Routledge
Taylor & Francis Group

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: www.tandfonline.com/journals/pcem20

Reshaping the happy face advantage with reinforcement learning

Tjits van Lent, Harm Veling, Rob W. Holland, Erik Bijleveld & Gijsbert Bijlstra

To cite this article: Tjits van Lent, Harm Veling, Rob W. Holland, Erik Bijleveld & Gijsbert Bijlstra (14 Oct 2025): Reshaping the happy face advantage with reinforcement learning, Cognition and Emotion, DOI: 10.1080/02699931.2025.2568553

To link to this article: https://doi.org/10.1080/02699931.2025.2568553

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2025.2568553

Routledge
Taylor & Francis Group

Reshaping the happy face advantage with reinforcement learning

Tjits van Lent [a]*, Harm Veling [a,b], Rob W. Holland[a], Erik Bijleveld [a] and Gijsbert Bijlstra [a]

[a]Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands; [b]Consumption and Healthy Lifestyles, Wageningen University and Research, Wageningen, The Netherlands

ABSTRACT

Recognising emotional expressions is important for successful social interactions. Prior research has demonstrated that emotion recognition is influenced by evaluative associations people have with different social categories. Here, we systematically investigate whether reinforcement learning can modify social category biases in emotion recognition. Previous research has shown that reinforcement learning is a promising method for altering evaluative associations. In Experiment 1 ($N = 40$), we replicated that the Happy Face Advantage is influenced by social category membership. People were faster at recognising happiness as happiness than anger as anger for White – Dutch faces, while no difference was found for Moroccan – Dutch faces. In Experiments 2–3 ($Ntotal = 144$), we used a reinforcement learning go/no-go task, in which people learned to act to images of Moroccan – Dutch faces to obtain rewards and to not act to images of White – Dutch faces to avoid punishments before participating in the emotion recognition task. Results show that reinforcement learning influences emotion recognition. Instead of the commonly observed interaction effect between social category and expression valence (e.g. in Experiment 1 and previous work), we consistently found a main effect of valence on emotion recognition. These findings suggest that aligning (in)actions with rewards/punishments changes emotion recognition.

ARTICLE HISTORY
Received 11 March 2025
Revised 25 August 2025
Accepted 25 September 2025

KEYWORDS
Emotion recognition; facial expressions; prejudice; social categorisation; instrumental learning

In everyday life, it is important to recognise emotional expressions of others. Through emotional expressions, people can show how they feel about specific situations, demonstrating their strong communicative function. For example, happiness indicates to others that they can approach, while anger may signal that others are better off keeping their distance (Adams et al., 2006; Marsh et al., 2005). The ease with which emotional expressions are recognised is important for successful social interactions (Erickson & Schulkin, 2003). For example, people who are better at recognising emotional expressions tend to exhibit more prosocial behaviours (Marsh et al., 2007), manage conflicts at work more effectively (Côté & Miners, 2006), and experience greater satisfaction in romantic relationships (Yoo & Noyes, 2016). Importantly, however, the ease with which emotional expressions are recognised is biased, e.g. by information on social categories available from faces (e.g. Elfenbein & Ambady, 2002; Hugenberg & Bodenhausen, 2003). Since emotion recognition is an important building block of social interactions, the present research explores whether biases in emotion recognition can be changed.

CONTACT Tjits van Lent t.vanlent@uu.nl Behavioural Science Institute, Radboud University, P.O. Box 9104, Nijmegen 6500 HE, The Netherlands
*Present address: Department of Social, Health and Organisational Psychology, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02699931.2025.2568553.

In general, people are faster at recognising positive emotional expressions than negative emotional expressions (Leppänen & Hietanen, 2004), a phenomenon called the Happy Face Advantage (HFA). Previous research, for example, shows that people are faster at recognising happiness than anger (Hugdahl et al., 1993), sadness (Crews & Harrison, 1994), disgust (Stalans & Wedding, 1985), or neutral faces (Hugdahl et al., 1993). Moreover, there is ample evidence that the HFA is influenced by social information available from a face, such as the social category to which a face belongs. This occurs for a wide range of social categories. That is, previous research finds a larger HFA for faces of women than men (Becker et al., 2007; Craig & Lipp, 2017; Hugenberg & Sczesny, 2006), White males than Black males (Craig et al., 2017; Hugenberg, 2005; Lipp et al., 2015), White – Dutch faces than Moroccan – Dutch faces (Bijlstra et al., 2010), and young than old men (Bijlstra et al., 2019b; Craig & Lipp, 2018). Taken together, previous research has reliably shown an interaction between social category – ranging from different ethnic groups, ages, and genders – and valence of the expression on emotion recognition (see S1 for an overview).

While there is abundant evidence that social category affects the HFA, an important question is what mechanisms are at play. One prominent account in the literature is the evaluative congruence account (e.g. Bijlstra et al., 2010; Craig et al., 2017; Hugenberg, 2005; Hugenberg & Sczesny, 2006). This account suggests that the recognition of emotional expressions is facilitated or inhibited by evaluative social category associations. Relatively more positive associations facilitate the recognition of positive emotional expressions, whereas relatively more negative associations inhibit the recognition of positive emotional expressions. For example, finding an HFA for White – Dutch faces but not for Moroccan – Dutch faces suggests more positive associative evaluations for White – Dutch than Moroccan – Dutch faces. Moreover, Hugenberg and Bodenhausen (2003) demonstrated that participants' evaluative associations with Black and White faces relate to the ease with which anger is perceived (for a similar finding related to stereotype associations, see also Bijlstra et al., 2014). In favour of this account, new evidence is accumulating that the magnitude of the HFA not only depends on the social information available from a face, but also on whether the social category of the face is an ingroup for participants (Martin et al., 2024; Tipples, 2023b), suggesting that positive evaluative associations related to ingroup faces influence the HFA. Thus, evaluative associations related to the social category of a face, relative to the other social category present, seem to facilitate or inhibit the recognition of positive emotional expressions.

People continuously learn about others through interactions with the world (e.g. Heyes, 1994; Olsson & Phelps, 2007). Therefore, these evaluative associations may be malleable. For example, consistent with this line of reasoning, previous research has demonstrated that evaluative associations of social categories can be modified through mere approach or avoidance behaviours (Kawakami et al., 2007). Literature on changing evaluative associations linked to social categories and its consequence for emotion recognition is absent. However, recently, attention has been devoted to the influence of behavioural information about unfamiliar individuals on emotion recognition (Lindeberg et al., 2019). Across two experiments, neutral White faces alongside positive or negative behavioural information about this specific individual were introduced. Subsequently, to test participants' memory, participants were asked to categorise whether an individual did something bad or good and received feedback for their decisions. Finally, participants were asked to recognise happy and angry emotional expressions of these same individuals. Findings indicated that associating new and unfamiliar individuals with positive or negative behaviours resulted in a larger HFA for individuals associated with positive information than negative information. This provides a first indication that *new* evaluative associations of individuals can be learned, and that they can subsequently alter the HFA. Whether it is possible to influence evaluative associations of *existing* social categories using learning processes, and consequently affect emotion recognition remains unclear.

In their experiments, Lindeberg et al. (2019) used reinforcement learning to test participants' memory of which *individual* is "bad" or "good". That is, participants were asked to categorise individuals by the behavioural information and receive feedback (correct or wrong) for their action decisions. Reinforcement learning is a form of learning in which an individual's responses lead to outcomes depending on that response that are perceived as rewarding or punishing. In many reinforcement learning studies, individuals learn through trial-and-error, where the outcomes of their responses shape the

responses toward an optimised response schedule (Sutton & Barto, 2018). Recent suggestions indicate that reinforcement learning is a promising strategy not only for shaping *newly* formed evaluative associations at the individual level but also for altering *existing* evaluative associations with social categories (Amodio, 2019; Amodio & Cikara, 2021). Therefore, we propose that the key to alter the HFA of existing social categories may also be found in these reinforcement learning processes.

There is surprisingly little research on how reinforcement learning affects evaluative associations related to social categories. Some recent work indicates that reinforcement learning affects impression formation for novel non-existing social categories (Allidina & Cunningham, 2021; Hackel et al., 2022) and existing social categories (Traast et al., 2024). Interestingly, these effects can be maximised when consequences align with (in)action decisions (Liu et al., 2025; van Lent et al., 2025). For example, van Lent et al. (2025) show that linking individual faces with actions and rewards during learning led to the most positive evaluative impressions, while linking individual faces with inactions and punishment avoidance during learning led to the least positive evaluative impressions. Thus, on an individual level, evaluative impressions are most strongly influenced when rewards and punishments are aligned with actions and inactions. Given that evaluative associations may underlie impressions, we believe that aligning consequences with (in)actions during learning can shape evaluative associations of social categories.

Here, we investigate whether reinforcement learning affects the recognition of positive and negative emotional expressions. We use reinforcement learning to influence the ease with which emotional expressions are recognised of two social categories: White – Dutch and Moroccan – Dutch faces. In light of existing negative evaluative associations related to Moroccan – Dutch social category members (Verkuyten & Zarembe, 2005), we aim to positively influence evaluative associations linked to Moroccan – Dutch faces and negatively influence evaluative associations linked to White – Dutch faces. These become visible when recognising emotional expressions: We explore whether recognising positive emotional expressions as positive for Moroccan – Dutch faces (compared to recognising negative emotional expressions as negative) is facilitated and recognising positive emotional expressions as positive for White – Dutch faces (compared to recognising negative emotional expressions as negative) is hindered.

The present research

In the current research, we investigate in three experiments how reinforcement learning shapes the HFA. First, in Experiment 1, we attempt to replicate whether social category membership moderates the HFA employing a direct replication of Bijlstra et al. (2010, Experiment 1). We chose Moroccan – Dutch faces because, in the Netherlands, the Moroccan – Dutch community is one of the most negatively prejudiced social categories (Verkuyten & Zarembe, 2005). We predict an HFA for White – Dutch faces and no HFA for Moroccan – Dutch faces. Second, to modify evaluative associations within the social categories, we introduce a reinforcement learning task (Guitart-Masip et al., 2012) in Experiments 2–3 before participants perform the emotion recognition task. Thus, from Experiment 2 onwards, we investigate our main research question of how reinforcement learning shapes the HFA. In the reinforcement learning task, we link Moroccan – Dutch faces to action and reward, as this led to the most positive evaluative impressions in previous research (van Lent et al., 2025). Conversely, we link White – Dutch faces to inaction and punishment avoidance, as this led to the least positive evaluative impressions. In doing so, we aim to maximally modify evaluative associations in favour of Moroccan – Dutch social category members. In Experiments 2 and 3, we will replicate the procedure from Experiment 1 and incorporate the reinforcement learning task before the emotion recognition task. Subsequently, we will test whether the standard interaction between social category and valence of the expression vanishes or even reverses.[1]

Transparency and openness

We report all manipulations, measures, and exclusions in these studies. All data, scripts, and analysis code are available via: https://osf.io/38qdk/. Stimulus materials are available upon request via www.rafd.nl. Data were analyzed using R (v4.3.1; R Core Team, 2023) and the packages afex (v1.3.0; Singmann et al., 2023), tidyverse (v2.0.0; Wickham et al., 2019), emmeans (v1.8.9; Lenth, 2023), lme4 (v1.1.34; Bates et al., 2015), parallel (v4.3.1; R Core Team, 2023), HLMdiag (v0.5.0; Loy & Hofmann, 2014), car (v3.1.2; Fox & Weisberg, 2019), lmerTest (v3.1.3; Kuznetsova et al., 2017),

DescTools (v0.99.50; Signorell, 2023), Rmisc (v1.5.1; Hope, 2022), ggpubr (v0.6.0; Kassambara, 2023) and patchwork (v1.1.2; Pedersen, 2024). The study design, planned sample size, inclusion/exclusion criteria and planned analyses of all experiments were preregistered at the Open Science Framework (Experiment 1: https://osf.io/mn523, Experiment 2: https://osf.io/9x46w, Experiment 3: https://osf.io/kd6yx). The Ethics Committee Social Sciences at Radboud University approved this study (ECSW-2023-070).

Experiment 1

In Experiment 1, we aimed to replicate whether social category membership moderates the HFA (Bijlstra et al., 2010, Experiment 1). That is, participants performed a speeded recognition task with happy and angry White – Dutch and Moroccan – Dutch male faces, and we expect an interaction between the social category of a face (White – Dutch or Moroccan – Dutch) and the valence of the emotional expression (positive or negative) on emotion recognition speed. More specifically, we predict a larger HFA for White – Dutch than Moroccan – Dutch faces.

Method

Sample size justification

An a priori power analysis using summary-statistics-based power analysis (Murayama et al., 2022) indicated that 20 participants were sufficient to find the interaction effect between social category and expression valence (power = .80, alpha = .05), given the data of four experiments (Bijlstra et al., 2010, Experiment 1; Bijlstra et al., 2019a, Experiment 2; Bijlstra et al., 2019b; Hugenberg, 2005, Experiment 1). For this calculation, we used the average *t* statistic (4.12) and average sample size ($N = 32.5$). However, we decided to be conservative and collected data from 40 participants. Participants were rewarded €5 or 0.5 credit point for participating.

Participants

In total, 40 Radboud University students participated ($M_{age} = 23.9$, $SD_{age} = 6.9$, 19–60 years old, 75% women, 25% men, 70% Dutch, 25% German, 5% Belgian). In all experiments, participants were recruited via the Radboud Research Participation System (Sona Systems, https://www.sona-systems.com) and we explicitly recruited participants who grew up in The Netherlands, Germany, or Belgium and who identified with the accompanying ethnicity (i.e. Dutch, German, or Belgian). In doing so, we aimed to create the intergroup context: The White – Dutch faces reflected the ingroup and the Moroccan – Dutch faces were the outgroup for participants. The appearance of the White – Dutch faces is very similar to faces from Belgium or Germany.

Materials and procedure

Emotion recognition task. Upon entering the lab, participants provided consent by signing the informed consent form. Next, we employed a speeded recognition task (e.g. Bijlstra et al., 2010; Hugenberg, 2005) and instructed participants that it was their task to recognise pictures of faces based on their emotional expression (angry or happy) as quickly and accurately as possible. The happy and angry faces used in the task were taken from the Radboud Faces Database (RaFD; Langner et al., 2010; frontal images of actors: 03, 05, 07, 09, 10, 15, 20, 21, 23, 24, 29, 30, 33, 35, 36, 38, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 59, 60, 67, 68, 69, 70, 71, 72, and 73, see Figure 1 for example stimuli). The task

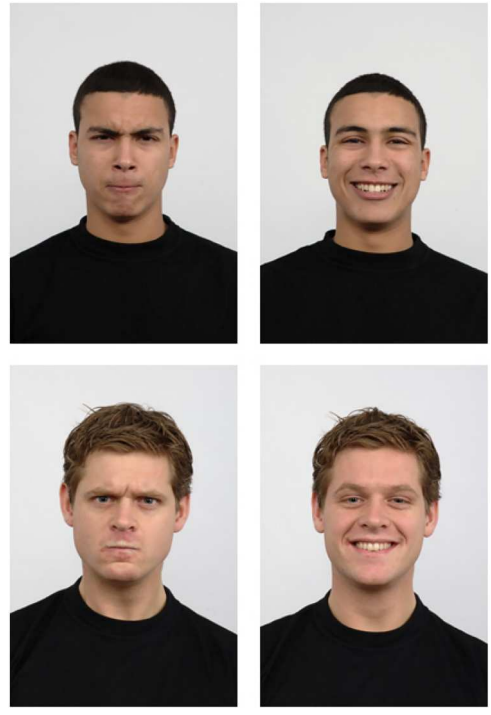

Figure 1. Examples of happy and angry faces.

Note: Example of angry Moroccan – Dutch face (top left), happy Moroccan – Dutch (top right), angry White – Dutch face (bottom left) and happy White – Dutch face (bottom right).

consisted of two experimental blocks of 72 trials (18 happy White – Dutch faces, 18 happy Moroccan – Dutch faces, 18 angry White – Dutch faces, and 18 angry Moroccan – Dutch faces). The faces were randomised within each block. Each trial started with a fixation cross presented for 1000 ms followed by a White – Dutch or Moroccan – Dutch face expressing either anger or happiness for 200 ms. In all blocks, participants were asked to recognise the emotional expression as either angry or happy by pressing the "A" or "L" key. Participants' response time is our main measure of interest. To closely adhere to the original paradigm by Bijlstra et al. (2010), the order of response mapping was counterbalanced within participants. Before each block, participants took part in eight practice trials to get familiarised with the task (using different faces than the ones used in the experimental trials). Upon finishing the experiment, participants were asked to fill in their demographics (i.e. age, gender and ethnicity). The task lasted approximately 20 min and was programmed using Inquisit (Inquisit 6, 2022).

Confirmatory analyses

Emotion recognition task. The response time needed to recognise emotional expressions served as the main dependent variable. As preregistered and similar to Bijlstra et al. (2010), we performed all confirmatory analyses on log-transformed response times due to the right-skewed distribution. To determine differences in response time needed to recognise happy and angry expressions between White – Dutch and Moroccan – Dutch faces, we conducted a linear mixed model. This model included the within-participant factors social category (White – Dutch/Moroccan – Dutch) and expression valence (positive/negative). Moreover, this model included a random intercept of stimulus and participant as well as random slopes for social category and expression valence for participant. All models have a maximal random-effects structure (Barr et al., 2013) and all fixed effects were coded using sum-to-zero contrasts.[2]

Results

Confirmatory analyses

Emotion recognition task. As preregistered and similar to Bijlstra et al. (2010), we excluded incorrect trials (8.21%) and response times below 200 ms (0.07%). In line with our preregistered hypothesis, there was a significant interaction between social category and expression valence on response time, $B = 0.007$, $SE = 0.003$, $F(1,27.97) = 4.78$, $p = .037$, 95% CI [0.001, 0.01]. As expected, responses to happy emotional expressions ($M = 480$, $SD = 38$) were faster than responses to angry emotional expressions ($M = 487$, $SD = 31$) when displayed by White – Dutch faces, $B = 0.02$, $SE = 0.01$, $p = .035$. No difference was found between response times to happy ($M = 490$, $SD = 26$) and angry ($M = 486$, $SD = 31$) emotional expressions displayed by Moroccan – Dutch faces, $B = -0.008$, $SE = 0.009$, $p = .385$ (see Figures 3 and 4). For the sake of completeness, responses to happy emotional expressions were faster when displayed by White – Dutch than Moroccan – Dutch faces, $B = -0.025$, $SE = 0.009$, $p = .008$, but there was no significant difference for angry emotional expressions, $B = 0.004$, $SE = 0.009$, $p = .684$. Although not hypothesised, there was no significant main effect of social category, $B = -0.005$, $SE = 0.003$, $F(1,26.80) = 2.87$, $p = .102$, 95% CI [– 0.01, 0.001], and expression valence, $B = 0.003$, $SE = 0.003$, $F(1,29.50) = 0.86$, $p = .361$, 95% CI [– 0.003, 0.01] on response time. Taken together, we replicated influences of social category on the HFA.

Discussion

Consistent with prior research (e.g. Bijlstra et al., 2010; Hugenberg, 2005; Lipp et al., 2015), we found evidence that social category membership moderates the HFA. That is, we find evidence that the HFA is present for White – Dutch faces, but not for Moroccan – Dutch faces. By replicating social category influences on the HFA, we provide further evidence that evaluative associations of social categories affect the speed of emotion recognition. This paves the way for the main purpose of this research: Investigating whether the effect found in Experiment 1 can be modified by reinforcement learning processes.

Experiment 2

In Experiment 2, we aimed to investigate whether the HFA for Moroccan – Dutch and White – Dutch faces can be modified through reinforcement learning. We expected an interaction between the social category of the face (White – Dutch or Moroccan – Dutch) and the valence of the expression (positive or negative) on participants' response time needed to recognise emotional expressions. However, different from what was observed in Experiment 1, we expect an

HFA for Moroccan – Dutch faces, and a smaller, reversed, or no HFA for White – Dutch faces. Lastly, as manipulation check, we measured explicit evaluations. If reinforcement learning affects evaluative associations, this is expected to be reflected in explicit evaluations. We expected participants' evaluation of faces to be more positive for Moroccan – Dutch faces than White – Dutch faces.

Method

Sample size justification

An a priori power analysis using simr (Green & MacLeod, 2016) indicated that 72 participants were sufficient to find an interaction between social category and expression valence (power = .80, alpha = .05) given the data of Experiment 1 ($B = 0.007$). Participants were rewarded 7.5 euro or 0.75 credit point for participating and could earn a bonus based on their performance in the Reinforcement Learning Go/No-Go Task (RL GNG Task; up to €3).

Participants

After recruiting 72 participants, we excluded two participants according to our preregistered exclusion criteria. One was excluded because she had a non-Dutch cultural background (Irish) and one because she performed the same action choice (e.g. always press the same key) more than or equal to 90% of the time in at least one of the four blocks of the RL GNG Task. Additionally, we excluded one participant who did not finish the experiment. We resampled the number of excluded participants to again reach a sample size of 72 ($M_{age} = 20.8$, $SD_{age} = 3.5$, 18–40 years old, 79.17% women, 20.83% men, 84.72% Dutch, 15.28% German).

Materials and procedure

Our aim was to investigate whether the HFA for in – and outgroup faces can be modified through reinforcement learning. To that end, we included a RL GNG Task before the speeded emotion recognition task in Experiment 2.

Reinforcement learning go/no-go task. To modify evaluative associations of social categories, we asked participants to first participate in a reinforcement learning task, aiming to positively change evaluative associations of Moroccan – Dutch faces and negatively change evaluative associations of White – Dutch faces, before participating in the emotion recognition task. After providing consent, participants were asked to fill in their demographics (i.e. age, gender, and ethnicity). The procedure of the RL GNG Task was adapted from Guitart-Masip et al. (2012; see also van Lent et al., 2025). Participants were shown pictures of fractals and faces of different categories. Each category required a specific response: Either go (press the spacebar) or no-go (do not press). Each response also led to a reward (i.e. gain one point), a neutral outcome (i.e. neither gain nor loss), or a punishment (i.e. lose one point). In total, there were four different categories: For two they could win points by either go (Go-To-Win) or no-go (No-Go-To-Win) and for two they could avoid losing points by either go (Go-To-Avoid-Losing) or no-go (No-Go-To-Avoid-Losing).

Based on the (in)action, participants received feedback. This feedback was probabilistic: If participants performed the correct action (e.g. pressed the spacebar when the category of the object on the picture required a "go" response), they received a reward (for Go-To-Win and No-Go-To-Win) or a neutral outcome (for Go-To-Avoid-Losing and No-Go-To-Avoid-Losing) in 80% of the trails. This means that 20% of the correct trials resulted in a neutral outcome (for Go-To-Win and No-Go-To-Win) or punishment (for Go-To-Avoid-Losing and No-Go-To-Avoid-Losing). Feedback was probabilistic to ensure learning was more like real-life learning, the task was not too easy and, in addition, partial reinforcement is more resistant to extinction. For every participant, pictures of neutral Moroccan – Dutch faces were linked to the Go-To-Win condition and pictures of neutral White – Dutch faces were linked to the No-Go-To-Avoid-Losing condition. This was done because previous research (van Lent et al., 2025) has shown that those conditions were the most effective in changing evaluative associations: Go-To-Win the most positive and No-Go-To-Avoid-Losing the least positive. The remaining two conditions, No-Go-To-Win and Go-To-Avoid-Losing, were linked to pictures of either orange or blue fractals. Although we are not necessarily interested in the No-Go-To-Win and Go-To-Avoid-Losing conditions, we decided to include them in the design to make the task less obvious and thereby aim to minimise social desirability influences. The fractals were adapted from Mathôt et al. (2015).

Finally, for each category participants had to learn trial and error based on the feedback what the best response is. Thus, the optimal response (go/no-go)

for each category was not instructed and had to be discovered based on the feedback. Trial and error learning is an important component of reinforcement learning, as it makes the learning active. Participants also learned that after completion of the task, the points would be converted to a monetary bonus ranging from 0 to 3 euro. In this way, learning was made consequential to stimulate the learning process. More specifically and unbeknownst to participants, participants with scores of 0 or below 0 points gained 0 euro bonus, participants with scores ranging between 1 and 30 gained 1 euro bonus, participants with scores ranging between 31 and 60 gained 2 euro bonus and participants with scores higher than 60 gained 3 euro bonus ($M_{points} = 32.18$, $SD_{points} = 20.39$; $M_{bonus} = 1.55$, $SD_{bonus} = 0.67$).

Each trial started with presenting one picture for 1500 ms, and during presentation, participants either had to press the spacebar (go) or withhold from pressing (no-go). There were five pictures linked to each condition: Five neutral Moroccan – Dutch faces, five neutral White – Dutch faces, five orange fractals, and five blue fractals. These pictures were kept constant throughout all trials. Moroccan – Dutch and White – Dutch faces were randomly selected per participant out of a list of 18 faces. After participants chose the action, they received feedback for 2000ms. They either received a reward (upwards pointing green arrow), a punishment (downwards pointing red arrow) or a neutral outcome (yellow bar). Each trial ended with an inter-trial interval (ITI) that varied from 1000 ms to 1750ms in steps of 150 ms (see Figure 2 for an overview).

In total, the task included four categories (each including five pictures) with 60 trials per category resulting in 240 trials. After every 60 trials (15 trials per RL condition; Go-To-Win, No-Go-To-Win, Go-To-Avoid-Losing, No-Go-To-Avoid-Losing), participants had a 20 s break. The trials within the blocks were randomised. Before starting the task, participants took part in 10 practice trials per condition to get familiarised with the task (using different pictures than the ones used in the actual RL GNG Task).

Emotion recognition task. The emotion recognition task remained the same as in Experiment 1.

Explicit evaluation task. Lastly, participants were asked to judge all 36 neutral faces on a 200-point scale (0 = *very negative*, 200 = *very positive*). The order of the faces was randomised per participant. Ratings were made using a slider; its starting position was always at 100 by default (neither positive nor negative). In total, all three tasks lasted approximately 45 min and participants were paid according to their performance. The task was programmed using PsychoPy (Peirce et al., 2019).

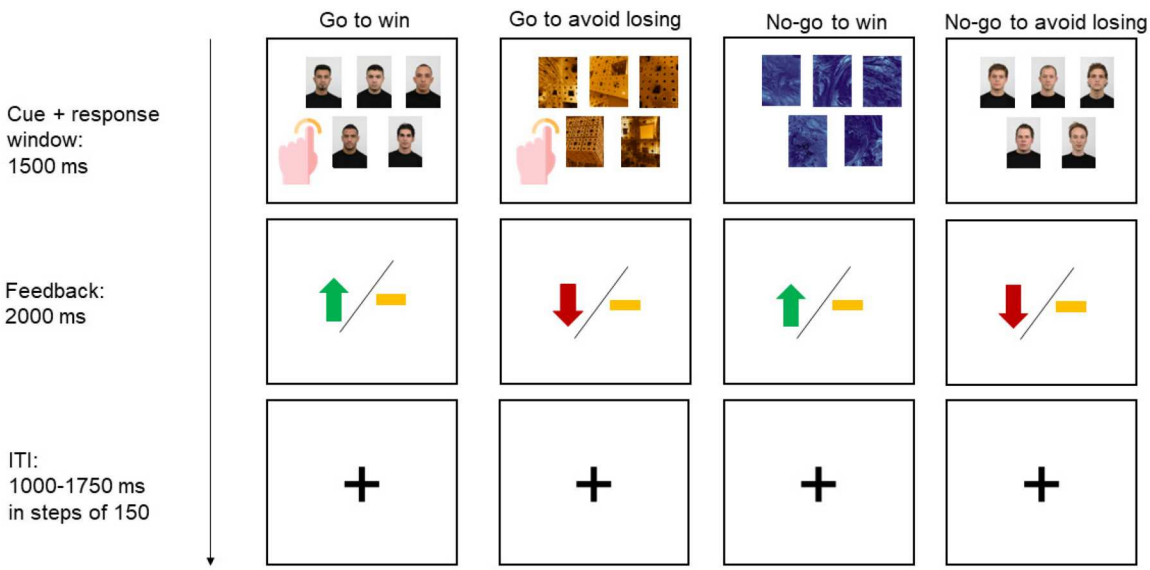

Figure 2. Overview of the reinforcement learning go/no-go task.

Note: Each trial started with the presentation of a face or fractal and was followed by response-dependent feedback. Rewards, punishments, and neutral outcomes were visualized by upwards green arrows, downwards red arrows and yellow bars, respectively.

Confirmatory analyses

Emotion recognition task. The analyses of the emotion recognition task were identical to Experiment 1.

Explicit evaluation task. To determine whether there is a difference in how the faces are evaluated, we conducted a linear mixed effects model. The model included the within-participant factor social category (Moroccan – Dutch and White – Dutch) and a random intercept of participant as well as a random slope for social category.[3] Moreover, a random intercept of stimulus was included.

Results

Confirmatory analyses

Emotion recognition task. As preregistered and similar to Bijlstra et al. (2010), we excluded incorrect trials (8.25%) and response times below 200 ms (0.77%). In contrast to our preregistered hypothesis, there was no significant interaction between social category and expression valence on response time, $B = 0.004$, $SE = 0.003$, $F(1,54.12) = 1.63$, $p = .208$, 95% CI [−0.002, 0.011]. There was no significant main effect of social category on response time, $B = -0.004$, $SE = 0.003$, $F(1,56.82) = 1.07$, $p = .305$, 95% CI [−0.01, 0.003]. There was, however, a significant main effect of expression valence on response time, $B = 0.01$, $SE = 0.003$, $F(1,57.65) = 10.63$, $p = .002$, 95% CI [0.005, 0.018], suggesting an overall HFA. Responses to happy emotional expressions ($M = 422$, $SD = 37$) were faster than responses to angry emotional expressions ($M = 432$, $SD = 43$) (see Figures 3 and 4).

Although we did not find an interaction between social category and expression valence on response time, we did zoom in on the response time differences within each social category for the sake of completeness. Responses to happy emotional expressions ($M = 418$, $SD = 26$) were faster than responses to angry emotional expressions ($M = 431$, $SD = 38$) when displayed by White – Dutch faces, $B = 0.03$, $SE = 0.01$, $p = .001$. Although response times to happy emotional expressions were numerically faster ($M = 426$, $SD = 44$) than response times to angry emotional expressions ($M = 433$, $SD = 49$) when displayed by Moroccan – Dutch faces, this difference was not significant, $B = 0.014$, $SE = 0.01$, $p = .175$. Moreover, for the sake of completeness, there was no significant difference between White – Dutch and Moroccan – Dutch faces for happy emotional expressions, $B = -0.02$, $SE = 0.01$, $p = .099$, and angry emotional expressions, $B = 0.002$, $SE = 0.01$, $p = .853$. Taken together, although not in the expected direction, the data pattern changed compared to what was found in Experiment 1.

Explicit evaluation task. In line with our preregistered hypothesis, there was a significant main effect of social category on explicit evaluation, $B = -4.25$, $SE = 1.83$, $F(1,49.66) = 5.41$, $p = .024$, 95% CI [−7.83, −0.67] (see Figure 5). Moroccan – Dutch faces ($M = 102.07$, $SD = 10.56$) were more positively evaluated than White – Dutch faces ($M = 93.57$, $SD = 10.56$).

Exploratory analyses

Emotion recognition task. We explored whether response latencies differed over the course of the experiment, to see whether any influences of the RL GNG Task are present at first, but extinct over time. To do this, we analyzed block 1 and block 2 separately.

Block 1. In block 1, there was no significant interaction between social category and expression valence on response time, $B = 0.003$, $SE = 0.004$, $F(1,52.44) = 0.66$, $p = .421$, 95% CI [−0.005, 0.012], and no significant main effect of social category on response time, $B = -0.002$, $SE = 0.004$, $F(1,55.66) = 0.31$, $p = .580$, 95% CI [−0.011, 0.006]. There was a significant main effect of expression valence on response time, $B = 0.02$, $SE = 0.005$, $F(1,68.96) = 14.35$, $p < .001$, 95% CI [0.009, 0.028], suggesting an overall HFA. Responses to happy emotional expressions ($M = 409$, $SD = 40$) were faster than responses to angry emotional expressions ($M = 428$, $SD = 55$).

Again, we zoomed in on the response time differences per emotion within each social category for the sake of completeness. Responses to happy emotional expressions ($M = 408$, $SD = 32$) were faster than responses to angry emotional expressions ($M = 427$, $SD = 51$) when displayed by White – Dutch faces, $B = 0.04$, $SE = 0.01$, $p < .001$, and responses to happy emotional expressions ($M = 411$, $SD = 48$) where faster than responses to angry emotional expressions ($M = 429$, $SD = 59$) when displayed by Moroccan – Dutch faces, $B = 0.03$, $SE = 0.01$, $p = .016$. Moreover, there was no significant difference between White – Dutch and Moroccan – Dutch faces for happy emotional expressions, $B = -0.01$, $SE = 0.01$, $p = .323$, and angry emotional expressions, $B = 0.002$, $SE = 0.01$, $p = .871$ (see Figure 3).

Block 2. In block 2, there was no significant interaction between social category and expression valence on response time, $B = 0.006$, $SE = 0.045$, $F(1,53.95) = 1.59$, $p = .213$, 95% CI [−0.003, 0.014], no significant main effect of social category on response time, $B = -0.004$, $SE = 0.004$, $F(1,43.42) = 1.39$, $p = .244$, 95% CI [−0.012, 0.003] and no significant main effect of expression valence on response time, $B = 0.003$, $SE = 0.004$, $F(1,56.89) = 0.62$, $p = .436$, 95% CI [−0.005, 0.013] (see Figure 3).

Discussion

In Experiment 2, we did not find the preregistered interaction between social category and valence of the expression on emotion recognition (note that we expected a reserved-shaped interaction). We, however, did find a main effect of valence of the expression, signalling an overall HFA. As preregistered, we found that participants' explicit evaluation of faces was more positive for Moroccan – Dutch faces than for White – Dutch faces.

Our findings suggest that being subjected to a reinforcement learning task adjusted the HFA. That is, after the reinforcement learning task, we did not find evidence for the standard interaction between social category and valence of the expression on emotion recognition. Instead, participants showed an overall HFA. Additionally, our data suggest participants hold a more positive explicit evaluation of Moroccan – Dutch faces than White – Dutch faces. Thus, these results suggest that evaluative associations of social categories are malleable with reinforcement learning, but that these effects are not as strong as expected. It seems to be very hard to negatively affect evaluative associations of ingroup faces with reinforcement learning.

After exploring the data, it seems that an overall HFA was only present in the first block. Counterbalancing the keys after the first block seems to disrupt the impact of learning in the reinforcement learning task on emotion recognition. To further investigate whether learning in the reinforcement learning task affects emotion recognition, we decided to replicate Experiment 2, but to counterbalance the order of response mapping between participants in Experiment 3 instead of within participants.

Experiment 3

Experiment 3 consisted of only one experimental block (144 trials) to test the influence of reinforcement learning on emotion recognition. Based on the results of Experiment 2, we changed our hypothesis for the emotion recognition task. We expected a main effect of expression valence on participants' response time needed to recognise emotional expressions. Moreover, regarding the explicit evaluations, our hypothesis remained the same: We expected that participants' evaluation of faces is more positive for Moroccan – Dutch faces than White – Dutch faces.

Method

Sample size justification

An a priori power analysis using simr (Green & MacLeod, 2016) indicated that 35 participants were sufficient to find a main effect of expression valence (power = 0.80, alpha = .05) given the main effect of expression valence in block 1 of Experiment 2 ($B = 0.02$). We decided to collect 72 participants, as the power analysis for Experiment 2 indicated that this number was necessary to detect an interaction effect between social category and expression valence. To reduce the risk of a type II error, we recruited 72 participants, ensuring a power of 0.95 to detect the main effect of expression valence.

Participants

After recruiting 72 participants, we excluded one participant according to our preregistered exclusion criteria. This participant had to be excluded because she performed the same action choice (e.g. always press the same key) more than or equal to 90% of the time in at least one of the four blocks of the RL GNG Task. We resampled the number of excluded participants to again reach a sample size of 72 ($M_{age} = 20.51$, $SD_{age} = 2.94$, 18–32 years old, 87.5% women, 12.5% men, 77.78% Dutch, 20.83% German, 1.39% Dutch/German).

Materials and procedure

The setup of the experiment was the same as Experiment 2. The only difference was that the order of the response mapping in the emotion recognition task was counterbalanced between participants, thereby removing the break. As this is a short task (10–15 min), we thought a break was not necessary. In the RL GNG Task participants again gained money ($M_{points} = 36.46$, $SD_{points} = 19.17$; $M_{bonus} = 1.65$, $SD_{bonus} = 0.73$).

Confirmatory analyses

The confirmatory analyses were the same as in Experiment 2.

Results

Confirmatory analyses

Emotion recognition task. As preregistered and similar to Experiments 1 and 2, we excluded incorrect trials (7.19%) and response times below 200 ms (0.44%). In line with our preregistered hypothesis, there was a significant main effect of expression valence on response time, $B = 0.01$, $SE = 0.004$, $F(1,81.88) = 6.50$, $p = .013$, 95% CI [0.003, 0.019], suggesting an overall HFA. Responses to happy emotional expressions ($M = 433$, $SD = 31$) were faster than responses to angry emotional expressions ($M = 442$, $SD = 30$). Again, there was no significant

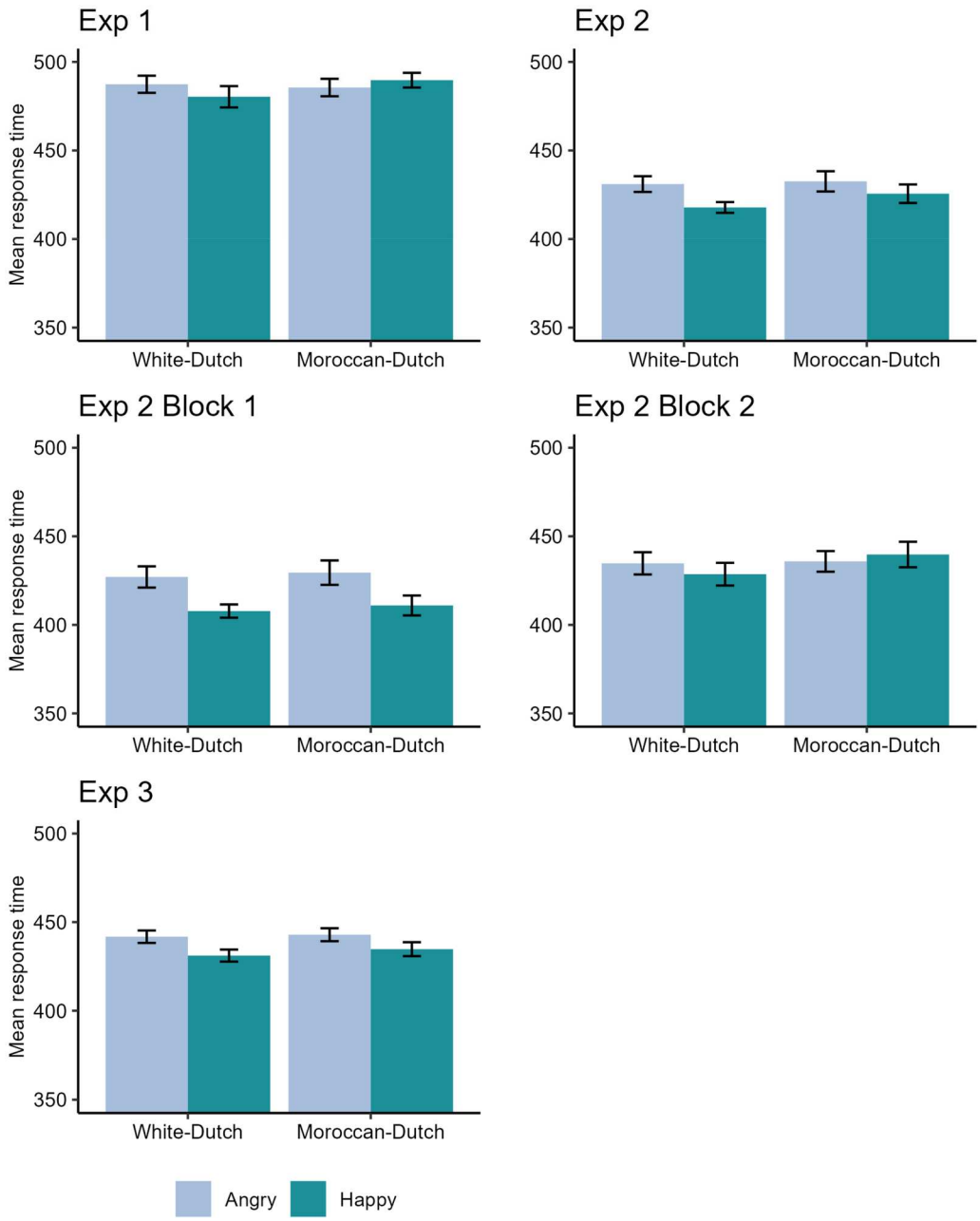

Figure 3. Results emotion recognition task.

Note: Mean response time in each of the conditions. Error bars reflect within-participants standard errors around those means.

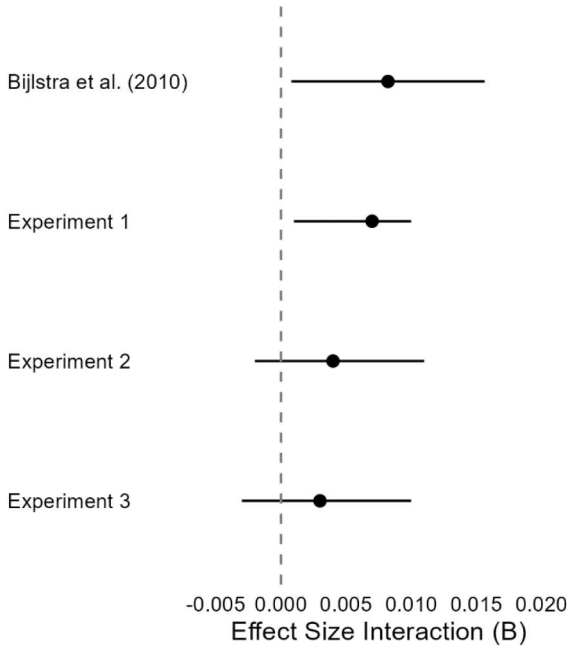

Figure 4. Forest plot emotion recognition task.

Note: Effect sizes of the interaction between social category and expression valence on response time for each experiment, including the original experiment by Bijlstra et al. (2010). A significant interaction effect was observed in Bijlstra et al. (2010) and Experiment 1, but not in Experiments 2 or 3. Error bars reflect confidence intervals around the effect size.

interaction between social category and expression valence on response time, $B = 0.003$, $SE = 0.004$, $F(1,53.62) = 1.23$, $p = .272$, 95% CI [−0.003, 0.010]. There was no significant main effect of social category on response time, $B = -0.001$, $SE = 0.003$, $F(1,55.62) = 0.21$, $p = .647$, 95% CI [−0.008, 0.005].

Although we did not find an interaction between social category and expression valence on response time, as was also the case in Experiment 2, we zoomed in on the response differences per emotion within each social category for the sake of completeness. Responses to happy emotional expressions ($M = 431$, $SD = 29$) were faster than responses to angry emotional expressions ($M = 442$, $SD = 30$) when displayed by White – Dutch faces, $B = 0.03$, $SE = 0.01$, $p = .007$. Although response times to happy emotional expressions were numerically faster ($M = 435$, $SD = 33$) than response times to angry emotional expressions ($M = 443$, $SD = 31$) when displayed by Moroccan – Dutch faces, this difference was not significant, $B = 0.015$, $SE = 0.01$, $p = .161$. Finally, there was no significant difference between White – Dutch and Moroccan – Dutch faces for happy emotional expressions, $B = -0.01$, $SE = 0.01$, $p = .266$, and angry emotional expressions, $B = 0.004$, $SE = 0.01$, $p = .655$ (see Figures 3 and 4).[4]

Explicit evaluation task. In contrast to our preregistered hypothesis, there was no significant main effect of social category on explicit evaluation, $B = -3.50$, $SE = 1.84$, $F(1,44.79) = 3.63$, $p = .063$, 95% CI [−7.10, 0.09] (see Figure 5). Although not significant,

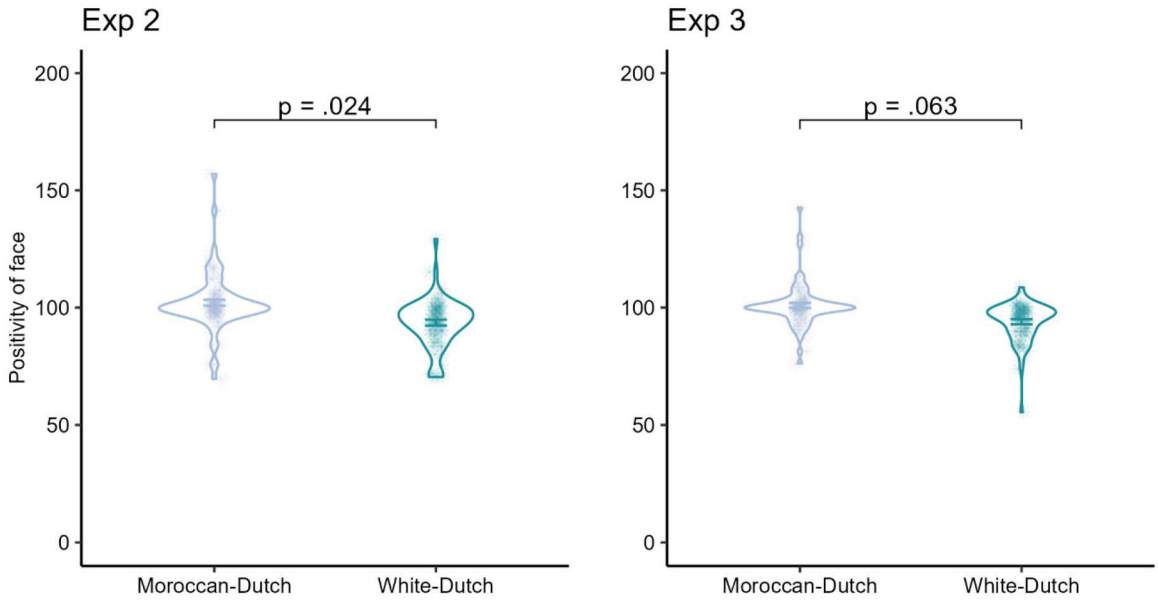

Figure 5. Results explicit evaluation task.

Note: Mean evaluation of the face for each social category. Error bars reflect within-participants standard errors around those means.

Moroccan – Dutch faces ($M = 100.92$, $SD = 9.07$) were numerically more positively evaluated than White – Dutch faces ($M = 93.92$, $SD = 9.07$).

Discussion

In Experiment 3 (and in line with results from Experiment 2), we found the preregistered main effect of valence of the expression on participants' response time needed to recognise emotional expressions, signalling an overall HFA. Contrary to our preregistered hypothesis, we did not find that participants' explicit evaluation of faces was more positive for Moroccan – Dutch faces compared to White – Dutch faces.

Based on these findings, we cautiously conclude that being subjected to a reinforcement learning task adjusted the HFA in the expected direction. That is, participants showed an overall HFA. These results suggest that evaluative associations of social categories are malleable with reinforcement learning, and that this affects emotion recognition.

General discussion

The current research explored whether reinforcement learning influences the HFA. We first replicated the commonly observed moderation effect of social categories on the HFA (Bijlstra et al., 2010; see also Becker et al., 2007; Bijlstra et al., 2019b; Craig et al., 2017; Craig & Lipp, 2017, 2018; Hugenberg, 2005; Hugenberg & Sczesny, 2006; Lipp et al., 2015). That is, we observed an HFA for White – Dutch faces and not for Moroccan – Dutch faces. Conducting this replication is important because if we do not find the original effect, attempting to influence it would be futile. Moreover, evaluative associations can change over time (Charlesworth & Banaji, 2019), and given that the original paper is 15 years old (Bijlstra et al., 2010), it is crucial to verify that the evaluative associations remain as expected in the current context.

Second, and importantly, our results suggest that reinforcement learning alters this differential effect of the HFA for different social categories. After participating in a reinforcement learning task, in which we linked Moroccan – Dutch faces to actions and rewards and White – Dutch faces to inactions and punishment avoidance, we no longer find evidence for the moderation effect of a face's social category on the HFA. Instead, we consistently show a main effect of the valence of the expression, such that responses to happy faces were faster than responses to angry faces, regardless of the social category of the face (White – Dutch or Moroccan – Dutch). This suggests that emotion recognition of existing social categories can be influenced by reinforcement learning.

These findings provide further support for the idea that evaluative social category associations are underlying HFA effects, i.e. the evaluative congruence account (e.g. Bijlstra et al., 2010; Hugenberg, 2005). Support for this account now comes from several sources. First, previous research shows that when evaluative associations are more accessible, people need less input to recognise evaluative consistent emotional expressions (Hugenberg & Bodenhausen, 2003). Second, Lindeberg et al. (2019) demonstrated the possibility of creating new evaluative associations for individual faces, influencing the HFA for these faces. Finally, the present research suggests that evaluative associations of existing social categories are malleable by a reinforcement learning task known to modify evaluative associations (Liu et al., 2025; van Lent et al., 2025), affecting subsequent emotion recognition. Together, these studies provide converging evidence for the idea that evaluative associations underlie the moderation of the HFA by social category membership of the target.

Initially, in Experiment 2, we expected reinforcement learning to affect both the HFA of White – Dutch and Moroccan – Dutch faces. In this, we expected an HFA for Moroccan – Dutch faces, and a smaller, reversal, or no HFA for White – Dutch faces. In two experiments, we found no evidence for this and observed an overall HFA instead. Apparently, it is very difficult to change learned positive evaluative associations people have with ingroup faces. Why would this be the case? One probable explanation is that people have a long learning history with ingroup faces. These are the faces people encounter most often. The longer the learning history, the stronger the evaluative association (Sherman, 1996). In addition, people prefer familiar faces (Zajonc, 1968), or new faces that look similar to previously seen ones (Zebrowitz et al., 2008), probably resulting in a more positive evaluative association for ingroup members. A complementary motivational explanation is that people strive to maintain a positive image of their ingroup (Tajfel & Turner, 1979; see also Tajfel, 1982). This may have contributed to the difficulty of making evaluative associations less positive. Overall, it seems that the learned positive associations with ingroup faces are either very strong, or people are highly motivated to maintain a positive view of

ingroup faces, or a combination of both. As a result, our computerised reinforcement learning task may not be powerful enough to change these learned or strongly motivated evaluative associations.

Instead of the reversed pattern of the often-observed interaction effect, we found consistent evidence for a general HFA. However, zooming in on the specific contrasts, there was no significant difference in response times for happy and angry faces among Moroccan – Dutch faces, while response times to happy faces were numerically faster than response times to angry faces. So, although the moderation of the HFA by social category membership appears to be statistically changed by reinforcement learning, the effects of reinforcement learning seem not that strong. Future research could investigate whether amplifying the effects of reinforcement learning, such as by increasing the number of learning trials, results in stronger effects on emotion recognition. In line with this idea, we found in Experiment 2 that the reinforcement learning task is initially effective, but after a short pause during which we swap the keys to recognise emotional expressions, the HFA pattern changes back to its original pattern. It seems that by inserting a pause in which we switch keys, we disrupt the effects of reinforcement learning on emotion recognition. One possible explanation is that task switching undermines the newly learned evaluative associations and thereby old associations come back to the fore more strongly (see Walther et al., 2019 for an overview of the impact of responses on attitudes).

Implications

The findings of the current research have both theoretical and practical implications. At the theoretical level, this research is the first to show that reinforcement learning affects emotion perception in members of different social categories. That is, we show that learning to act to Moroccan – Dutch faces to obtain rewards and learning to not act to White – Dutch faces to avoid punishments affects emotion perception on the category level. Importantly, we show that the effect of reinforcement learning processes generalises from neutral to emotional expressions: People learn about neutral emotional expressions during the reinforcement learning task, and the learning effects translate to different response times for happy and angry emotional expressions. Moreover, learning also generalises to new faces that were not present in the reinforcement learning task: People learn about five faces during the reinforcement learning task, and these learning effects translate to new and unfamiliar faces (13 faces) from the same social category when recognising emotional expressions.

At a more practical level, our findings further highlight the importance of positive contact situations for reducing prejudice (Allport, 1954; Paolini et al., 2024; Pettigrew & Tropp, 2006). While previous research suggests that the absence of contact can perpetuate negative evaluative associations (Allidina & Cunningham, 2021), the current study, on the other hand, adds to the large body of literature by demonstrating that positive contact may lead to more positive evaluative associations. Specifically, the reinforcement learning task we used can be seen as an abstract version of contact (or no contact) with a social category. Here, contact (depicted as go action decisions) that results in rewards seems to positively affect evaluative associations of outgroup members. That being said, monetary rewards are very different from rewards in everyday social interactions. Future research could investigate whether this effect persists in a more ecologically valid context when more social rewards, such as receiving smiles, are used.

Strengths and limitations

The present research has important strengths. All experiments are preregistered, we included a replication study, and the analysis strategies used are relatively new in the literature on recognising emotional expressions. This research also introduces theoretical advancements in the reinforcement learning literature, specifically in the area of aligning consequences with (in)actions decisions (i.e. action – valence asymmetries in learning). While most research in this area focuses on understanding these action – valence asymmetries itself (Guitart-Masip et al., 2012, 2014), our research examines the outcomes of these learning processes for emotion recognition. To date, only a few studies have explored the consequences of action – valence asymmetries in learning. For example, Liu et al. (2025) investigated its impact on food choice, and van Lent et al. (2025) examined its effects on individual impression formation. Here, we demonstrate for the first time that action – valence asymmetries in learning influence perceptions of social categories, and that these effects generalise to targets that participants did not learn about.

The current research also has its limitations. Most importantly, we did not test in a between-participant design whether the HFA of Moroccan – Dutch faces statistically differed after being subjected to the reinforcement learning task or not. We deliberately chose the current design without a control condition because including one led to a sample size that was far too large ($N = 2000$). However, we consistently provide novel evidence for a general HFA after participants conducted the reinforcement learning task.

Moreover, this experimental work was conducted in a controlled lab environment and not in real-life situations, limiting the external validity. Next, due to the experimental design in Experiments 2–3, the results obtained in the Explicit Evaluation Task are not as informative; it is unknown whether the difference in explicit evaluations can be attributed to other factors, such as demand characteristics. Additionally, we cannot exclude the possibility that the speed of emotion recognition might be influenced by other factors (such as the architecture of the face itself; Becker et al., 2007) besides the underlying evaluative associations related to social categories. Furthermore, there may be more optimal ways to analyze response time data, and future research could benefit from applying such alternative approaches (Tipples, 2022, 2023a; see also S2). Finally, it is unknown whether the absence of an HFA predicts discrimination in the real world. Therefore, the exact implications of a changed HFA pattern remain unclear. Future research could investigate whether HFA has predictive value for discriminatory behaviour.

Conclusion

Taken together, our results suggest that reinforcement learning affects evaluative associations, influencing subsequent emotion recognition: Combining actions with rewards for Moroccan – Dutch faces and combining inactions with avoidance of punishments for White – Dutch faces adjusted the HFA pattern. Since a large body of literature has consistently shown social category influences on the HFA – more negative evaluative associations lead to slower recognition of positive emotional expressions as positive – it is striking that we were able to adjust emotion recognition using a basic learning task. These findings provide new insights in the role of learning mechanisms to change emotion recognition. Our results suggest that aligning actions with rewards changes evaluative associations of outgroup members, consequently affecting emotion recognition.

Notes

1. Although it would be the most optimal design, we decided to avoid using a between-participants design that introduced the reinforcement learning task as a manipulation and no learning task as a control condition. A sample size calculation showed that this approach would require 2000 participants to detect the three-way interaction, which we deemed neither feasible (costs around 15,000 euros) nor ethical given the resources involved, and the benefits of conducting such a study did not outweigh the costs involved.
2. To give more insight into the data and for the sake of completeness, we report additional exploratory analyses for all experiments in S2. Here, we have also included alternative ways of analyzing the data that were recommended by a reviewer.
3. The random slope for social category was by mistake not preregistered. However, we deemed it more correct to include the random slope for social category and therefore we included it. The results with and without the random slope are the same.
4. For completeness, and despite being severely underpowered for this analysis, we compared Experiment 1 with Experiments 2–3 (see S2). This analysis revealed no significant three-way interaction between social category, expression valence, and experiment on response time, indicating that there is no evidence that the interaction between social category and expression valence differs between Experiment 1 and Experiments 2–3. Given the required sample size (see Footnote 1), this finding is not unexpected.

Acknowledgements

We thank Thijs Verwijmeren from Radboud University for helping with programming Experiments 2–3. We thank Erik van den Berge from Radboud University for helping with color adjustments to the fractal stimuli in Experiments 2–3. We thank Andrea van Langerak, Céline Epars, and Indy Slagmeulen from Radboud University for helping with data collection. We thank members of the BRACE lab group and RSCL lab group from Radboud University for useful discussions.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

Adams, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion*, *30*(2), 177–186. https://doi.org/10.1007/s11031-006-9020-2

Allidina, S., & Cunningham, W. A. (2021). Avoidance begets avoidance: A computational account of negative stereotype persistence. *Journal of Experimental Psychology: General*, *150*(10), 2078–2099. https://doi.org/10.1037/xge0001037

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

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., & Cikara, M. (2021). The social neuroscience of prejudice. *Annual Review of Psychology*, *72*(1), 439–469. https://doi.org/10.1146/annurev-psych-010419-050928

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

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

Bijlstra, G., Holland, R. W., Dotsch, R., & Wigboldus, D. H. J. (2019a). Stereotypes and prejudice affect the recognition of emotional body postures. *Emotion*, *19*(2), 189–199. https://doi.org/10.1037/emo0000438

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, *46*(4), 657–663. https://doi.org/10.1016/j.jesp.2010.03.006

Bijlstra, G., Kleverwal, D., van Lent, T., & Holland, R. W. (2019b). Evaluations versus stereotypes in emotion recognition: A replication and extension of Craig and Lipp's (2018) study on facial age cues. *Cognition and Emotion*, *33*(2), 386–389. https://doi.org/10.1080/02699931.2018.1526778

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Côté, S., & Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly*, *51*(1), 1–28. https://doi.org/10.2189/asqu.51.1.1

Craig, B. M., Koch, S., & Lipp, O. V. (2017a). The influence of social category cues on the happy categorisation advantage depends on expression valence. *Cognition and Emotion*, *31*(7), 1493–1501. https://doi.org/10.1080/02699931.2016.1215293

Craig, B. M., & Lipp, O. V. (2017). The influence of facial sex cues on emotional expression categorization is not fixed. *Emotion*, *17*(1), 28–39. https://doi.org/10.1037/emo0000208

Craig, B. M., & Lipp, O. V. (2018). Facial age cues and emotional expression interact asymmetrically: Age cues moderate emotion categorisation. *Cognition and Emotion*, *32*(2), 350–362. https://doi.org/10.1080/02699931.2017.1310087

Craig, B. M., Zhang, J., & Lipp, O. V. (2017b). Facial race and sex cues have a comparable influence on emotion recognition in Chinese and Australian participants. *Attention, Perception, & Psychophysics*, *79*(7), 2212–2223. https://doi.org/10.3758/s13414-017-1364-z

Crews, W. D., & Harrison, D. W. (1994). Cerebral asymmetry in facial affect perception by women: Neuropsychological effects of depressed mood. *Perceptual and Motor Skills*, *79*(3), 1667–1679. https://doi.org/10.2466/pms.1994.79.3f.1667

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Erickson, K., & Schulkin, J. (2003). Facial expressions of emotion: A cognitive neuroscience perspective. *Brain and Cognition*, *52*(1), 52–60. https://doi.org/10.1016/S0278-2626(03)00008-3

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences*, *18*(4), 194–202. https://doi.org/10.1016/j.tics.2014.01.003

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, *62*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Hackel, L. M., Kogon, D., Amodio, D. M., & Wood, W. (2022). Group value learned through interactions with members: A reinforcement learning account. *Journal of Experimental Social Psychology*, *99*, 104267. https://doi.org/10.1016/j.jesp.2021.104267

Heyes, C. M. (1994). Social learning in animals: Categories and mechanisms. *Biological Reviews*, *69*(2), 207–231. doi:10.1111/j.1469-185X.1994.tb01506.x

Hope, R. M. (2022). *Rmisc: Ryan miscellaneous* (Version 1.5.1) [R package]. https://CRAN.R-project.org/package=Rmisc

Hugdahl, K., Iversen, P. M., & Johnsen, B. H. (1993). Laterality for facial expressions: Does the sex of the subject interact with the sex of the stimulus face? *Cortex*, *29*(2), 325–331. https://doi.org/10.1016/S0010-9452(13)80185-2

Hugenberg, K. (2005). Social categorization and the perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion*, *5*(3), 267–276. https://doi.org/10.1037/1528-3542.5.3.267

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*(6), 640–643. https://doi.org/10.1046/j.0956-7976.2003.psci_1478.x

Hugenberg, K., & Sczesny, S. (2006). On wonderful women and seeing smiles: Social categorization moderates the happy face response latency advantage. *Social Cognition*, *24*(5), 516–539. https://doi.org/10.1521/soco.2006.24.5.516

Inquisit 6. (2022). [Computer software]. https://www.millisecond.com

Kassambara, A. (2023). *ggpubr: "ggplot2" Based publication ready plots* (Version 0.6.0) [R package]. https://CRAN.R-project.org/package=ggpubr

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*(6), 957–971. https://doi.org/10.1037/0022-3514.92.6.957

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* (Version 1.8.9) [R package]. https://CRAN.R-project.org/package=emmeans

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research*, *69*(1), 22–29. https://doi.org/10.1007/s00426-003-0157-2

Lindeberg, S., Craig, B. M., & Lipp, O. V. (2019). 2:0 for the good guys: Character information influences emotion perception. *Emotion*, *19*(8), 1495–1499. https://doi.org/10.1037/emo0000530

Lipp, O. V., Craig, B. M., & Dat, M. C. (2015). A happy face advantage with male caucasian faces: It depends on the company you keep. *Social Psychological and Personality Science*, *6*(1), 109–115. https://doi.org/10.1177/1948550614546047

Liu, H., Quandt, J., Zhang, L., Kang, X., Blechert, J., van Lent, T., Holland, R. W., & Veling, H. (2025). Shaping food choices with actions and inactions with and without reward and punishment. *Appetite*, *208*, 107950. https://doi.org/10.1016/j.appet.2025.107950

Loy, A., & Hofmann, H. (2014). HLMdiag: A suite of diagnostics for hierarchical linear models in R. *Journal of Statistical Software*, *56*(5), 1–28. https://doi.org/10.18637/jss.v056.i05

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach-and avoidance-related behaviors. *Emotion*, *5*(1), 119–124. https://doi.org/10.1037/1528-3542.5.1.119

Marsh, A. A., Kozak, M. N., & Ambady, N. (2007). Accurate identification of fear facial expressions predicts prosocial behavior. *Emotion*, *7*(2), 239–251. https://doi.org/10.1037/1528-3542.7.2.239

Martin, D., Hutchison, J., Konopka, A. E., Dallimore, C. J., Slessor, G., & Swainson, R. (2024). Intergroup processes and the happy face advantage: How social categories influence emotion categorization. *Journal of Personality and Social Psychology*, *126*(3), 390–412. https://doi.org/10.1037/pspa0000386

Mathôt, S., Siebold, A., Donk, M., & Vitu, F. (2015). Large pupils predict goal-driven eye movements. *Journal of Experimental Psychology: General*, *144*(3), 513–521. https://doi.org/10.1037/a0039168

Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modelling. *Psychological Methods*, *27*(6), 1014–1038. https://doi.org/10.1037/met0000330

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*(9), 1095–1102. https://doi.org/10.1038/nn1968

Paolini, S., Gibbs, M., Sales, B., Anderson, D., & McIntyre, K. (2024). Negativity bias in intergroup contact: Meta-analytical evidence that bad is stronger than good, especially when people have the opportunity and motivation to opt out of contact. *Psychological Bulletin*, *150*(8), 921–964. https://doi.org/10.1037/bul0000439

Pedersen, T. L. (2024). *patchwork: The composer of plots* (Version 1.1.2) [R package]. https://CRAN.R-project.org/package=patchwork

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

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

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology*, *70*(6), 1126–1141. doi:10.1037/0022-3514.70.6.1126

Signorell, A. (2023). *DescTools: Tools for descriptive statistics* (Version 0.99.50) [R package]. https://CRAN.R-project.org/package=DescTools

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2023). *afex: Analysis of factorial experiments* (Version 1.3-1) [R package]. https://CRAN.R-project.org/package=afex

Stalans, L., & Wedding, D. (1985). Superiority of the left hemisphere in the recognition of emotional faces. *International Journal of Neuroscience*, *25*(3–4), 219–223. https://doi.org/10.3109/00207458508985373

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*(1), 1–39. https://doi.org/10.1146/annurev.ps.33.020182.000245

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Tipples, J. (2022). No need to collect more data: Ex-Gaussian modelling of existing data (Craig & Lipp, 2018) reveals an interactive effect of face race and face sex on speeded expression recognition. *Cognition and Emotion*, *36*(7), 1440–1447. https://doi.org/10.1080/02699931.2022.2120850

Tipples, J. (2023a). Analyzing facial expression decision times: Reaction time distribution matters. *Emotion*, *23*(3), 688–707. https://doi.org/10.1037/emo0001098

Tipples, J. (2023b). When men are wonderful: A larger happy face facilitation effect for male (vs. female) faces for male participants. *Emotion*, *23*(7), 2080–2093. https://doi.org/10.1037/emo0001221

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

van Lent, T., Bijlstra, G., Holland, R. W., Bijleveld, E., & Veling, H. (2025). On rewarded actions and punishment-avoidant inactions: The action – valence asymmetry in face perception. *Journal of Experimental Social Psychology*, *119*, 104754. https://doi.org/10.1016/j.jesp.2025.104754

Verkuyten, M., & Zarembe, K. (2005). Interethnic relations in a changing political context. *Social Psychology Quarterly*, *68*(4), 375–386. https://doi.org/10.1177/019027250506800405

Walther, E., Blask, K., Halbeisen, G., & Frings, C. (2019). An action control perspective of evaluative conditioning. *European Review of Social Psychology*, *30*(1), 271–310. https://doi.org/10.1080/10463283.2019.1699743

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Yoo, S. H., & Noyes, S. E. (2016). Recognition of facial expressions of negative emotions in romantic relationships. *Journal of Nonverbal Behavior*, *40*(1), 1–12. https://doi.org/10.1007/s10919-015-0219-3

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zebrowitz, L. A., White, B., & Wieneke, K. (2008). Mere exposure and racial prejudice: Exposure to other-race faces increases liking for strangers of that race. *Social Cognition*, *26*(3), 259–275. https://doi.org/10.1521/soco.2008.26.3.259