, D. M. (2024). *Transmission of social bias through observational learning*. PsyArXiv. https://doi.org/10.31234/osf.io/7ec3u

Schultner, D. T., Stillerman, B. S., Lindström, B. R., Hackel, L. M., Hagen, D. R., Jostmann, N. B., & Amodio, D. M. (2024). Transmission of societal stereotypes to individual-level prejudices through instrumental learning. *Proceedings of the National Academy of Sciences*, *121* (45), e2414518121. https://doi.org/10.1073/pnas.2414518121

Shelton, J. N., & Richeson, J. A. (2006). Interracial interactions: A relational approach. *Advances in Experimental Social Psychology*, *38*, 121–181. https://doi.org/10.1016/S0065-2601(06)38003-3

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. *IEEE Transactions on Neural Networks*, *9*(5), 1054. https://doi.org/10.1109/tnn.1998.712192

Thomas, A., Doyle, A., & Vaughn, D. (2007). Implementation of a computer based implicit association test as a measure of attitudes toward individuals with disabilities. *Journal of Rehabilitation*, *73*(2), 3–14. https://search.proquest.com/openview/f235a2cebb3b07565a944d24a12d65d1/1?pq-origsite=gscholar&cbl=37110

Traast, I. J., Schultner, D. T., Doosje, B., & Amodio, D. M. (2024). Race effects on impression formation in social interaction: An instrumental learning account. *Journal of Experimental Psychology: General*, *153*(12), 2985–3001. https://doi.org/10.1037/xge0001523

U.S. Census Bureau. (2024). *QuickFacts. United States. Population estimates*. https://www.census.gov/quickfacts/fact/table/US/RHI225223

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920. https://doi.org/10.1037/a0023853

Van Ginkel, R. (1996). Foreigners' views of the Dutch: Past and present. *Dutch Crossing*, *20*(1), 117–131. https://doi.org/10.1080/03096564.1996.11784059

van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E.-J. (2011). Does the Name–Race Implicit Association Test measure racial prejudice? *Experimental Psychology*, *58*(4), 271–277. https://doi.org/10.1027/1618-3169/a000093

Vaughn, E. D., Thomas, A., & Doyle, A. L. (2011). The Multiple Disability Implicit Association Test: Psychometric analysis of a multiple administration IAT measure. *Rehabilitation Counseling Bulletin*, *54*(4), 223–235. https://doi.org/10.1177/0034355211403008

Verkuyten, M., & Thijs, J. (2002). Racist victimization among children in the Netherlands: The effect of ethnic group and school. *Ethnic and Racial Studies*, *25*(2), 310–331. https://doi.org/10.1080/01419870120109502

Verkuyten, M., & Thijs, J. (2010). Ethnic minority labeling, multiculturalism, and the attitude of majority group members. *Journal of Language and Social Psychology*, *29*(4), 467–477. https://doi.org/10.1177/0261927X10377992

Wingfield, A. H., & Chavez, K. (2020). Getting in, getting hired, getting sideways looks: Organizational hierarchy and perceptions of racial discrimination. *American Sociological Review*, *85*(1), 31–57. https://doi.org/10.1177/0003122419894335