SHAPING SOCIAL EVALUATIONS AND BEHAVIOR THROUGH INSTRUMENTAL LEARNING

Tjits van Lent

Behavioural
Science
Institute

Shaping Social Evaluations and Behavior Through Instrumental Learning

Tjits van Lent

ISBN 978-94-6473-985-5

Design and lay-out
Paulien Varkevisser | fotografie & vormgeving, Nijmegen
www.paulienvarkevisser.com

Printing
Ipskamp Printing B.V., Enschede, the Netherlands

Shaping Social Evaluations and Behavior Through Instrumental Learning

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 8 januari 2026
om 14.30 uur precies

door

Tjits van Lent
geboren op 16 augustus 1994
te Amsterdam

Promotoren:
Prof. dr. R.W. Holland
Prof. dr. H.P. Veling (Wageningen University & Research)

Copromotoren:
Dr. G. Bijlstra
Dr. E.H. Bijleveld

Manuscriptcommissie:
Prof. dr. K. Roelofs
Prof. dr. D.M. Amodio (Universiteit van Amsterdam)
Prof. dr. K. Kawakami (York University, Canada)

Shaping Social Evaluations and Behavior Through Instrumental Learning

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board
to be defended in public on

Thursday, January 8, 2026
at 2:30 pm

by

Tjits van Lent
born on August 16, 1994
in Amsterdam (the Netherlands)

PhD supervisors:
Prof. dr. R.W. Holland
Prof. dr. H.P. Veling (Wageningen University & Research)

PhD co-supervisors:
Dr. G. Bijlstra
Dr. E.H. Bijleveld

Manuscript Committee:
Prof. dr. K. Roelofs
Prof. dr. D.M. Amodio (University of Amsterdam)
Prof. dr. K. Kawakami (York University, Canada)

Table of contents

PREFACE 9

CHAPTER 1 **General introduction** 10

CHAPTER 2 **Instrumental learning shapes individual evaluations and preferences** 22
Supplemental materials 57

CHAPTER 3 **Instrumental learning shapes social category evaluations and emotion recognition** 58
Supplemental materials 83

CHAPTER 4 **Instrumental learning shapes consideration sets and preferences** 90
Supplemental materials 107

CHAPTER 5 **General discussion** 112

CHAPTER 6 **Miscellaneous** 124
References 126
Research data management, transparency, and privacy 147
English summary 148
Dutch summary | Nederlandse samenvatting 151
Acknowledgements | Dankwoord 154
About the author 160
List of academic publications 161

PREFACE

When I was a kid on vacation, I used to play the same song repeatedly on my mp3 player so that listening to it later would bring back that vacation feeling. For example, during my first flying experience (a vacation to Rome), I played the song ‘Everyday I Work on the Road’ by Voicst on repeat during takeoff. Much later, on a vacation in Greece, I played the song ‘Back Down South’ by Kings of Leon over and over. These songs still bring back that vacation feeling. Little did I know that back then, I was already engaging with learning theory.

My interest in how people learn began at a young age, even if I didn’t realize it at the time. The way people learn and its consequences are exactly what my dissertation is about. In my research, I have shifted my focus from how I learn to how people learn about others. I explored this within the context of an important societal theme: prejudice and discrimination.

CHAPTER 1

General introduction

"No one is born hating another person because of the color of his skin, or his background, or his religion. People must learn to hate, and if they can learn to hate, they can be taught to love, for love comes more naturally to the human heart than its opposite."

– Nelson Mandela (1994)

Prejudice and discrimination remain widespread both globally and within the Netherlands. For example, recent research indicates discrimination in hiring decisions (Quillian & Lee, 2023), housing decisions (Auspurg et al., 2019), judicial decisions (Galvan et al., 2024), and police stop-and-check decisions (Brown & van Eijk, 2021). Prejudice refers to general affective evaluations[1] (likes and dislikes) toward a social category[2] and its members (Amodio, 2014; Dovidio et al., 1997; Fazio et al., 1995). When someone talks about prejudice in everyday life, they are usually referring to a negative affective evaluation—for example, a negative evaluation of individuals from a certain ethnic category. If someone acts upon their prejudice, this is called discrimination: Unjust and differential treatment of members of different social categories (Allport, 1954). Discrimination has severe consequences for those affected, impacting their physical and mental health (Andriessen et al., 2020; Centraal Bureau voor de Statistiek, 2024) as well as their financial situation (Centraal Bureau voor de Statistiek, 2024). For example, discrimination is associated with higher levels of depression, anxiety, and psychological stress (Andriessen et al., 2020), as well as higher risks of cardiovascular diseases, obesity, and high blood pressure (see Williams et al., 2019 for a review). Thus, unfortunately, discrimination and prejudice are highly prevalent and negatively impact lives.

Hence, a crucial question is how to effectively mitigate these phenomena. The foremost intervention strategy to reduce prejudice is promoting intergroup contact (Paluck et al., 2019). Unsurprisingly, intergroup contact as an intervention strategy has received considerable attention (Allport, 1954; Lemmer & Wagner, 2015; Paluck et al., 2019; Pettigrew & Tropp, 2006). Since intergroup contact has mostly been researched in an intervention context, the focus has been more on positive contact rather than on both positive and negative contact (Paolini & McIntyre, 2019). Recently, new attention has been devoted to the valence (positive vs. negative) of contact, suggesting that contact with positive consequences typically reduces prejudice, while contact with negative consequences increases prejudice (Paolini et al., 2024; see also Aberson & Gaffney, 2009). Still, the precise way in which positive and negative consequences of behavior influence prejudice remains poorly understood, as the specific preconditions of contact that increase or decrease prejudice are still unknown (English, 2024; Paluck et al., 2019).

The consequences for behavior caused by contact can be understood as a form of instrumental learning (Hackel et al., 2015; Lott & Lott, 1974). That is, people learn

[1] In this dissertation, I use different terms to describe evaluations, adjusting the language based on what is appropriate in the field of publishing. In Chapters 2 and 4, I refer to evaluations as (subjective) values. In Chapter 3, I refer to evaluations as evaluative associations.

[2] In this dissertation, I use the terms social category and social group interchangeably.

about others through the positive or negative consequences of their behavior. These consequences shape people's evaluations of others and future behaviors toward them, thereby presumably influencing prejudice and discrimination. In line with this thinking, instrumental learning has been proposed as a strong candidate mechanism to understand and change prejudice (Amodio, 2019; Amodio & Cikara, 2021). **In general, the main aim of this dissertation is to gain a better understanding of the instrumental learning processes that contribute to how people evaluate individuals and social categories.** In this introductory chapter, I will outline the theoretical considerations that inspired the experiments in the empirical chapters of this dissertation, introduce both chapter-specific and overarching research questions, and provide a brief overview of each chapter.

Learning About Others Through Consequences of Actions

Instrumental learning is a learning process in which people learn optimal responses based on outcomes that follow prior responses through trial-and-error (Sutton & Barto, 2018). In this, responses followed by pleasant consequences are likely to be repeated, whereas responses followed by unpleasant consequences are less likely to be repeated (operant conditioning; Skinner, 1938; the law of effect; Thorndike, 1927). This form of trial-and-error learning is beneficial as it allows people to flexibly adapt to their environment. Instrumental learning is an active learning process: Rather than passively receiving information, people perform actions and learn from the consequences of those actions. Such active learning boosts cognitive processing, such as memory encoding (Yebra et al., 2019), memory recall (Katzman & Hartley, 2020) and visual attention (Bamford et al., 2020), potentially resulting in strong effects on behavior (cf. Dubinsky & Hamid, 2024).

Theoretically, instrumental learning may play an important role in shaping evaluations and behaviors toward people. First, it may play a role at the individual level. During social interactions, people learn about others through the responses of their interaction partners. The nature of this response potentially influences the evaluations people hold of their interaction partners, which, in turn, shapes their subsequent responses. For example, when a stranger smiles in response to your greeting, you may be more likely to like the stranger and greet them again in the future. Conversely, if a stranger ignores you, the opposite may be true. Second, instrumental learning may play a role at the social category level. During social interactions, individuals can learn about the social category to which their interaction partner belongs. For example, when an outgroup member smiles in response to your greeting, you may think more positively about the outgroup as a whole and greet other members of that outgroup in the future. In this way, the nature of the response generalizes to the group level and potentially influences the evaluations individuals hold of the social category, which, in turn, shapes their responses to novel members of the same social category.

Recent empirical research has examined the interplay between instrumental learning and the shaping of evaluations and responses to individuals from different social categories (for a review see Amodio, 2025). This collection of studies provides initial evidence that instrumental learning influences both evaluations and behavior. Regarding evaluations, research shows that consequential interactions change evaluations of the interaction partner itself (Hackel et al., 2019), as well as evaluations of the social category (Hackel, Kogon, et al., 2022). Moreover, regarding intergroup behavior, research demonstrates that consequential interactions with members of a social category influence preferences for the interaction partner itself (Traast et al., 2024, 2025) and future interactions with novel members of the same social category with whom participants had never interacted (Allidina & Cunningham, 2021; Hackel, Kogon, et al., 2022). Thus, there are indications that instrumental learning changes people's evaluations and behavior toward social categories and their members.

The Role of Inactions in Learning

Although recent research has examined the influence of having interactions on evaluations and behavior, little is known about the effects of having *no* interactions or *no* contact on evaluations and behavior. Often, individuals choose *not* to interact with others. For example, people may decide not to greet a stranger, which could lead to certain consequences (e.g., a frown), thereby potentially influencing evaluations (e.g., dislike of the stranger or the stranger's social category) and future behavior (e.g., the decision to greet or not greet the stranger or another person from the same social category). Various lines of literature provide indications that such inactions play an important and distinct role in social interactions (e.g., Allidina & Cunningham, 2021; Chotpitayasunondh & Douglas, 2018; Denrell, 2005; Hehman & Neel, 2024; Plant & Devine, 2003; Szekeres et al., 2023; Zadro & Gonsalkorale, 2014). However, a systematic examination of the direct impact of consequential inactions on evaluations and behavior is lacking.

The idea that inactions play an important and distinct role in social interactions aligns with findings from the basic learning processes literature. That is, some of this literature focuses specifically on consequential inaction learning and how it differs from consequential action learning (e.g., Guitart-Masip et al., 2012). Note that these studies investigate the learning process *itself*, rather than the effects of consequential inactions on evaluations and behaviors. Evidence has accumulated that learning (in)actions can be enhanced when different types of consequences (rewards versus punishments) are matched to actions versus inactions (Cavanagh et al., 2013; Chowdhury et al., 2013; Guitart-Masip, Duzel, et al., 2014; Guitart-Masip, Economides, et al., 2014; Guitart-Masip et al., 2012; Mkrtchian et al., 2017; Moutoussis et al., 2018; Richter et al., 2014; Swart et al., 2017). More specifically, this work has revealed an *action–valence asymmetry* in learning. In this, people learn to act (e.g., respond to a stimulus) more easily to stimuli that elicit rewards, while they have relatively high difficulty learning actions to stimuli that lead to avoidance of punishments. Conversely, people learn inactions (e.g., withholding a response to a stimulus) better when they lead to the avoidance of punishments, while they have relatively high difficulty learning inactions to stimuli that elicit rewards.

The action–valence asymmetry in learning is explained by the interaction between instrumental learning and a hard-wired *Pavlovian bias* (Guitart-Masip, Duzel, et al., 2014). Due to the Pavlovian bias, reward-predictive stimuli prepare a tendency to *act*, while punishment-predictive stimuli prepare a tendency to *not act* (Boureau & Dayan, 2011; Glickman & Schiff, 1967; Huys et al., 2011). For example, individuals tend to act toward a friendly-looking person, while they tend to not act toward an unfriendly-looking person. Pavlovian bias is essential for basic human functioning as it helps individuals select actions quickly, with low computational costs and minimal errors. Because Pavlovian responses are automatically triggered in the presence of stimuli, they are relatively inflexible and could potentially result in inappropriate behavioral responses (Breland & Breland, 1961; Hershberger, 1986). For example, sometimes it is necessary to not act toward a friendly-looking person, such as when that friendly-looking person has bad intentions, but this may not be one's first response. Thus, due to the Pavlovian bias, rewards hinder the learning of inactions, whereas punishments hinder the learning of actions, leading to the action–valence asymmetry in learning.

Although this literature on basic learning processes provides insight into the nature of inaction learning itself, the social consequences of inaction learning remain unknown. Specifically, it is unclear how learning consequential inactions translates into social evaluations and behavior. Due to the action–valence asymmetry in learning, the impact of inactions on social evaluations and behavior may depend on the nature of the corresponding consequence. Aligning inactions with the avoidance of punishment when learning about people may have a strengthening effect on social evaluations and behavior, since research from the food domain shows that adding consequences to approach–avoidance training strengthens the effect on evaluations and choices (Van Dessel et al., 2018) and rewarding actions and avoiding punishment for inactions amplifies the effects on preferences (Liu et al., 2025). Yet, how consequential inactions influence evaluations and behavior in the social domain remains unexplored. **Therefore, to contrast and compare consequential inactions with actions, the first research question of this dissertation focuses on the unique contribution of inactions, over and above reward and punishment, in shaping evaluations and behavior toward individuals and social categories.**

Generalization

Returning to research on intergroup contact, a main challenge with using intergroup contact as an intervention is generalization (see e.g., Lowe, 2025). Studies on contact observe more individual than social category effects; that is, contact is more effective at changing evaluations and behavior toward individuals met in person than toward a social category. For intergroup contact to be an effective intervention in reducing prejudice and discrimination, it is of central importance to focus on generalization. Thus, it is important to investigate whether the consequences of instrumental learning generalize to new targets from the same social category. For example, whether a positive social interaction with an individual from a particular social category positively influences evaluations of or behaviors toward new individuals from the same social category whom someone has not yet encountered, or whether a negative social interaction with an individual from a particular social category negatively influences evaluations of a social category as a whole, which may lead to *not* engaging in future social interactions with new individuals from the same social category whom someone has not yet encountered. **Thus, a second research** question is whether instrumental learning generalizes to evaluations and behavior toward others from the same social category.

Consequences of Instrumental Learning for Evaluations and Behavior

There are various ways to measure evaluations and behavior in the context of prejudice and discrimination, which enable the investigation of the aforementioned research questions. The most straightforward way to measure prejudice is to ask for someone's evaluation of members of another social category using self-report measures, such as survey questions. Here, evaluations are explicit statements regarding, for example, how positively participants rate members of a particular social category. These evaluations can predict behavior, such as preferences, i.e., which option someone prefers over various alternatives (cf. Hascher et al., 2021). However, sometimes there are reasons to believe that people may be unwilling or unable to report their true evaluations. For example, asking someone to evaluate a member of another social category may elicit socially desirable responses, as respondents might feel it is unacceptable to express their true feelings. Therefore, researchers use not only explicit self-reports but also behavioral measures of prejudice and discrimination that are less controllable (Fazio & Olson, 2003), such as speeded responses (cf. Chen et al., 2019). Investigating such behavioral outcome measures allows to contribute to psychology as the science of behavior and broadens its implications beyond mere evaluations.

In speeded responses, the presence of time pressure makes answers less controllable and, consequently, less prone to social desirability. Therefore, under greater time pressure, people base their decisions on more basic information that is readily accessible (Fazio, 1990; Linn et al., 2024), such as their more automatic evaluation of a particular social category. For example, previous research shows that under time pressure, doctors behaved more in line with their racial biases (Stepanikova, 2012), and hiring managers behaved more in line with their gender biases (Lucas et al., 2021). There are many ways to use speeded responses as an outcome measure (Greenwald & Lai, 2020). Here, I will elaborate on two methods. First, a straightforward approach, namely, speeded preferences, where people indicate their preference for one option (e.g., one person) over another. Previous work has shown that this task is a valid and reliable way to tap into automatic processes (Veling et al., 2017). A second method for using speeded responses to investigate the effects of instrumental learning on prejudice and discrimination is to focus on emotion recognition. Emotion recognition is an important building block of social interactions. Emotional expressions serve a strong communicative function, as people can convey their feelings about certain situations through these expressions. Importantly, the ease with which emotional expressions are recognized is influenced by the social category of the target (e.g., Elfenbein & Ambady, 2002; Hugenberg & Bodenhausen, 2003). Why is this the case? The most prominent theoretical account in the literature suggests that the recognition of emotional expressions is facilitated or inhibited by the accessibility of evaluations related to social categories (e.g., Bijlstra et al., 2010; Craig, Zhang, et al., 2017; Hugenberg, 2005; Hugenberg & Sczesny, 2006). Relatively stronger positive evaluations facilitate the recognition of positive emotional expressions, whereas relatively stronger negative evaluations inhibit the recognition of positive emotional expressions. Since the recognition of emotional expressions is influenced by the social category of the target, and the ease of recognition is less prone to demand characteristics than explicit evaluations, it may serve as a reliable outcome measure related to prejudice.

Consequences of Instrumental Learning for Consideration Sets

To broaden the knowledge about the implications of instrumental learning and gain a better understanding of the basic instrumental learning processes underlying discrimination in real-world decisions, such as in recruitment procedures (Lancee, 2019; Quillian & Lee, 2023; Thijssen et al., 2021), the role of instrumental learning in the *pre-decision phase* is examined. The *pre-decision phase* refers to how people consider their choice options before making a decision. When making decisions, people are unable to carefully evaluate all available choice options due to cognitive limitations and time constraints (Krajbich et al., 2010; Oeberst & Imhoff, 2023; Simon, 1955). Consequently, they only evaluate a few choice options out of all possible choice options. That is, when making a decision, people create a *consideration set* of choice options they deem relevant to the decision from all possible choice options (Howard & Sheth, 1969). How do people create this consideration set? Previous research in the food domain demonstrates that positively evaluated choice options are more likely to enter the consideration set than negatively evaluated choice options (Morris et al., 2021; see also Posavac et al., 1997), even when the evaluations are irrelevant to the decision at hand.

This raises questions about how this principle may apply when making decisions about people. When constructing a consideration set in a social setting, two types of evaluations can influence the consideration set. The first type is evaluations based on someone's social category (i.e., prejudice or social category-based evaluations; Fiske & Neuberg, 1990). The second type is evaluations based on someone's individual characteristics, such as their traits or behavior (i.e., individual-based evaluations). Together, learning about these two types of evaluations may influence the construction of the consideration set. How do the two types of evaluations interplay? For example, with equal individual-based evaluations, does learning about members from a negatively evaluated social category lead to a lower probability of their inclusion in the consideration set compared to learning about members from a more positively evaluated social category? **To gain a better understanding of this pre-decision phase, a third research question examines whether learning social category-based and individual-based evaluations influences the construction of the consideration set.** This provides more insight into discriminatory decision-making and offers guidance on which stage of the decision-making process to intervene—thereby aiming to reduce biases in decision-making.

The Role of Prejudice in Learning Actions and Inactions

So far, the consequences of instrumental learning on evaluations and behavior toward individuals and social categories have been discussed. However, when focusing on instrumental learning *itself*, is it dependent on the social category of the target? If reward-predictive stimuli prepare actions and punishment-predictive stimuli prepare inactions due to the Pavlovian bias, this raises the question of whether social category-based evaluations of others (positive or negative) influence the learning *itself*, specifically the learning of actions and inactions. Following Pavlovian bias logic, if a target is more negatively evaluated based on their social category, one may anticipate a punishment, which could presumably hinder action learning and facilitate inaction learning. Conversely, if a target is more positively evaluated based on their social category, one may anticipate a reward, which could presumably hinder inaction learning and facilitate action learning. This implies that people would perform worse at learning actions toward people from negatively evaluated social categories, while they may perform relatively better at learning inactions toward these individuals. The reverse holds for people from positively evaluated social categories. Whether social category-based evaluations generalize to learning (in) actions regarding novel members of the same social category with whom people have never interacted remains unexplored. **To explore this idea, the fourth research question examines whether the target's social category influences (in)action learning itself.** For example, do people find it more difficult to learn actions toward a target from a negatively evaluated social category? Conversely, do people find it more difficult to learn inactions toward a target from a positively evaluated social category?

The Present Dissertation

The present dissertation aims to gain a better understanding of the instrumental learning processes that contribute to how people evaluate individuals and social categories. To that end, this dissertation systematically and experimentally investigates instrumental learning and its consequences for social evaluations and behavior through four research questions. Table 1.1 provides an overview of the relationship between the empirical chapters and the research questions that this dissertation aims to address. The order of the research questions in this table does not match the order of the empirical chapters, as the research questions are arranged according to their importance. Please note that I do not measure *actual* social interactions; instead, I aim to study processes related to social interactions within an experimental setup, allowing me to maintain full control over all parameters. To achieve the main aim, I have conducted 11 experiments, which are presented in the three empirical chapters of this dissertation. All research is preregistered, and all data, R scripts, experimental scripts, and, where possible, stimulus materials are publicly available (see Research data management for repository links). The chapters of this dissertation are based on articles that have been published or submitted to peer-reviewed scientific journals. Consequently, each chapter can be read independently and, in any order, leading to some overlap in the descriptions of the theoretical underpinnings.

In light of rigorous science, each chapter begins with a replication, upon which I subsequently build. A replication "refers to testing the reliability of a prior finding with different data" (p. 721; Nosek et al., 2022). In other words, it involves repeating the same experiment to see if the same result occurs. I ran these replications for several reasons. First, I believe that a solid foundation is necessary before further building upon knowledge (Nosek et al., 2022). Conducting replications gives me confidence in the credibility of previous findings and confirms the continued relevance of the results in the present. Moreover, replication provides evidence of generalizability (Nosek & Errington, 2020) and convergence. In sum, by conducting these replications, I aim to contribute to trustworthy and solid science.

In all chapters, instrumental learning and its consequences for social evaluations and/or behavior are central. In **Chapter 2,** I investigate instrumental learning and its consequences for social evaluations and behavior at the *individual* level. White–Dutch and Moroccan–Dutch targets are investigated here because, in the Netherlands, the Moroccan–Dutch community is one of the most negatively prejudiced social categories (Centraal Bureau voor de Statistiek, 2024; Verkuyten & Zarembe, 2005; the same applies to Chapter 3). Specifically, I investigate whether the action–valence asymmetry in learning influences evaluations and speeded preferences regarding both Moroccan–Dutch and White–Dutch individual targets across four experiments (RQ1). Crucially, I focus on inactions because they may have a unique effect on evaluations and behavior and have largely been ignored in social psychological literature. I contrast and compare the consequences of rewarded actions, punishment-avoidant actions, rewarded inactions, and punishment-avoidant inactions on individual evaluations and behavior. In doing so, I aim to provide new evidence for the influence of consequential inactions on social evaluations and behavior, while conceptually replicating the influence of consequential actions. In an exploratory fashion, I investigate whether the social category of the target influences learning (RQ4).

Second, I expand on this in **Chapter 3**, where I investigate instrumental learning and its consequences for social evaluations and behavior at the *social category* level, rather than at the individual level. In addition to measuring evaluations, I assess social behavior, specifically the recognition of emotional expressions. Across three experiments, I investigate whether instrumental learning influences the recognition of emotions of multiple ingroup (White–Dutch) and outgroup (Moroccan–Dutch) targets (RQ1). To achieve the strongest possible effect, I use the most impactful instrumental learning conditions from Chapter 2. I link the condition with the strongest positive effect to outgroup targets and the condition with the strongest negative effect to ingroup targets. Importantly, I investigate whether potential learning effects generalize beyond the target with whom someone interacts (RQ2).

Next, I move to investigating instrumental learning and its consequences in social decision-making. Specifically, I examine whether instrumental learning influences the pre-decision phase, that is, how people are considered when making decisions in **Chapter 4**. Across four experiments, I investigate the role of individual-based and social category-based evaluations of individuals from *non-existing* social categories in the construction of consideration sets (RQ3). For example, are members with more positive individual-based or social category-based evaluations more likely to be considered than those with less positive evaluations? Additionally, when individual-based evaluations are equal, do people consider individuals from positively evaluated social categories more than those from negatively evaluated social categories? In addition to measuring consideration sets, I assess speeded preferences, where participants indicate their preference for novel targets from the social categories they have just learned about, aiming to examine potential generalization effects (RQ2).

Finally, in **Chapter 5,** I provide a summary of the main findings, discuss and integrate the most relevant findings of this dissertation, and highlight future directions. Taken together, this dissertation presents systematic experimental work to increase understanding of the instrumental learning processes that contribute to how people evaluate individuals and social categories.

Table 1.1

Overview of the Research Questions and the Empirical Chapters That Address Them

Research Questions	Chapter 2	Chapter 3	Chapter 4
1. What is the unique contribution of inactions over and above reward and punishment in shaping evaluations and behavior?	x	x	
2. Does instrumental learning generalize to evaluations and behaviors toward others from the same social category?		x	x
3. Does instrumental learning about social category-based and individual-based evaluations influence the construction of consideration sets?			x
4. Does the target's social category influence (in)action learning?	x		

CHAPTER 2

Instrumental learning shapes individual evaluations and preferences

This chapter is based on:

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

Abstract

Although social interactions are ubiquitous, people often choose *not* to interact with others—for example, people may choose to not greet a stranger, to not talk to a colleague at work, or to ignore a text message from a friend. Here, we systematically investigate how people's actions, inactions, and their consequences (rewards and punishments) affect impressions. In four preregistered experiments (N = 240), we used a reinforcement learning go/no-go task, in which people learned to act or not act to images of fractals/faces to obtain rewards or avoid punishments. Findings replicated the action–valence asymmetry in learning (Experiments 2.1–2.4): People more easily learned to act when acting led to the attainment of rewards (vs. the avoidance of punishments), while people learned inactions more easily when these inactions led to the avoidance of punishments (vs. the attainment of rewards). Our experiments demonstrate that these action–valence asymmetries extend to social stimuli (Experiment 2.2 ingroup faces; Experiment 2.3 outgroup faces; Experiment 2.4 ingroup and outgroup faces) and that they affect subsequent impressions. That is, people evaluated faces most positively when acting had previously led to the attainment of rewards; people evaluated faces most negatively when *not* acting had previously led to the avoidance of punishment. We discuss our findings in light of the approach–avoidance literature. This work has implications for our understanding of the role of inactions in social contexts: It shows evidence that inactions lead to less positive impressions than actions, over and above the effect of punishment signals.

Keywords: reinforcement learning, intergroup relations, Pavlovian bias, value-based decision making, impression formation

Open Science Practices: Open Data, Open Materials, Preregistered

Introduction

People's lives are filled with social interactions: On average, people have 12 to 16 social interactions per day (Del Valle et al., 2007; Zhaoyang et al., 2018). In the past decades, it has become clear that these interactions strongly affect how people evaluate others. For example, engaging in social interactions leads to more liking of others (Hackel et al., 2019; Jones et al., 2011; Lott & Lott, 1974), more positive attitudes and behaviors toward outgroup members (Lowe, 2021; Mousa, 2020; Pettigrew & Tropp, 2006; Shook & Fazio, 2008), and more interpersonal attraction (Insko & Wilson, 1977; Reis et al., 2011). This body of research shows that interactions—and their positive and negative consequences—affect evaluations of other people. Yet, although social interactions are frequent, people also often choose *not* to interact with someone—for example, they may choose to not greet a stranger, to not talk to a colleague at work, or to ignore a text message from a friend—which may lead to rewards (or punishments) as well. This feature of social interactions raises the question of how the combination of actions, inactions, and consequences shapes impressions of other people. Is there perhaps something special about inactions?

Research from cognitive psychology and neuroscience shows that inactions are special, at least in the sense that they are subject to distinct learning mechanisms than actions (Collins & Frank, 2014; Frank, 2005). It seems plausible that inactions play an important, distinct role in social interactions (Allidina & Cunningham, 2021; Chotpitayasunondh & Douglas, 2018; Denrell, 2005; Fazio et al., 2004; Plant, 2004; Plant & Devine, 2003; Szekeres et al., 2023; Zadro & Gonsalkorale, 2014). Yet, the basic mechanisms that underpin the role of inactions are not yet well understood. Here, building on prior work on inactions, we contrast and compare the consequences of rewarded actions, punished actions, rewarded inactions, and punished inactions. We focus on a central outcome in person perception: How do people's actions, inactions, and their consequences affect impressions of other people's faces?

Non-consequential (in)actions as Value Input

Decision-making researchers generally assume that people, when they make value-based decisions (e.g., decisions to interact with other people), first assign subjective value to different choice options (Berkman, 2018; Berkman et al., 2017; Gluth et al., 2012; Gold & Shadlen, 2007; Rangel et al., 2008; Verbruggen et al., 2014), and then choose the highest-valued choice alternative. Interestingly, mere (in)actions[3] without any external reinforcement are sufficient to influence subjective value of choice options (Veling et al., 2022). Executing cued non-consequential actions and inactions toward stimuli has been shown to influence a variety of value-based decisions, including soda consumption (Eberly et al., 2022), alcohol consumption (Di Lemma & Field, 2017), smartphone app use (Johannes et al., 2021), evaluation of food items (Chen et al., 2019), and evaluation of faces (Driscoll et al., 2021; Fenske et al., 2005). Generally, not responding to attractive stimuli makes these stimuli less positive; responding to less appetitive stimuli can make

[3] When we write "(in)actions", we always refer to people's decisions whether to (not) respond to a stimulus.

them more positive (e.g., Chen et al., 2019; Veling et al., 2022). Taken together, mere (in)actions affect impressions people have of others.

Recent theorizing on how non-consequential (in)actions influence subjective values holds that these (in)actions are best viewed as decisions, and that people aim to align the subjective values of stimuli with the performed (in)action decisions to reduce prediction errors (Veling et al., 2022). Due to a *Pavlovian bias*, exposure to highly valued items likely prepares a tendency to act, while negatively valued items prepares a tendency not to act (Ereira et al., 2021; Guitart-Masip, Duzel, et al., 2014). Forced action and inaction decisions toward the items (i.e., by presenting a cue to act or not act) may conflict with these inherent (in)action tendencies, resulting in prediction errors (e.g., when a cue to not act is presented near a rewarding stimulus). This may prompt updating the value of the items to bring prepared reaction tendencies in line with the stimulus values (to aid task execution). Thus, even without consequences, it is possible to change subjective values through (in)action decisions.

Consequential (in)actions as Value Input

In real life, (in)actions in social interactions often come with consequences, which can be experienced as rewarding or punishing. Consider our example of greeting a stranger: A smile from a stranger after (not) greeting them may be perceived as rewarding, while a frown may be perceived as punishing. How do rewarding and punitive consequences combined with the decision to (in)act affect people's impressions of others? For instance, are effects of (in)actions reduced once rewards and punishments are involved? Answering this question requires understanding the interplay between effects of action and inaction decisions on the one hand, and reward and punishment effects on the other hand. Ample evidence suggests that the mere co-occurrence of a positive/negative stimulus with another stimulus (e.g., a face) changes subjective values (De Houwer, 2007; Hofmann et al., 2010; Moran et al., 2023). Thus, co-occurrence of reward and punishment related information with faces changes people's impressions of those faces.

In addition, several studies have examined the influence of consequential *actions* on subjective values (Hackel et al., 2015; Hackel, Mende-Siedlecki, et al., 2022; Lott & Lott, 1974). For example, research suggests that action–reward learning about individuals influences attitudes, i.e., participants adopted a more positive attitude toward targets associated with greater reward (Hackel et al., 2019). However, the impact of consequential *inactions* on subjective values has remained unexplored. Research from the related domain of approach–avoidance research shows that adding consequences to approach–avoidance training strengthens the effect (Van Dessel et al., 2018). This suggests that adding consequences to (in)actions may have a similar strengthening effect.

Approach–Avoidance with or without Action

An important related research domain is work on approach–avoidance responses.[4] Approach responses are responses instrumental in obtaining rewards as outcomes, and avoidance responses are executed to avoid punishment as outcomes (Cacioppo & Berntson, 1994; Elliot, 2008; Krieglmeyer et al., 2013; Miller, 1959). To date, the influence of obtaining rewards (approach) and avoiding punishments (avoidance) on subjective values of stimuli, including faces, has been investigated in combination with approach and avoidance actions: People evaluate stimuli (or people) they have previously actively approached more positively than stimuli they have actively avoided (Kawakami et al., 2007, 2008; Phills et al., 2011; Slepian et al., 2012; Wiers et al., 2011; Woud et al., 2013; cf. Cacioppo et al., 1993).

Interestingly, research shows that whether an action is classified as approach or avoidance depends on the outcomes of that action and not the direction of the action itself (Krieglmeyer et al., 2010). Specifically, pulling a joystick can signify avoidance (i.e., retracting one's hand from the object) or approach (i.e., drawing the object closer to oneself). On the other hand, pushing a joystick can signify avoidance (i.e., pushing the object away) or approach (i.e., stretching one's hand toward the object; Markman & Brendl, 2005; Seibt et al., 2008). This raises the question of whether similar approach–avoidance effects on subjective value occur in the context of inactions. After all, one can also create situations in which doing nothing (inaction) leads to obtaining rewards or doing nothing leads to avoiding punishment.

The Action–Valence Asymmetry in Learning

In the context of inaction decisions, researchers have investigated consequential (in)actions within the domain of basic learning processes (e.g., Guitart-Masip et al., 2012). These studies focus on the learning process itself, and not so much on the impact of consequential (in)action decisions on subsequent evaluations of these stimuli. Evidence has accumulated that learning (in)actions can be amplified when different types of consequences (reward versus punishment) align with action versus inaction decisions (Cavanagh et al., 2013; Chowdhury et al., 2013; Guitart-Masip, Duzel, et al., 2014; Guitart-Masip, Economides, et al., 2014; Guitart-Masip et al., 2012; Mkrtchian et al., 2017; Moutoussis et al., 2018; Richter et al., 2014; Swart et al., 2017). Specifically, this work has demonstrated a basic action–valence asymmetry in learning, such that people more easily learn to act (e.g., respond to a stimulus) toward stimuli that elicit rewards, whereas they have relative difficulty to learn to act toward stimuli that lead to the avoidance of punishments. Conversely, people learn inactions (e.g., withhold a response to a stimulus) toward stimuli better when these lead to the avoidance of punishments, whereas they have relative difficulty to learn inactions toward stimuli that elicit rewards. The action–valence asymmetry is explained by an interaction between instrumental learning and a hard-wired Pavlovian bias (Guitart-

[4] Incorporating the approach–avoidance literature is something that predominantly emerged during the review process, which is why questions raised by approach–avoidance work are not reflected in the preregistrations and confirmatory tests.

Masip, Duzel, et al., 2014). According to the Pavlovian bias, exposure to rewards prepares a tendency to act, while exposure to punishments prepares a tendency to not act. In doing so, rewards hinder inaction learning while punishments hinder action learning, resulting in the action–valence asymmetry in learning. Here, we investigate whether and how this action–valence asymmetry in learning translates into evaluative impressions of faces. Moreover, we explore whether and how this may depend on whether the faces are from an ingroup or outgroup member.

The Present Research

In four experiments, we investigate how learning to act or not act toward faces to obtain rewards or avoid punishments shapes the subjective values of faces. All experiments consist of two phases. In the first phase, participants learn to act or not act toward faces to obtain rewards and avoid punishments. Here, we hypothesized to replicate previous work (Guitart-Masip et al., 2012), such that people would learn to act better when a face would elicit a reward and to not act when acting would result in punishment compared to respectively acting to avoid punishments or not acting to obtain rewards.

In the second phase, i.e., after learning, we measured the subjective value of the faces. We hypothesized a main effect of both action and valence on impressions, such that participants would rate faces as most positive when acting to these faces was learned to elicit rewards and most negative when people learned that acting would lead to punishment. Note that this means people may be more negative about faces for which they received less punishment. This is because, according to the action–valence asymmetry in learning, people should learn more easily to *not* take actions to avoid punishment (leading to less punishment) than to take actions to avoid punishment (leading to more punishment).

Subjective value will be measured by (1) preferences for one face over the other in value-based decisions and (2) explicit evaluations of each face. We use two measures for subjective value to increase reliability. Our four experiments are built up as follows: First, in Experiment 2.1, we replicate the Reinforcement Learning Go/No-Go Task (RL task) to test whether the findings from the original lab experiment (Guitart-Masip et al., 2012) generalize to an online environment. This task allows us to quantify action–valence asymmetries with fractals as stimuli. Then, in Experiments 2.2–2.4, we adjust the paradigm to a social context by using different White–Dutch faces (Experiment 2.2), Moroccan–Dutch faces (Experiment 2.3), and both groups of faces (Experiment 2.4) as stimuli. This set-up allows us to investigate our main research goal: The effects of action–valence asymmetries in learning on subjective values of faces. To investigate the consequences of action–valence asymmetries, it is essential that these asymmetries in learning are present. Therefore, while this is not our primary research aim, we also investigate whether action–valence asymmetries in learning generalize to faces. Although we expect to replicate action–valence asymmetries for faces, this is not evident. Learning about meaningless fractals is one thing; learning about faces, which are meaningful and therefore more complex, may be different. We use faces from different social groups to explore whether both the action–valence asymmetry in learning itself and its consequences on the subjective value of faces (1) generalize to different social groups and (2) are sensitive to group membership. We chose

Moroccan–Dutch faces as the outgroup because the Moroccan–Dutch community is one of the most negatively stereotyped social groups in the Netherlands (Verkuyten & Zarembe, 2005).

Experiment 2.1 is a direct replication of previous work on action–valence asymmetries with meaningless fractals as stimuli to verify whether our materials and procedures are reliable. Experiments 2.2–2.4 test our main research question, i.e., they address how (in)actions and its associated consequences shape subjective value of faces. This research has been reviewed independently by the Ethics Committee Social Sciences (ECSS) of the Radboud University and there is no formal objection to this study (ECSW-2021-037).

Open Practices

We report how we determined our sample sizes, all participant exclusions (if any), all manipulations, and all measures in the experiments. All data, all analysis code, all experiment scripts, and stimulus materials of Experiment 2.1 are available via https://osf.io/52k73/. Stimulus materials of Experiments 2.2–2.4 are available upon request via www.rafd.nl. Data were analyzed using R, version 4.1.2 (R Core Team, 2023).

We used the following packages programmed in R: lme4 (version 1.1.27.1; Bates et al., 2015), tidyverse (version 1.3.1; Wickham et al., 2019), Hmisc (version 4.6.0; Harrell, 2023), emmeans (version 1.7.2; Lenth, 2023), data.table (version 1.14.2; Dowle & Srinivasan, 2023), pbkrtest (version 0.5.1; Halekoh & Højsgaard, 2014), afex (version 1.0.1; Singmann et al., 2023), Rmisc (version 1.5; Hope, 2022), car (version 3.0.12; Fox & Weisberg, 2019), ggpubr (version 0.4.0; (Kassambara, 2023), patchwork (version 1.1.2; Pedersen, 2024), parallel (R Core Team, 2023) and psych (version 2.1.9; Revelle, 2023). All experiments were preregistered at the Open Science Framework (OSF). Preregistrations are available via https://osf.io/tvuq5 (Experiment 2.1), https://osf.io/587fp (Experiment 2.2), https://osf.io/c6z2j (Experiment 2.3), and https://osf.io/t7bk2 (Experiment 2.4).

Experiment 2.1

In Experiment 2.1, we replicated the RL task (Guitart-Masip et al., 2012) investigating the action–valence asymmetry in learning. We tested our confirmatory hypothesis that participants better learn to perform actions toward stimuli to obtain rewards (Go-To-Win) than to avoid punishments (Go-To-Avoid-Losing). Furthermore, we predicted participants to better learn to perform inactions toward stimuli to avoid punishments (No-Go-To-Avoid-Losing) than to obtain rewards (No-Go-To-Win).

Method

Sample Size Justification

The sample size was based on previous work (*N* = 47, Guitart-Masip et al., 2012). Note that we fully acknowledge the limitations and problems with basing a sample size on previous work (Anderson et al., 2017). However, in this specific case, we deemed it justified because we planned to perform formal power analyses based on this initial data set for Experiments 2.2–2.4. Moreover, we decided to be conservative, and intended to sample data of 60 participants. Participants were rewarded 5 euro or 0.5 credit point for participating and could earn more based on their performance (up to €3).

Participants

After recruiting 60 participants, we excluded four participants according to our preregistered exclusion criterion (performing the same action choice more than or equal to 90% of the time in at least one of the four blocks) and one participant because she did not finish the experiment. We resampled the number of excluded participants to again reach a sample size of 60 (M_{age} = 24.57 years, SD_{age} = 8.01 years, 45 women, 15 men). In all experiments, we recruited Radboud University students.

Materials and Procedure

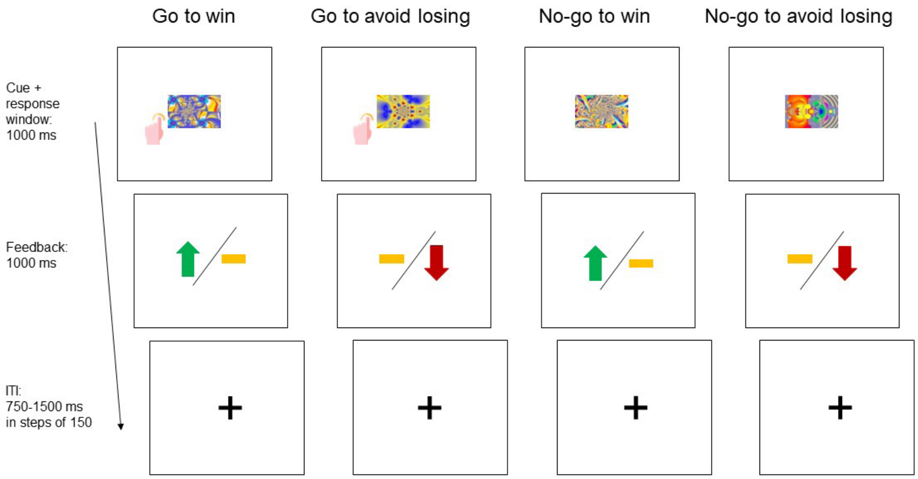
Reinforcement Learning Go/No-Go Task.

After providing consent, participants were asked to fill in their demographics (i.e., age, gender). The procedure used was adapted from Guitart-Masip and colleagues (2012). Participants were instructed that they would be shown different fractals and each fractal has one correct response (go; press spacebar, or no-go; not press). Moreover, each response could lead to a reward, a neutral outcome, or a punishment. Rewards led to a gain of 1 point and punishments led to a loss of 1 point. The feedback they received is probabilistic: If participants performed the right action, they received a reward (for Go-To-Win and No-Go-To-Win) or no punishment (for Go-To-Avoid-Losing and No-Go-To-Avoid-Losing) in 80% of the trials. Feedback is probabilistic to better mimic real-life conditions, where outcomes are often uncertain. In total, there were four different fractals: 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). Finally, for each fractal they had to learn trial and error based on the feedback what the best response is. Thus, the optimal response (go/no-go) for each fractal was not instructed and should be learned from the feedback. Participants also learned that after completion of the task, the points would be converted to a monetary bonus ranging from 0 to 3 euro. 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} = 36.05, SD_{points} = 22.24; M_{bonus} = 1.7, SD_{bonus} = 0.85).

Each trial started with the presentation of one of four fractals (randomly assigned per participant from a list of 12 fractals) for 1000 ms (image size: 200 × 125 pixels, 300 dpi). The visual angle of all stimuli was not controlled. During the presentation, participants either had to press spacebar (go) or withhold from pressing (no-go). The fractals were identical to what was employed in Guitart-Masip and colleagues (2012). After participants chose the action, they received feedback for 1000 ms. 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 750 ms to 1500 ms in steps of 150 ms (see Figure 2.1 for an overview).

In total, the task included four fractals, with 60 trials per fractal, resulting in a total of 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-second break. The trials within the blocks were randomized. Before starting the task, participants took part in 10 practice trials per fractal to get familiarized with the speed requirements (using different fractals than the ones used in the actual task). The task lasted approximately 25 minutes and after participating participants were paid conform their performance.

Figure 2.1



Note. Each trial started with the presentation of a 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

Reinforcement Learning Go/No-Go Task.

The probability of an optimal action choice (go or no-go) per RL condition served as dependent variable. This enabled us to investigate whether the learning of optimal action choice differs between RL conditions. Higher probabilities reflect better learning, whereas lower probabilities reflect worse learning. To determine the difference in performing the optimal action choice between Go-To-Win and Go-To-Avoid-Losing as well as between No-Go-To-Win and No-Go-To-Avoid-Losing, we conducted a binomial generalized linear mixed model. The model included the within-participant factors action (go/no-go) and valence (win/avoid losing). All fixed effects were coded using sum-to-zero contrasts. Moreover, this model included a random intercept of participant as well as random slopes for action and valence to keep a maximal random-effects structure (Barr et al., 2013).

Results

Confirmatory Analyses

Reinforcement Learning Go/No-Go Task.

The interaction effect of action and valence on optimal action choice was significant, *B* = -0.51, *SE* = 0.10, $\chi^2(1) = 20.05$, *p* <.001, 95% CI [-.72,-.30], OR = 0.60. As expected, follow-up pairwise comparisons revealed that the probability of performing the optimal action choice (go) was significantly higher in the Go-To-Win RL condition (*M* = 0.83*, SD* = 0.19) than in the Go-To-Avoid-Losing RL condition (*M* = 0.69*, SD* = 0.22), *B* = -1.24, *SE* = 0.28, *p* <.001, OR = 0.29. Moreover, the probability of performing the optimal action choice (no-go) was significantly higher in the No-Go-To-Avoid-Losing RL condition (*M* = 0.81*, SD* = 0.17) than in the No-Go-To-Win RL condition (*M* = 0.67*, SD* = 0.20), *B* = 0.81, *SE* = 0.20, *p* <.001, OR = 2.25. Taken together, we replicated the action–valence asymmetry in learning. See Figure 2.2 for the results.

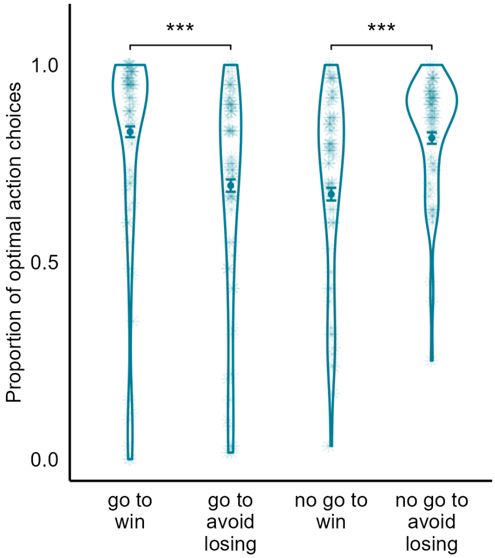
Exploratory follow-up pairwise comparisons revealed that the probability of performing the optimal action choice was significantly higher in the Go-To-Win RL condition (*M* = 0.83*, SD* = 0.19) than in the No-Go-To-Win RL condition (*M* = 0.67*, SD* = 0.20), *B* = 1.3, *SE* = 0.32, *p* <.001, OR = 3.68. Moreover, the probability of performing the optimal action choice was significantly higher in the No-Go-To-Avoid-Losing RL condition (*M* = 0.81*, SD* = 0.17) than in the Go-To-Avoid-Losing RL condition (*M* = 0.69*, SD* = 0.22), *B* = -0.75, *SE* = 0.23, *p* = .001, OR = 0.47.

Discussion

In Experiment 2.1, in line with Guitart-Masip and colleagues (2012), we found evidence for the action–valence asymmetry in learning. This replication paves the way for the next step: Investigating whether the action–valence asymmetry influences subjective values of faces.

Figure 2.2

Mean Proportion of Optimal Responses



Note. A violin plot showing the mean proportion of optimal responses in each of the four RL conditions. Error bars reflect within-participants confidence intervals around those means.

Experiment 2.2

In Experiment 2.2, we replaced fractals with faces of White-Dutch males[5] and examined whether the action–valence asymmetry generalizes to social stimuli. We again expected an asymmetry in learning. Importantly—to test whether and how the action–valence asymmetry in learning influences subjective value of the faces—participants performed a Value-Based Decision Task (VBD task) after the RL task. The goal of the VBD task is to unravel participants' preference for a face (see Chen et al., 2019; Schonberg et al., 2014; Veling et al., 2017). We expected that the learning process affects preferences for faces. More specifically, we expected that the probability of choosing a go face (a face that is assigned to the Go-To-Win or Go-To-Avoid-Losing condition) is higher in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair than in the other three experimental choice pairs. This exact hypothesis was preregistered. Additionally, we asked participants for their evaluation of each face, to explore whether the learning process in the RL task influences evaluations.

[5] We are aware of the issues of single gender designs (Grady, 1981). However, we selected only male stimuli because—looking ahead to Experiment 2.4—in the Netherlands, prejudice effects are presumed to be stronger for Moroccan–Dutch males than for Moroccan–Dutch females since males are, for example, more strongly associated with crime (e.g., Bovenkerk & Fokkema, 2016), and we aimed to examine whether effects would be different for a negatively stereotyped group.

Method

Sample Size Justification

An a priori power analysis using simr (Green & MacLeod, 2016) indicated that 27 participants were sufficient to find the action–valence asymmetry (power = .82, alpha = .05), given the data of Experiment 2.1. In Experiment 2.2, we made two adjustments: (1) we added the VBD task and the explicit evaluations and (2) we changed the stimuli from fractals to faces. Therefore, it is difficult to estimate whether 27 participants are indeed sufficient. Consequently, we decided to be conservative and collected data from 60 participants. Additionally, a sensitivity power analysis using simr (Green & MacLeod, 2016) indicated that 60 participants allowed us to find the action–valence asymmetry effect as small as *B* =-0.3 (power = .80, alpha = .05), given the data of Experiment 2.1. Participants were rewarded 7.5 euro or 0.75 credit point for participating and could earn more based on their performance in the RL task (the same bonus structure as in Experiment 2.1; M_{points} = 37.87, SD_{points} = 20.26; M_{bonus} = 1.73, SD_{bonus} = 0.66).

Participants

After recruiting 60 participants, we excluded two participants according to our preregistered exclusion criterion (performing the same action choice more than or equal to 90% of the time in at least one of the four blocks of the RL task), one participant because she participated twice (we kept her first try) and one participant because she reported to be 17 (we preregistered a minimum age of 18 years). We resampled the number of excluded participants to again reach the sample size of 60 (M_{age} = 21.22 years, SD_{age} = 3.92 years, 51 women, 9 men, see Supplemental materials for information on participant ethnicity).[6]

Materials and Procedure

Experiment 2.2 consisted of three parts: the RL task, the VBD task and the explicit evaluation task.

Reinforcement Learning Go/No-Go Task.

The first task was similar to the task in Experiment 2.1 but with different stimuli (i.e., faces of White–Dutch males). The faces were taken from the Radboud Faces Database (RaFD, Langner et al., 2010) and matched on valence and attractiveness. We used the RaFD because it is of high quality and the faces are well matched. We used the neutral, frontal images of actors 07, 23, 30, 71 (image size: 205 × 308 pixels, 96 dpi). Faces were randomly assigned

[6] Since the vast majority of the participants in Experiments 2.2–2.4 were women, we checked whether the descriptives of all analyses were similar for men. We concluded that the patterns were similar. This gives us confidence that the effects are not driven by the gender of the participants, but seem solid independent of participants' gender.

to RL condition (Go-To-Win, Go-To-Avoid-Losing, No-Go-To-Win, No-Go-To-Avoid-Losing) per participant. Before starting the RL task, participants took part in 10 practice trials per RL condition with different faces than the ones used in the actual task.

Value-Based Decision Task.

After the RL task, participants were introduced to the VBD task. Participants were instructed that they would repeatedly be presented with two faces. On each trial, they needed to decide which of the two faces appeared most positive to them at this moment. In the task, participants repeatedly observed two faces for 1500 ms. During the presentation, they had to press either the left arrow key (indicating that they chose the left face) or the right arrow key (indicating that they chose the right face). Once participants made their choice, the chosen face was surrounded by a blue frame for 500 ms to confirm that a choice had been made (see Figure 2.3 for an overview). Blue was chosen because it was not yet present in the RL task and it is not associated with correct or incorrect (e.g., cf. green or red). If participants failed to make a choice within the given time, that choice pair would be presented again at the end of the task.

Participants were consecutively presented with six different choice pairs of two faces, representing all possible combinations of four faces. Four choice pairs were experimental and were included to test the hypothesis: 'Go-To-Win vs No-Go-To-Avoid-Losing', 'Go-To-Avoid-Losing vs No-Go-To-Avoid-Losing', 'Go-To-Win vs No-Go-To-Win', and 'Go-To-Avoid-Losing vs No-Go-To-Win'. For two choice pairs, we did not have confirmatory hypotheses: 'Go-To-Win vs Go-To-Avoid-Losing' and 'No-Go-To-Avoid-Losing' vs No-Go-To-Win'. These six different choice pairs were presented in eight blocks, resulting in a total of 48 trials. There was no break between the blocks. The trials within the blocks were randomized, and the order of the faces was counterbalanced.

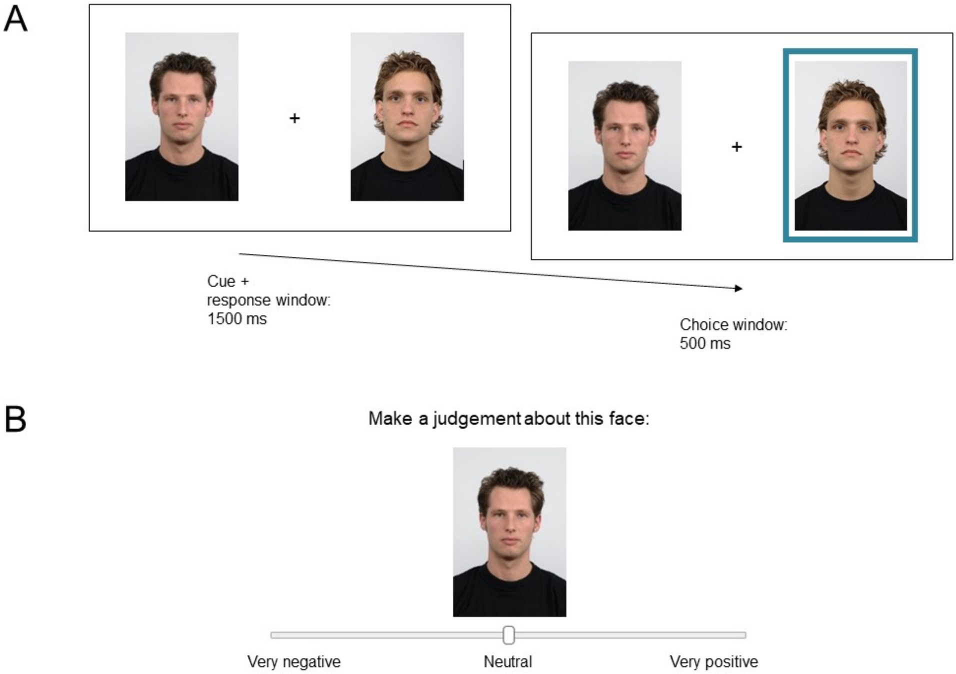
Explicit Evaluation Task.

After the VBD task, participants were asked to rate the four faces ('make a judgement about this face') on a scale from very negative (0) to very positive (200) (see Figure 2.3 for an overview). The order of the faces was randomized per participant. Ratings were made using a slider; its starting position was always at zero by default (neither positive nor negative).

Moreover, in addition to the demographics asked in Experiment 2.1, we asked participants with which ethnicity they identified at the end of the experiment best to get a more detailed overview of the participant characteristics. In total, all three tasks lasted approximately 40 minutes.

Figure 2.3

Overview of the Value-Based Decision Task and the Explicit Evaluation Task



Note. (A) The VBD task: Participants receive a series of binary choices between two faces. (B) The Explicit Evaluation task: Participants evaluate each face.

Confirmatory Analyses

Reinforcement Learning Go/No-Go Task. The analyses of the RL task were identical to Experiment 2.1.

Value-Based Decision Task. The probability of choosing a go face served as the dependent variable. In doing so, we were able to compare the differences in preference between choice pairs. A higher probability reflects a larger difference in preference, with a stronger preference for a go face (Go-To-Win / Go-To-Avoid-Losing condition) over a no-go face (No-Go-To-Win / No-Go-To-Avoid-Losing condition). To determine the difference in preference between the four experimental choice pairs, we conducted a binomial generalized linear mixed model. The model included the within-participant factor choice pair (Go-To-Win vs. No-Go-To-Avoid-Losing / Go-To-Avoid-Losing vs. No-Go-To-Avoid-Losing / Go-To-Win vs. No-Go-To-Win / Go-To-Avoid-Losing vs. No-Go-To-Win). Moreover, this model included a random intercept of participant as well as a random slope for choice pair.

Exploratory Analyses

Value-Based Decision Task.

We conducted these analyses to compare our results to the approach–avoidance literature. Action conditions reflect approach (Go-To-Win) and avoid (Go-To-Avoid-Losing). To explore whether approach–avoidance effects on preferences differ between action and inaction conditions, we conducted a binomial generalized linear mixed model. The probability of choosing a 'win' face served as the dependent variable. This model included the within-participant factor choice pair (Go-To-Win vs. Go-To-Avoid-Losing / No-Go-To-Win vs. No-Go-To-Avoid-Losing). Moreover, this model included a random intercept of participant as well as a random slope for choice pair.

Explicit Evaluation Task.

The evaluation per face served as the dependent variable. In doing so, we were able to compare the difference in how the faces are evaluated. 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 RL condition (Go-To-Win / No-Go-To-Win / Go-To-Avoid-Losing / No-Go-To-Avoid-Losing) and a random intercept of participant.

Results

Confirmatory Analyses

Reinforcement Learning Go/No-Go Task.

In line with Experiment 2.1, the interaction effect of action and valence on optimal action choice was significant, *B* =-0.52, *SE* = 0.08, $\chi^2(1) = 33.58$, *p* < .001, 95% CI [-.66,-.35], OR = 0.59. As expected, follow-up pairwise comparisons revealed that the probability of performing the optimal action choice (here: go) was significantly higher in the Go-To-Win RL condition (*M* = 0.91*, SD* = 0.14) than in the Go-To-Avoid-Losing RL condition (*M* = 0.76*, SD* = 0.15), *B* = -1.43, *SE* = 0.18, *p* <.001, OR = 0.24. Moreover, the probability of performing the optimal action choice (here: no-go) was significantly higher in the No-Go-To-Avoid-Losing RL condition (*M* = 0.78*, SD* = 0.11) than in the No-Go-To-Win RL condition (*M* = 0.63, *SD* = 0.22), *B* = 0.66, *SE* = 0.19, *p* <.001, OR = 1.93. See Figure 2.4 for the results.

Exploratory follow-up pairwise comparisons revealed that the probability of performing the optimal action choice was significantly higher in the Go-To-Win RL condition (*M* = 0.91*, SD* = 0.14) than in the No-Go-To-Win RL condition (*M* = 0.63, *SD* = 0.22), *B* = 1.94, *SE* = 0.26, *p* <.001, OR = 6.96. There was no significant difference in the probability of performing the optimal action choice between the No-Go-To-Avoid-Losing RL condition (*M* = 0.78*, SD* = 0.11) and the Go-To-Avoid-Losing RL condition (*M* = 0.76, *SD* = 0.15), *B* =-0.15, *SE* = 0.16, *p* = .334, OR = 0.86.

Value-Based Decision Task.

The main effect of choice pair on the probability of choices for go faces was significant, $\chi^2(1) = 16.94$, *p* = .002. As expected, follow-up pairwise comparisons revealed that the difference in preference was significantly higher in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair (*M* = 0.73*, SD* = 0.28) than in the other three choice pairs, respectively the 'Go-To-

Avoid-Losing vs No-Go-To-Avoid-Losing' (*M* = 0.67*, SD* = 0.25), *B* = -1.06, *SE* = 0.40, *p* = .038, OR = 0.35*,* the 'Go-To-Win vs No-Go-To-Win' (*M* = 0.57*, SD* = 0.27), *B* = 1.88, *SE* = 0.47, *p* < .001, OR = 6.55, and the 'Go-To-Avoid-Losing vs No-Go-To-Win' choice pair (*M* = 0.46*, SD* = 0.32), *B* = -2.69, *SE* = 0.64, *p* < .001, OR = 0.07. This confirms our hypothesis that the strongest difference in preference occurs in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair. See Figure 2.5 for the results.

Exploratory Analyses

Value-Based Decision Task.

The difference in preference for win outcomes is stronger in inaction contexts (No-Go-To-Win vs. No-Go-To-Avoid-Losing; *M* = 0.71*, SD* = 0.27) than in action contexts (Go-To-Win vs. Go-To-Avoid-Losing; *M* = 0.62*, SD* = 0.27)[7], *B* = 0.79, *SE* = 0.38, $\chi^2(1)$ = 4.42, *p* = .040, OR = 2.2, demonstrating that approach–avoidance effects on preferences are stronger in inaction than action conditions.

Explicit Evaluation Task.

The main effect of RL condition on explicit evaluation of the faces was significant, *F*(3,236) = 19.28, *p* <.001.[8] Follow-up comparisons revealed that the Go-To-Win face was the most positively evaluated face (*M* = 132.27*, SD* = 47.80) and that this face differed significantly from the Go-To-Avoid-Losing face (*M* = 103.93*, SD* = 40.00), *B* = -28.30, *SE* = 8.01, *p* = .003, β =-0.58[9]*,* and the No-Go-To-Avoid-Losing face (*M* = 75.48*, SD* = 53.92), *B* = 56.80, *SE* = 8.01, *p* <.001, β = 1.17. There was no significant difference between the Go-To-Win and No-Go-To-Win face (*M* = 121.87*, SD* = 43.11), *B* = 10.40, *SE* = 8.01, *p* =.565, β = 0.21. Moreover, the No-Go-To-Avoid-Losing face was evaluated the least positive, and in addition to differing from the Go-To-Win face the evaluation of this face differed significantly from the Go-To-Avoid-Losing face, *B* = 28.40, *SE* = 8.01, *p* = .003, β = 0.59, and No-Go-To-Win face, *B* = -46.40, *SE* = 8.01, *p* <.001, β =-0.95. There was no difference between the Go-To-Avoid-Losing and No-Go-To-Win face, *B* = -17.90, *SE* = 8.01, *p* =.117, β =-0.37. See Figure 2.6 for the results.

Discussion

Experiment 2.2 shows an asymmetry in learning and shows how this asymmetry translates into preferences for faces. Specifically, and as hypothesized, we found that the probability of choosing a go face is higher in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair

[7] We additionally tested whether these preferences differ from chance level (50 %). This is the case for all No-Go-To-Win vs. No-Go-To-Avoid-Losing and Go-To-Win vs. Go-To-Avoid-Losing choice pairs in Experiments 2.2–2.4.

[8] This model had a singular fit warning. This might be due no variation between participants or to negative ICC (intra class correlation). We double checked the results with an ANOVA. The conclusions were the same and therefore, we keep reporting this model.

[9] There are no generally accepted ways to compute standardized effect sizes for mixed effects models. Therefore, we included beta weights as a rough indicator of effect sizes for linear mixed effects models.

than in the other three experimental choice pairs. Moreover, exploratory analyses indicated that the asymmetries influence evaluations of faces. One thing is noteworthy: The No-Go-To-Avoid-Losing face was evaluated more negatively than the Go-To-Avoid-Losing face. This is striking as both RL conditions received punishment as feedback to the same degree. So mere co-occurrence effects cannot explain this difference. This points toward an additive effect of action and valence such that faces associated with No-Go-To-Avoid-Losing are particularly devalued. Since this finding is exploratory, we examined whether this replicates in the subsequent experiments.

In Experiment 2.3, we are interested to see whether the influence of action–valence asymmetries on preference for a face generalizes to Moroccan–Dutch targets.

Experiment 2.3

Experiment 2.3 is a replication of Experiment 2.2, but with Moroccan–Dutch male faces as stimuli.

Method

The procedure and measures were identical to Experiment 2.2. In addition to demographics (similar as in Experiment 2.2), we asked participants whether they have lived in the Netherlands all their life. This offers the possibility to get a more detailed overview of the participant characteristics. Participants were rewarded 7.5 euro or 0.75 credit point for participating and could earn more based on their performance in the RL task (the same bonus structure as in Experiment 2.2: M_{points} = 38.53, SD_{points} = 17.40; M_{bonus} = 1.75, SD_{bonus} = 0.68).[10] As stimuli, we used Moroccan–Dutch male faces. Faces were again taken from the RaFD (Langner et al., 2010) and were matched with each other and with the White–Dutch faces on valence and attractiveness. We used the neutral, frontal images of actors 29, 50, 53, 70 (image size: 205 × 308 pixels, 96 dpi).

Participants

In total, 60 participants participated (M_{age} = 20.9 years, SD_{age} = 2.84 years, 48 women, 10 men, 2 non-binary, see Supplemental materials for information on participant ethnicity).[11] Since this is a replication of Experiment 2.2, the power analyses described in Experiment 2.2 are also applicable here. According to our preregistered exclusion criteria, no participants had to be excluded.

[10] Due to a programming mistake in Experiments 2.1–2.3 some participants received less money than they were entitled to. Therefore, we recalculated the bonusses and we paid the concerned participants the right amount.

[11] We checked whether the descriptives of all analyses were similar when analyzing data of the participants that identified with Dutch, German or Caucasian. We concluded that the patterns were similar.

Results

Confirmatory Analyses

Reinforcement Learning Go/No-Go Task.

In line with Experiment 2.1 and 2.2, the interaction effect of action and valence on optimal action choice was significant, *B* = -0.47, *SE* = 0.08, $\chi^2(1) = 28.88$, *p* <.001, 95% CI [-.63,-.32], OR = 0.63.[12] As expected, follow-up pairwise comparisons revealed that the probability of performing the optimal action choice (here: go) was significantly higher in the Go-To-Win RL condition (*M* = 0.89*, SD* = 0.17) than in the Go-To-Avoid-Losing RL condition (*M* = 0.76*, SD* = 0.12), *B* = -1.43, *SE* = 0.21, *p* <.001, OR = 0.24. Moreover, the probability of performing the optimal action choice (here: no-go) was significantly higher in the No-Go-To-Avoid-Losing RL condition (*M* = 0.79*, SD* = 0.11) than in the No-Go-To-Win RL condition (*M* = 0.69*, SD* = 0.19), *B* = 0.48, *SE* = 0.17, *p* =.005, OR = 1.62. See Figure 2.4 for the results.

Exploratory follow-up pairwise comparisons revealed that the probability of performing the optimal action choice was significantly higher in the Go-To-Win RL condition (*M* = 0.89*, SD* = 0.17) than in the No-Go-To-Win RL condition (*M* = 0.69*, SD* = 0.19), *B* = 1.66, *SE* = 0.26, *p* <.001, OR = 5.26. There was no significant difference in the probability of performing the optimal action choice between the No-Go-To-Avoid-Losing RL condition (*M* = 0.79*, SD* = 0.11) and the Go-To-Avoid-Losing RL condition (*M* = 0.76*, SD* = 0.12), *B* =-0.24, *SE* = 0.13, *p* = .074, OR = 0.79.

Value-Based Decision Task.

The main effect of choice pair on the probability of choices for go faces was significant, $\chi^2(1) = 33.87$, *p* <.001. As expected, follow-up pairwise comparisons revealed that the difference in preference was significantly higher in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair (*M* = 0.83*, SD* = 0.25) than in the other three choice pairs, respectively the 'Go-To-Avoid-Losing vs No-Go-To-Avoid-Losing' (*M* = 0.67*, SD* = 0.25), *B* = -1.31, *SE* = 0.30, *p* < .001, OR = 0.27, the 'Go-To-Win vs No-Go-To-Win' (*M* = 0.63*, SD* = 0.27), *B* = 1.66, *SE* = 0.35, *p* < .001, OR = 5.26*,* and the 'Go-To-Avoid-Losing vs No-Go-To-Win' choice pair (*M* = 0.44*, SD* = 0.25), *B* = -2.74, *SE* = 0.41, *p* < .001, OR = 0.06. See Figure 2.5 for the results.

Exploratory Analyses

Value-Based Decision Task.

There is no difference in preference for win outcomes between inaction (No-Go-To-Win vs. No-Go-To-Avoid-Losing; *M* = 0.67*, SD* = 0.25) and action contexts (Go-To-Win vs. Go-To-Avoid-Losing; *M* = 0.66*, SD* = 0.25), *B* = 0.48, *SE* = 0.47, $\chi^2(1) = 1.14$, *p* = .287, OR = 1.61.

Explicit Evaluation Task.

The main effect of RL condition on explicit evaluation of the faces was significant, *F*(3,236) = 23.69, *p* <.001 (see Footnote 8). Follow-up comparisons revealed that the Go-To-Win

[12] For this analysis, we preregistered to determine p-values using the parametric bootstrap method. This method, however, is very time-consuming. Since it is as reliable as likelihood ratio tests according to Luke (2017), we chose to use likelihood ratio tests from now on to determine p-values.

face was the most positively evaluated face (*M* = 141.83*, SD* = 37.94) and that this face differed significantly from the Go-To-Avoid-Losing face (*M* = 109.53*, SD* = 35.18), *B* = -32.30, *SE* = 7.05, *p* <.001, β =-0.74*,* and the No-Go-To-Avoid-Losing face (*M* = 84.57*, SD* = 42.41), *B* = 57.30, *SE* = 7.05, *p* <.001, β = 1.31. There was no significant difference between the Go-To-Win and No-Go-To-Win face, (*M* = 124.50*, SD* = 43.56), *B* = 17.30, *SE* = 7.05, *p* =.070, β = 0.40. Moreover, the No-Go-To-Avoid-Losing face was evaluated the least positive, and in addition the differing from the Go-To-Win face the evaluation of this face differed significantly from the Go-To-Avoid-Losing face, *B* = 25.00, *SE* = 7.05, *p* =.003, β = 0.57, and No-Go-To-Win face, *B* = -39.90, *SE* = 7.05, *p* <.001, β =-0.91. There was no difference between the Go-To-Avoid-Losing and No-Go-To-Win face, *B* = -15.00, *SE* = 7.05, *p* =.150, β =-0.34. See Figure 2.6 for the results.

Discussion

Experiment 2.3 replicates the results of Experiment 2.2: The action–valence asymmetry and its influence on subjective value generalizes to Moroccan–Dutch faces. Again, exploratory analyses indicated that the No-Go-To-Avoid-Losing face was evaluated more negatively than the Go-To-Avoid-Losing face, suggesting an additive effect of action and valence. Now we obtained evidence that the action-valence asymmetry influences subjective value for both White–Dutch and Moroccan–Dutch faces similarly, we extended the research paradigm using both White–Dutch and Moroccan–Dutch faces in a within-participant design in a Dutch/German sample in Experiment 2.4. By displaying both social groups (instead of one), we aim to push the intergroup context more strongly (Young et al., 2009). If the experiment is sensitive to intergroup differences, we expect that effects should emerge more quickly in this within-participant design.

Experiment 2.4

In Experiment 2.4, we examined the influence of the action–valence asymmetry on subjective value for in- and outgroup members. We used both White–Dutch and Moroccan–Dutch male faces in the RL task. After the RL task, participants again carried out the VBD task and evaluated all faces. Since it is unknown whether the experiment is sensitive to intergroup differences, the intergroup comparison was exploratory.

Within each group condition (ingroup and outgroup), we did have predictions, which we preregistered for each group condition *separately*. The hypotheses for both the RL task and the VBD task are the same as described in Experiments 2.2–2.3. In addition, this time, we preregistered the hypotheses for the explicit evaluation task. We expected that the learning processes in the RL task affect the evaluation of face. More specifically, we expected that participants' evaluation of the Go-To-Win face is the most positive and significantly more positive than the No-Go-To-Avoid-Losing face. Moreover, we expected that the participants' evaluation of the Go-To-Avoid-Losing face is significantly more positive than the No-Go-To-Avoid-Losing face.

Method

Sample Size Justification

Two a priori power analyses using simr (Green & MacLeod, 2016) indicated that respectively 34 (given the data from Experiment 2.2, power = .80, alpha = .05) and 27 participants (given the data from Experiment 2.3, power = .82, alpha = .05) were sufficient to find the smallest hypothesized effect, which was the difference in participants' evaluation of a face between the 'Go-To-Avoid-Losing' and 'No-Go-To-Avoid-Losing' condition in the explicit evaluation task. In Experiment 2.4 we made one major adjustment: Because we included both the stimuli of Experiment 2.2 and 2.3, we doubled all tasks included. Therefore, it was difficult to estimate whether 27 to 34 participants were indeed sufficient. Consequently, again, we decided to be conservative and to collect data from 60 participants. Additionally, two sensitivity power analyses using simr (Green & MacLeod, 2016) indicated that 60 participants allowed us to find an effect on participants' evaluation of a face between the 'Go-To-Avoid-Losing' and 'No-Go-To-Avoid-Losing' condition in the explicit evaluation task as small as *B* = 10 (given the data from Experiment 2.2, power = .80, alpha = .05) and *B* = 9.5 (given the data from Experiment 2.3, power = .80, alpha = .05).

Participants

After recruiting 60 participants, we excluded two participants according to our preregistered exclusion criterion (performing the same action choice more than or equal to 90% of the time in at least one of the six blocks of the RL task) and one participant because she participated twice (we kept her first try). We resampled the number of excluded participants to again reach the sample size of 60 (M_{age} = 22.02 years, SD_{age} = 5.34 years, 47 women, 13 men, 48 Dutch, 12 German). For this experiment, we explicitly recruited participants that were grown up in The Netherlands, Germany, or Belgium and that 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 as familiar for Dutch participants as it is for German and Belgian participants. Thus, we anticipate that the White–Dutch faces are observed in the same way for both German and Belgian as well as Dutch participants, thereby reflecting an ingroup. Participants were rewarded 10 euro or 1 credit point for participating and could earn more based on their performance in the RL task (up to €4).

Materials and Procedure

Experiment 2.4 consisted of three parts: the RL task, the VBD task and the explicit evaluation task.

Reinforcement Learning Go/No-Go Task.

We combined Experiments 2.2–2.3 in the first task. This resulted in eight different RL conditions: Go-To-Win Moroccan face, Go-To-Avoid-Losing Moroccan face, No-Go-To-Win

Moroccan face, No-Go-To-Avoid-Losing Moroccan face, Go-To-Win Dutch face, Go-To-Avoid-Losing Dutch face, No-Go-To-Win Dutch face, and No-Go-To-Avoid-Losing Dutch face. The set up and the stimuli were the same as in Experiments 2.2–2.3, only now with both Moroccan and Dutch faces as stimuli. Faces were randomly assigned to RL conditions within each group (ingroup or outgroup). Thus, there were eight RL conditions with 60 trials per RL condition, resulting in a total of 480 trials. After every 80 trials (10 trials per RL condition), participants had a 20-second break. The trials within the blocks were randomized.

Based on their learning performance, participants could (as in Experiments 2.1–2.3) earn a monetary bonus. The monetary bonus ranged from 0 to 4 euro. More specifically, and unbeknownst to participants, participants with scores of 0 or below 0 points gained 0 euro bonus, participants with scores ranging 1 and 40 gained 1 euro bonus, participants with scores ranging between 41 and 76 gained 2 euro bonus, participants with scores ranging between 77 and 112 gained 3 euro bonus and participants with scores higher than 112 gained 4 euro bonus (M_{points} = 50.55, SD_{points} = 33.12; M_{bonus} = 1.78, SD_{bonus} = 0.94).

Value-Based Decision Task.
After the RL task, participants participated in the VBD task. This task was the same as Experiments 2.2–2.3, but now with more choice pairs. Participants were consecutively presented with 28 different choice pairs of two faces, representing all possible combinations of eight faces. Eight choice pairs were experimental: 'Go-To-Win vs No-Go-To-Avoid-Losing', 'Go-To-Avoid-Losing vs No-Go-To-Avoid-Losing', 'Go-To-Win vs No-Go-To-Win' and 'Go-To-Avoid-Losing vs No-Go-To-Win' for both White–Dutch and Moroccan–Dutch faces. For the other choice pairs, we did not have confirmatory hypotheses. These 28 different choice pairs were presented in eight blocks, resulting in a total of 224 trials. After 112 trials, participants had a 20-second break. The trials within the blocks were randomized, and the order of the faces was counterbalanced.

Explicit Evaluation Task.
After the VBD task, participants were again asked to rate the eight faces ('make a judgement about this face') on a scale from very negative (0) to very positive (200). The order of the faces was random per participant. In total, all three tasks lasted approximately 55 minutes.

Results

Confirmatory Analyses

Reinforcement Learning Go/No-Go Task — Moroccan–Dutch Faces.

The interaction effect of action and valence on optimal action choice was significant, *B* = -0.54, *SE* = 0.09, $\chi^2(1)$ = 25.95, *p* <.001, 95% CI [-.73,-.35], OR = 0.58. As expected, follow-up pairwise comparisons revealed that the probability of performing the optimal action choice (here: go) was significantly higher in the Go-To-Win RL condition (*M* = 0.80, *SD* = 0.22) than in the Go-To-Avoid-Losing RL condition (*M* = 0.69, *SD* = 0.16), *B* = -1.03, *SE* = 0.21, *p* <.001, OR = 0.36. Moreover, the probability of performing the optimal action choice (here: no-go) was significantly higher in the No-Go-To-Avoid-Losing RL condition (*M* = 0.72, *SD* = 0.17) than in the No-Go-To-Win RL condition (*M* = 0.50, *SD* = 0.26), *B* = 1.12, *SE* = 0.23, *p* <.001, OR = 3.06. See Figure 2.4 for the results.

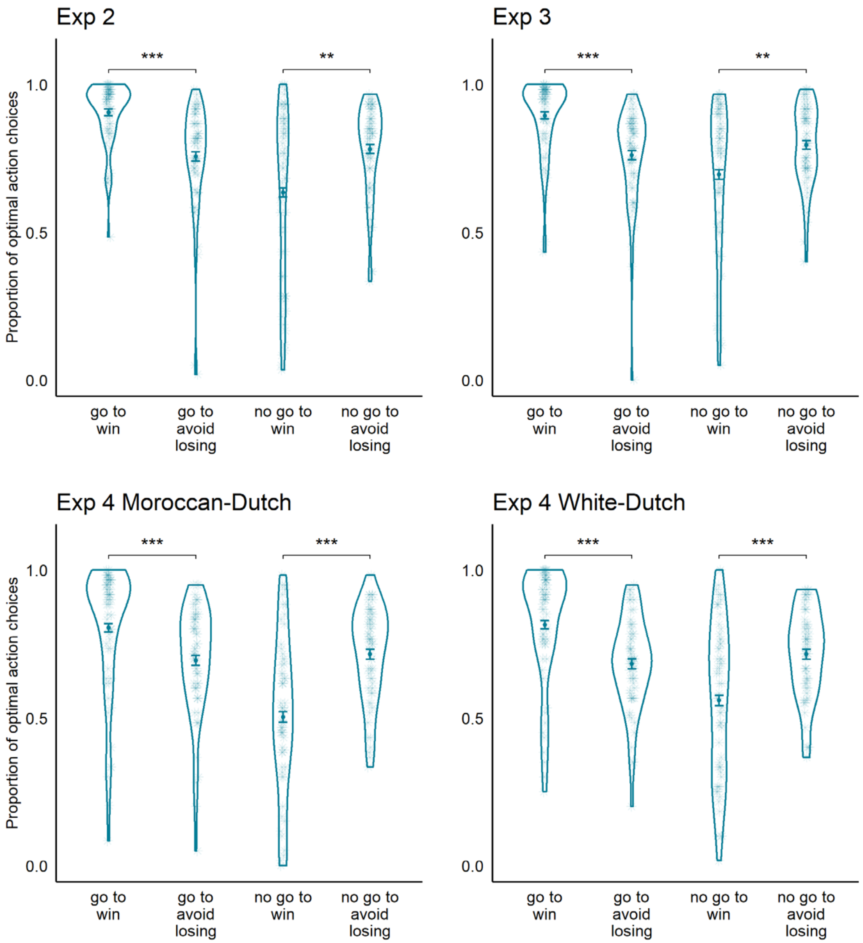
Exploratory follow-up pairwise comparisons revealed that the probability of performing the optimal action choice was significantly higher in the Go-To-Win RL condition (*M* = 0.80, *SD* = 0.22) than in the No-Go-To-Win RL condition (*M* = 0.50*, SD* = 0.26), *B* = 2.05, *SE* = 0.35, *p* <.001, OR = 7.73. There was no significant difference in the probability of performing the optimal action choice between the No-Go-To-Avoid-Losing RL condition (*M* = 0.72, *SD* = 0.17) and the Go-To-Avoid-Losing RL condition (*M* = 0.69, *SD* = 0.16), *B* =-0.10, *SE* = 0.16, *p* = .535, OR = 0.91.

Reinforcement Learning Go/No-Go Task — White–Dutch Faces.
The interaction effect of action and valence on optimal action choice was significant, *B* = -0.49, *SE* = 0.08, χ^2 (1) = 28.99, *p* <.001, 95% CI [-.65,-.33], OR = 0.61. As expected, follow-up pairwise comparisons revealed that the probability of performing the optimal action choice (here: go) was significantly higher in the Go-To-Win RL condition (*M* = 0.81, *SD* = 0.20) than in the Go-To-Avoid-Losing RL condition (*M* = 0.68, *SD* = 0.13), *B* = -1.24, *SE* = 0.22, *p* <.001, OR = 0.29*.* Moreover, the probability of performing the optimal action choice (here: no-go) was significantly higher in the No-Go-To-Avoid-Losing RL condition (*M* = 0.72, *SD* = 0.14) than in the No-Go-To-Win RL condition (*M* = 0.56, *SD* = 0.22), *B* = 0.71, *SE* = 0.18, *p* <.001, OR = 2.03. See Figure 2.4 for the results.

Exploratory follow-up pairwise comparisons revealed that the probability of performing the optimal action choice was significantly higher in the Go-To-Win RL condition (*M* = 0.81, *SD* = 0.20) than in the No-Go-To-Win RL condition (*M* = 0.56, *SD* = 0.22), *B* = 1.78, *SE* = 0.29, *p* <.001, OR = 5.90. There was no significant difference in the probability of performing the optimal action choice between the No-Go-To-Avoid-Losing RL condition (*M* = 0.72, *SD* = 0.14) and the Go-To-Avoid-Losing RL condition (*M* = 0.68, *SD* = 0.13), *B* =-0.17, *SE* = 0.13, *p* = .166, OR = 0.84.

Figure 2.4

Results Reinforcement Learning Go/No-Go Task



Note. Mean proportion of optimal responses in each of the four RL conditions. Error bars reflect within-participants confidence intervals around those means.

Value-Based Decision Task — Moroccan–Dutch Faces.

The main effect of choice pair on the probability of choices for go faces was significant, χ^2 (3) = 26.20, *p* <.001. As expected, follow-up pairwise comparisons revealed that the difference in preference was significantly higher in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair (*M* = 0.79*, SD* = 0.22) than the 'Go-To-Avoid-Losing vs No-Go-To-Avoid-Losing' (*M* = 0.64*, SD* = 0.28), *B* = -0.87, *SE* = 0.32, *p* =.033, OR = 0.42*,* and the 'Go-To-Avoid-Losing vs No-Go-To-Win' choice pair (*M* = 0.42*, SD* = 0.28), *B* = -2.37, *SE* = 0.42, *p* <.001, OR = 0.09*.* Contrary to our hypothesis, we did not find a difference in the 'Go-To-Win vs

No-Go-To-Win' choice pair (*M* = 0.65*, SD* = 0.29), *B* = 0.86, *SE* = 0.34, *p* =.054, OR = 2.36. See Figure 2.5 for the results.

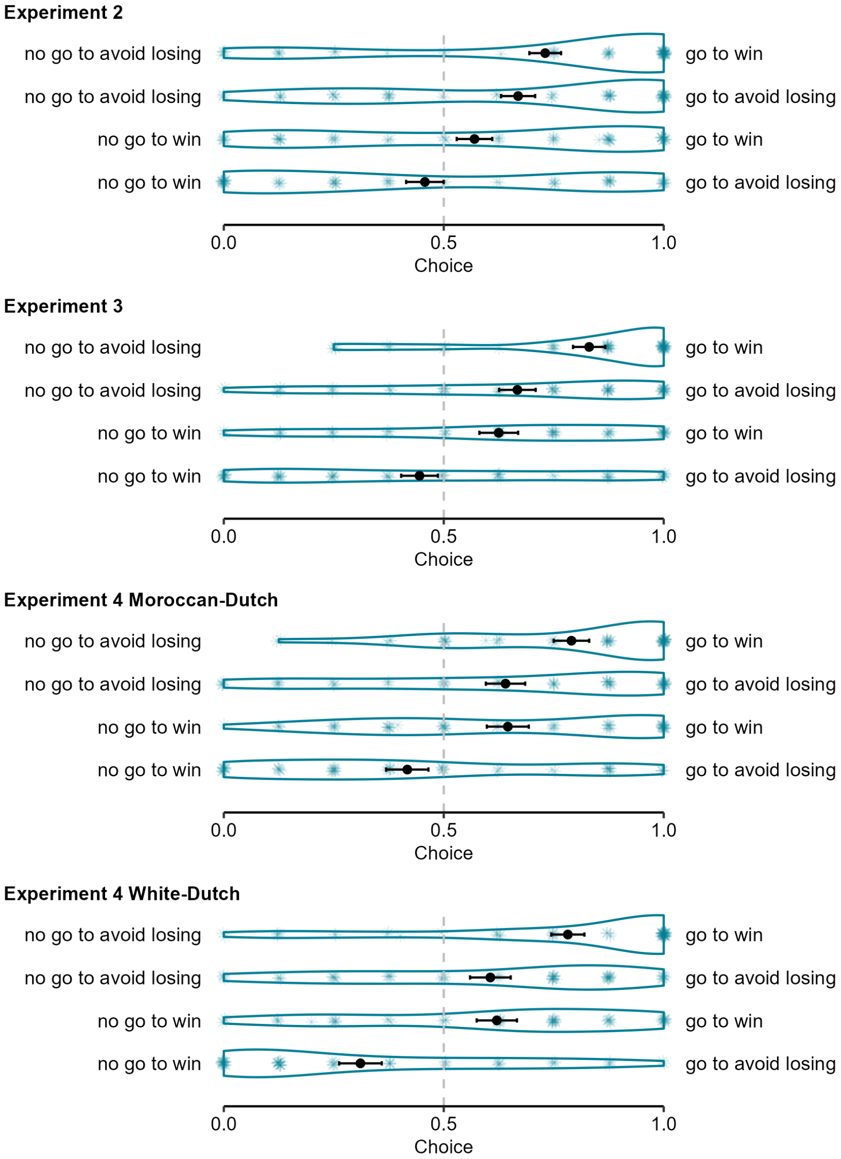
Exploratory analyses indicated that there is no difference in preference for win outcomes between inaction (No-Go-To-Win vs. No-Go-To-Avoid-Losing; *M* = 0.68*, SD* = 0.24) and action contexts (Go-To-Win vs. Go-To-Avoid-Losing; *M* = 0.69*, SD* = 0.24), *B* = -0.41, *SE* = 0.62, χ^2(1) = 0.46, *p* = .496, OR = 0.66.

Value-Based Decision Task — White–Dutch Faces.
The main effect of choice pair on the probability of choices for go faces was significant, χ^2 (3) = 34.84, *p* <.001. As expected, follow-up pairwise comparisons revealed that the difference in preference was significantly higher in the 'Go-To-Win vs No-Go-To-Avoid-Losing' choice pair (*M* = 0.78*, SD* = 0.25) than in the other three choice pairs, respectively the 'Go-To-Avoid-Losing vs No-Go-To-Avoid-Losing' (*M* = 0.61*, SD* = 0.25), *B* = -1.94, *SE* = 0.40, *p* <.001, OR = 0.14*,* the 'Go-To-Win vs No-Go-To-Win' (*M* = 0.62*, SD* = 0.23), *B* = 1.88, *SE* = 0.41, *p* <.001, OR = 6.55, and the 'Go-To-Avoid-Losing vs No-Go-To-Win' choice pair (*M* = 0.31*, SD* = 0.32), *B* = -3.80, *SE* = 0.60, *p* <.001, OR = 0.02*.* See Figure 2.5 for the results.

Exploratory analyses indicated that there is no difference in preference for win outcomes between inaction (No-Go-To-Win vs. No-Go-To-Avoid-Losing; *M* = 0.71, *SD* = 0.19) and action contexts (Go-To-Win vs. Go-To-Avoid-Losing; *M* = 0.71, *SD* = 0.19), *B* = 0.03, *SE* = 0.49, χ2(1) = 0.003, *p* = .957, OR = 1.03.

Figure 2.5

Results Value-Based Decision Task



Note. Mean proportion of chosen ‘go faces’ in each of the four experimental choice pairs. Error bars reflect within-participants confidence intervals around those means.

Explicit Evaluation Task — Moroccan–Dutch Faces.

As expected, the main effect of RL condition on the explicit evaluation of the faces was significant, *F*(3,236) = 20.01 , *p* <.001 (see Footnote 8). The descriptive order of the evaluations was also as hypothesized: the 'Go-To-Win' face was the most positive (*M* = 139.37*, SD* = 44.54), followed by the 'No-Go-To-Win' face (*M* =123.55*, SD* = 43.44), followed by the 'Go-To-Avoid-Losing' face (*M* = 110.18*, SD* = 43.91) and lastly followed by the 'No-Go-To-Avoid-Losing' face (*M* = 77.52*, SD* = 57.86). Moreover, we confirm the hypothesis that the 'Go-To-Win' face is evaluated more positively than the 'No-Go-To-Avoid-Losing' face, *B* = 61.90, *SE* = 8.31, *p* <.001, β = 1.22. Lastly, we confirm the hypothesis that the 'Go-To-Avoid-Losing' face is evaluated more positively than the 'No-Go-To-Avoid-Losing' face, *B* = 32.70, *SE* = 8.31, *p* <.001, β = 0.65.

In an exploratory fashion, we additionally tested all the other follow-up comparisons. This revealed that the 'Go-To-Win' face was evaluated more positively than the 'Go-To-Avoid-Losing' face, *B* = -29.20, *SE* = 8.31, *p* =.003, β =-0.58. There was no significant difference between the 'Go-To-Win' and 'No-Go-To-Win' face, *B* = 15.80, *SE* = 8.31, *p* =.231, β = 0.31. Moreover, the 'No-Go-To-Avoid-Losing' face was evaluated less positively than the 'No-Go-To-Win' face, *B* = -46.00, *SE* = 8.31, *p* <.001, β =-0.91. There was no difference between the 'Go-To-Avoid-Losing' and 'No-Go-To-Win' face, *B* = -13.40, *SE* = 8.31, *p* =.377, β =-0.26. See Figure 2.6 for the results.

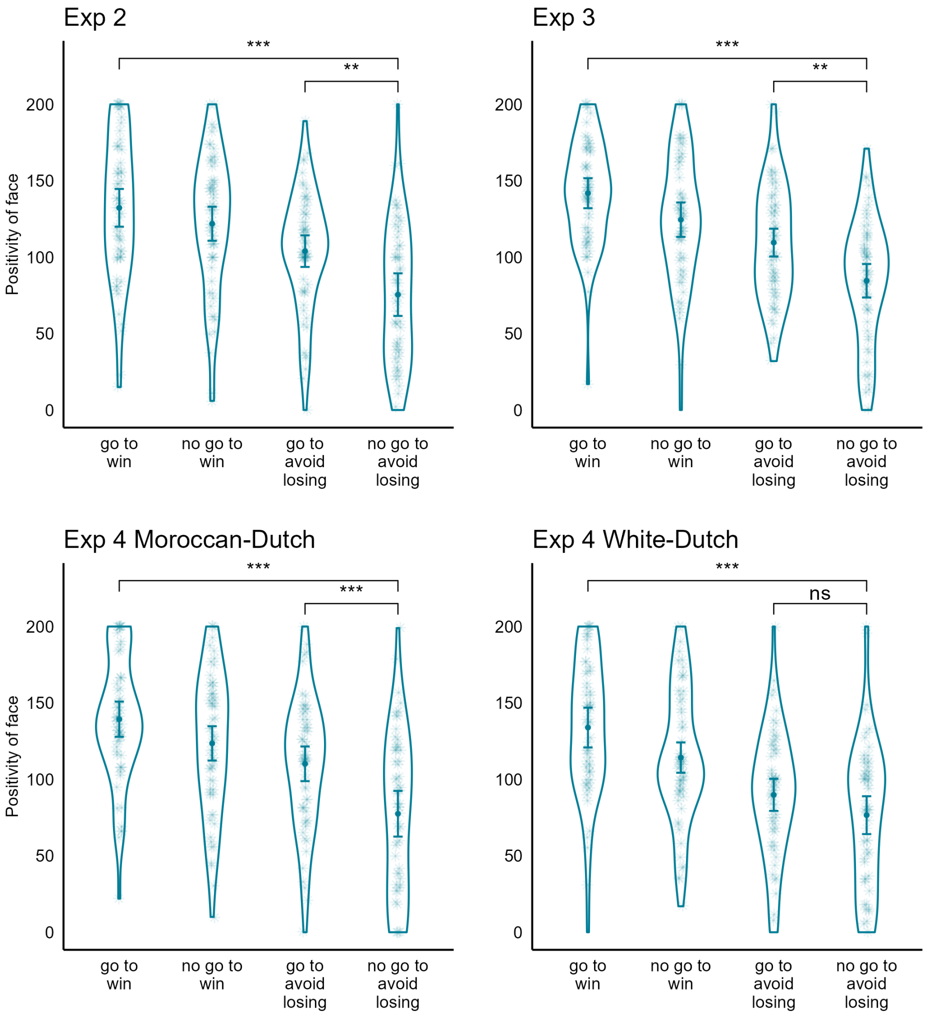
Explicit Evaluation Task — White–Dutch Faces.

As expected, the main effect of RL condition on the explicit evaluation of the faces was significant, *F*(3,236) = 19.99 , *p* <.001 (see Footnote 8). The descriptive order of the evaluations was also as hypothesized: the 'Go-To-Win' face was the most positive (*M* =133.90*, SD* = 50.34), followed by the 'No-Go-To-Win' face (*M* =114.25*, SD* = 38.38), followed by the 'Go-To-Avoid-Losing' face (*M* = 89.88*, SD* = 40.62), and lastly followed by the 'No-Go-To-Avoid-Losing' face (*M* = 76.60*, SD* = 47.97). Moreover, we confirm the hypothesis that the 'Go-To-Win' face is evaluated more positively than the 'No-Go-To-Avoid-Losing' face, *B* = 57.30, *SE* = 8.06, *p* <.001, β = 1.17. In contrast with our hypothesis and different than Experiments 2.2–2.3 and Moroccan–Dutch faces in Experiment 2.4, we do not find evidence that the 'Go-To-Avoid-Losing' face is evaluated more positively than the 'No-Go-To-Avoid-Losing' face, *B* =13.30, *SE* = 8.06, *p* = .355, β = 0.27.

Exploratory, we additionally tested all the other follow-up comparisons. This revealed that the 'Go-To-Win' face was evaluated more positively than the 'Go-To-Avoid-Losing' face, *B* = -44.00, *SE* = 8.06, *p* <.001, β =-0.90. There was no significant difference between the 'Go-To-Win' and 'No-Go-To-Win' face, *B* = 19.60, *SE* = 8.06, *p* = .074, β = 0.40. Moreover, the 'No-Go-To-Avoid-Losing' face was evaluated less positively than the 'No-Go-To-Win' face, *B* = -37.60, *SE* = 8.06, *p* <.001, β =-0.77. Lastly, the 'Go-To-Avoid-Losing' face was evaluated less positively than the 'No-Go-To-Win' face, *B* = -24.40, *SE* = 8.06, *p* =.015, β = -0.50. See Figure 2.6 for the results.

Figure 2.6

Results Explicit Evaluation Task



Note. Mean evaluation of the face in each of the four RL conditions. Error bars reflect within-participants confidence intervals around those means.

Additional Exploratory Analyses

To investigate whether there are intergroup differences in the RL task, the VBD task and explicit evaluation task we added the within-participant factor group condition (ingroup/outgroup). Moreover, the model of the RL task and VBD task included a random slope for group condition.

Reinforcement Learning Go/No-Go Task.

The three-way interaction of action, valence and group on optimal action choice was nonsignificant, *B* = 0.02, *SE* = 0.03, $\chi^2(1) = 0.46$, *p* =.50, OR = 1.02. In our data, we do not find evidence that there is an intergroup difference for learning the optimal action choice.

Value-Based Decision Task.

The interaction between group and choice pair on choices for go faces was nonsignificant, $\chi^2(41) = 6.80$, *p* =.079. In our data, we do not find evidence that there is an intergroup difference for preference for a face between the choice pairs.

Explicit Evaluation Task.

The interaction between group and RL condition on evaluation of the faces was nonsignificant, *F(*3,472) = 1.02, *p* =.382 (see Footnote 8). In our data, we do not find evidence that there is an intergroup difference for evaluations. We do, however, find a significant main effect for group condition, *B* = -9, *SE* = 4.09, *F*(1,472) = 4.83, *p* =.028, β =-0.18, with Moroccan faces (*M* = 112.65, *SD* = 69.80) being in general more positively evaluated than Dutch faces (*M* = 103.66, *SD* = 66.10).

Discussion

In Experiment 2.4, we replicate the results of Experiments 2.2–2.3, showing the influence of the action–valence asymmetry on subjective value for both White–Dutch and Moroccan–Dutch faces. These results again suggest an additive effect of (in)actions and consequences. Regarding the evaluations, we replicated the pattern that the Go-To-Win face is more positively evaluated than the No-Go-To-Avoid-Losing face. Moreover, in line with Experiment 2.3, the Go-To-Avoid-Losing face is more positively evaluated than the No-Go-To-Avoid-Losing face but only for outgroup faces. In contrast to our preregistered hypothesis, we do not find evidence for this effect for ingroup faces.

We also explored the action–valence contingencies, the preferences, and explicit evaluations for faces between the two group conditions and did not find any evidence for intergroup differences. Taken together, we conclude that action–valence asymmetries in learning and its influence on preference and explicit evaluations also occur in a within-participant design with two social groups, and that these effects are similar for ingroup and outgroup faces.

Exploratory Pooled Analysis

To further explore the idea that (in)actions and consequences lead to additive effects, and to provide our best estimate of these effects, we conducted an individual-participants meta-analysis on data from all experiments combined.

Explicit Evaluation Task

We re-analyzed the explicit evaluation data to examine whether there are two main effects present of action and valence. We fitted a linear mixed effects model. The model included the within-participant factors action (go/no-go) and valence (win/avoid losing) and the between-participant factor experiment. Moreover, this model included a random intercept of participant.

The main effect of action on the explicit evaluation of the faces was significant, *B* = 10.15, *SE* = 1.47, *F*(1,948) = 47.69, *p* <.001, β = 0.21. Action faces (*M* = 120.11, *SD* = 44.42) were more positively evaluated than inaction faces (*M* = 99.79, *SD* = 49.79). Moreover, the main effect of valence on explicit evaluations was significant, *B* =-18.78, *SE* = 1.47, *F*(1,948) = 163.16, *p* <.001, β =-0.39. Reward faces (*M* = 128.94, *SD* = 43.45) were more positively evaluated than avoid losing faces (*M* = 90.96, *SD* = 45.27). There was no significant interaction between action and valence on explicit evaluations, *B* = 2.58, *SE* = 1.47, *F*(1,948) = 3.07, *p* = .08, β = 0.05. That said, follow-up pairwise comparisons revealed that within the avoid losing condition go faces were significantly more positively evaluated than no-go faces, *B* = 25.5, *SE* = 4.16, *p* < .001, β = 0.53. Within the win condition go faces were also significantly more positively evaluated than no-go faces, *B* = 15.2, *SE* = 4.16, *p* < .001, β = 0.31.

Additionally, we also re-ran the one-factor linear mixed effects model that we used throughout this paper. The model included the within-participant factor RL condition, the between-participant factor experiment and a random intercept of participant.

The main effect of RL condition on the explicit evaluation of the faces was significant, *F*(3,948) = 71.31, *p* < .001. Follow-up comparisons revealed that all conditions differed significantly from one another. More specifically, the Go-To-Win face was the most positively evaluated face (*M* = 136.84*, SD* = 45.27) and this face differed significantly from the No-Go-To-Win face (*M* = 121.04*, SD* = 44.75), *B* = 15.2, *SE* = 4.16, *p* = .002, β = 0.31, the Go-To-Avoid-Losing face (*M* = 103.38, *SD* = 42.62), *B* = -32.4, *SE* = 4.16, *p* < .001, β =-0.67 , and the No-Go-To-Avoid-Losing face (*M* = 78.54, *SD* = 52.51), *B* = 57.9, *SE* = 4.16, *p* < .001, β = 1.2. Moreover, the No-Go-To-Avoid-Losing face was evaluated the least positive, and in addition to differing from the Go-To-Win face the evaluation of this face differed significantly from the Go-To-Avoid-Losing face, *B* = 25.5, *SE* = 4.16, *p* < .001, β = 0.53, and the No-Go-To-Win face, *B* = -42.7, *SE* = 4.16, *p* < .001, β =-0.89. Finally, there was a significant difference between the Go-To-Avoid-Losing and No-Go-To-Win face, *B* = -17.3, *SE* = 4.16, *p* < .001, β =-0.36.

Discussion

Our individual-participants meta-analysis shows that combining action and valence leads to additive effects on subjective values. That is, actions matter over and above consequences; this finding was most robust in the conditions in which participants avoided punishments (rather than attained rewards) through action or inaction.

General Discussion

The current research investigated how actions and inactions, and their associated consequences, influence subjective values of faces. In line with our preregistered hypotheses, we find consistent evidence for action–valence asymmetries during learning. After replicating the action–valence asymmetry in learning for fractals (Guitart-Masip et al., 2012), we found evidence that action–valence asymmetries in learning generalize to faces. Thus, we demonstrated that action–valence asymmetries extend to meaningful social stimuli. Importantly, and as preregistered, the action–valence asymmetry observed during learning transfers into effects on subjective values of faces. We showed that there is an interplay between the effects of actions and reward in shaping impressions of faces: Combining action with reward during learning led to the most positive impressions, while combining inaction with avoiding punishment during learning led to the least positive impressions.

We started off with the question whether (in)actions would still influence impressions of other people's faces when combined with affective consequences of reward and punishment. Thus, an important question raised by the current findings is whether the effects can be explained by the (frequency of) mere affective consequences (reward or punishment) with the faces. The most robust and convincing evidence against this explanation—and in favor of an explanation pointing to the role of (in)actions amplifying effects of affective consequences—can be found in the two avoiding punishment conditions (Go-To-Avoid-Losing and No-Go-To-Avoid-Losing). Here, we observed that (in)actions matter over and above punishment in forming impressions: Go-To-Avoid-Losing faces are evaluated more positively than No-Go-To-Avoid-Losing faces (except for White–Dutch faces in Experiment 2.4, although the pattern is in the expected direction) and Go-To-Avoid-Losing faces are chosen more in the Value-Based Decision Task ('Go-To-Avoid-Losing vs No-Go-To-Avoid-Losing' choice pair). In these two conditions participants learned the outcome contingencies approximately equally well in the Reinforcement Learning (RL) task (we do not find evidence for a difference in learning between these two conditions), meaning that they were exposed to similar face–punishment relations in both conditions. The frequency of punishments and faces cannot explain these results, but additive effects of (in)actions in learning can: (In)actions matter over and above the effect of punishment signals. In other words, there appears to be something special about punishing people for acting on faces, above and beyond the punishment signal itself. This amplified negativity of the action–punishment combination may be attributed to a hard-wired Pavlovian bias (Guitart-Masip, Duzel, et al., 2014). This bias during learning aids people to quickly learn to prevent actions to punishment-related stimuli (e.g., touching a flame). This bias may also be responsible for making these faces evaluated particularly negatively, which may further help people to prevent acting when they are confronted with these faces.

There are two potential psychological mechanisms underlying these additive effects of action and valence. Why are No-Go-To-Avoid-Losing faces evaluated the most negatively? One explanation is a low-level explanation, where avoidance of punishment is negative input, and inaction is also intrinsically negative input, resulting in an additive effect. Another explanation is a high-level inference explanation, where people have learned that No-Go-To-Avoid-Losing situations are particularly aversive. These might represent extreme negative situations in the ecological context, leading to overgeneralization such that the faces presented in this context are evaluated especially negatively. Future research could focus on gaining a better understanding of the underlying psychological mechanisms.

Absence of Evidence for Intergroup Differences

In the RL task, participants learned to respond or not respond to images of ingroup or outgroup members to obtain rewards or avoid punishments. We were interested in exploring differences in the action–valence asymmetry between ingroup and outgroup faces. We find evidence that the action–valence asymmetry in learning generalizes to both ingroup and outgroup faces to the same degree. One potential explanation for this null result is that the RL task stimulated individuation: Each individual was coupled with one specific RL condition. In doing so, participants learn (in)action outcome contingencies about one specific individual, and generalization to group membership is less valuable–i.e., it might lead to lower task performance. Similarly, prior studies have shown that individuation may reduce the activation of group stereotypes or category-based information (e.g., Rubinstein et al., 2018; Wheeler & Fiske, 2005), which in turn may decrease the likelihood of finding effects of group membership on performance.
To further investigate possible intergroup effects, future research could adapt the paradigm to the social category level instead of the individual level.

Implications

The findings of the current research have both theoretical and practical implications. On a theoretical level, this research is among the first to show evidence that inactions lead to less positive impressions than actions, over and above the effect of punishment signals. This has implications for the approach–avoidance literature. Similar to approach–avoidance research, we found that approach actions (Go-To-Win) led to more positive impressions of faces than avoidance actions (Go-To-Avoid-Losing; e.g., Kawakami et al., 2007; Phills et al., 2011; Slepian et al., 2012; Woud et al., 2013). However, note that as we manipulated outcomes instead of mere actions (c.f. Kawakami et al., 2007), more positive evaluations of the win faces compared to the lose faces is not necessarily an effect of the approach or avoidance action, but can also be explained by the mere co-occurrence of a reward or punishment with the face. Importantly, within the approach context (win conditions), action (Go-To-Win) leads to more positive impressions than inaction (No-Go-To-Win).

Similarly, within the avoidance context (avoid losing conditions), inaction (No-Go-To-Avoid-Losing) leads to more negative impressions than action (Go-To-Avoid-Losing). These latter findings demonstrate that approach–avoidance effects manipulated via the outcomes of instrumental (in)actions can be strengthened by aligning (in)actions with consequences.

On a practical level, the most evident practical implication is that when someone aims to positively influence impressions of individuals, it is most effective to combine actions with rewards. Moreover, on a more distal level, this research may have implications for the effectiveness of intergroup contact (Allport, 1954). It may help to inform under what circumstances it may or may not be effective in changing impressions on the group level. First, our research is in line with a large body of research that suggests that positive contact is an important precondition for better intergroup relations. In our study, although somewhat removed from real social interactions, we show causal evidence that actions with rewarding consequences create positive impressions (Go-To-Win). Second, our research gives some indication why negative contact is particularly damning (which is what prior correlational research shows, Aberson & Gaffney, 2009). In our study, when actions were punished (i.e., as happened in the No-Go-To-Avoid-Losing condition), this caused the least positive impression (even more so when inactions were punished; as in Go-To-Avoid-Losing). Thus, we speculate there is a strong, distinct aversion to being punished after performing actions: The same punishment may have a bigger impact when it follows an action (e.g., reaching out to an underrepresented group member) than when it follows an inaction (e.g., ignoring an underrepresented group member). Joining other recent insights on action and inaction in intergroup contexts (Allidina & Cunningham, 2021), the present work further suggests why negative contact is particularly problematic.

Outstanding Questions

There are several outstanding questions. First, our experimental design does not allow for generalizing effects to other faces than the faces used in the paradigm. Therefore, it is unknown whether the findings in the current research generalize to other Moroccan–Dutch or White–Dutch male faces, or to the social group in general (including among others women and children). An intergroup contact intervention is more effective if a positive impression generalizes to the group level. Therefore, future research could focus on the generalizability of the effect.

Second, our experimental design is not suitable for investigating long-term effects. Based on the current research it is unknown whether the effects are lasting. Future research could investigate the longevity of the effects.

Third, in this research we reduce social interactions to two essential components, i.e., (in)actions and consequences. In doing so, it is not clear whether the effects work out the same way in richer contexts, such as real social interactions. It could be the case, for example, that monetary consequences cannot fully substitute for social consequences, such as receiving smiles (Schmitz et al., 2020). Future research could investigate the effects in a more ecologically valid context.

Strengths and Limitations

The present research has several strengths. Our preregistered, theory-driven research revealed robust effects and new insights: The action–valence asymmetry in learning generalizes to different social groups and affects impressions. We show the latter for two measures, i.e., preferences and explicit evaluations, suggesting robustness of our results on different measures. The current research also has limitations. Although the research has high internal validity due to the controlled experimental set-up, this set-up also has its downsides. First, we used student samples so the findings cannot be directly generalized to other populations. Since the aim of the present research project is to gain fundamental insights (and not to generalize to the population level), we deemed it justified to collect data from a homogenous population. Second, this experimental work was conducted in a controlled environment and not in real-life situations which limits the external validity. Our operationalization of social interactions (action and inactions to win or avoid punishment) is very abstract and thus different from real-world social interactions. Consequently, the findings of this study cannot be generalized to actual social interactions. Future work can examine whether similar effects occur in situations with high ecological validity. Additionally, it is possible that this detachment from actual social interactions and/or individuation in the task prevented the occurrence of any intergroup effects. Finally, although we have ideas about the underlying psychological processes, we do not know how participants experience and interpret the tasks because we did not inquire about it.

Conclusion

Taken together, in the current research, we demonstrate that action–valence asymmetries during learning translate to additive effects on impression formation: Combining action with reward leads to the most positive impressions, while combining inaction with avoid punishment leads to the least positive impression. Moreover, we demonstrate that inactions lead to less positive impressions than actions, when the consequences are kept similar. This finding was most robust for the consequence of avoiding punishment. Generally, these findings provide new insights in the role of inactions in person perception. We suggest that impressions can best be changed by aligning (in)actions with consequences.

Acknowledgements

We thank Hubert Voogd (Radboud University) for programming the experiments. We thank dr. Marc Guitart-Masip (Karolinska Institute) for sharing the stimulus materials for Experiment 2.1. We thank members of the BRACE lab group (Radboud University) for useful discussions.

Supplemental materials

Table 2.1

Participant Characteristics Experiment 2.2

Ethnicity	Frequency
Dutch	33
German	15
Croatian	2
Ecuadorian	1
German American	1
Hungarian	1
Hungarian Cypriot	1
Italian	1
Portuguese	1
Polish	2
Turkish	1
Turkish Dutch	1

Table 2.2

Participant Characteristics Experiment 2.3

Ethnicity	Frequency
Dutch	27
German	7
Indonesian	1
American Hungarian	1
Arab	1
Bulgarian	1
Caucasian	2
Czech	1
Czech Vietnamese	1
Dutch Hungarian	1
Dutch Portuguese	1
Filipino	1
Greek	1
Hungarian	1
Indian	2
Iranian	1
Irish	1
Italian	2
Italian Dutch	1
Lithuanian	1
Luxemburg	1
Russian	1
South Asian Indian	1
Spanish	1
Thai	1

CHAPTER 3

Instrumental learning shapes social category evaluations and emotion recognition

This chapter is based on:

van Lent, T., Veling, H., Holland, R. W., Bijleveld, E., & Bijlstra, G. (2025). Reshaping the happy face advantage with reinforcement learning. *Cognition and Emotion*, 1–17. https://doi.org/10.1080/02699931.2025.2568553.

Abstract

Recognizing 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 3.1 (N = 40), we replicated that the Happy Face Advantage is influenced by social category membership. People were faster at recognizing happiness as happiness than anger as anger for White–Dutch faces, while no difference was found for Moroccan–Dutch faces. In Experiments 3.2–3.3 (N_{total} = 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 3.1 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.

Keywords: emotion recognition, facial expressions, prejudice, social categorization, instrumental learning

Open Science Practices: Open Data, Open Materials, Preregistered

Introduction

In everyday life, it is important to recognize 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 recognized is important for successful social interactions (Erickson & Schulkin, 2003). For example, people who are better at recognizing emotional expressions tend to exhibit more prosocial behaviors (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 recognized 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.

In general, people are faster at recognizing 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 recognizing 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, Koch, 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, Kleverwal, et al., 2019; Craig & Lipp, 2018a). 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 S3.1 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, Zhang, 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 favor 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 behaviors (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 behavioral information about unfamiliar individuals on emotion recognition (Lindeberg et al., 2019). Across two experiments, neutral White faces alongside positive or negative behavioral information about this specific individual were introduced. Subsequently, to test participants' memory, participants were asked to categorize whether an individual did something bad or good and received feedback for their decisions. Finally, participants were asked to recognize happy and angry emotional expressions of these same individuals. Findings indicated that associating new and unfamiliar individuals with positive or negative behaviors 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 and colleagues (2019) used reinforcement learning to test participants' memory of which *individual* is 'bad' or 'good'. That is, participants were asked to categorize individuals by the behavioral 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 optimized 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, Kogon, et al., 2022) and existing social categories (Traast et al., 2024). Interestingly, these effects can be maximized when consequences align with (in)action decisions (Liu et al., 2025; van Lent et al., 2025). For example, van Lent and colleagues (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 recognized 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 recognizing emotional expressions: We explore whether recognizing positive emotional expressions as positive for Moroccan–Dutch faces (compared to recognizing negative emotional expressions as negative) is facilitated and recognizing positive emotional expressions as positive for White–Dutch faces (compared to recognizing 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 3.1, we attempt to replicate whether social category membership moderates the HFA employing a direct replication of Bijlstra and colleagues (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 3.2–3.3 before participants perform the emotion recognition task. Thus, from Experiment 3.2 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 favor of Moroccan–Dutch social category members. In Experiments 3.2 and 3.3, we will replicate the procedure from Experiment 3.1 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.[13]

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; Van 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 3.1: https://osf.io/mn523, Experiment 3.2: https://osf.io/9x46w, Experiment 3.3: https://osf.io/kd6yx). The Ethics Committee Social Sciences at Radboud University approved this study (ECSW-2023-070).

Experiment 3.1

In Experiment 3.1, 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, Holland, et al., 2019, Experiment 2; Bijlstra, Kleverwal, et al., 2019; Hugenberg, 2005, Experiment 1). For this calculation, we used the average t statistic (4.12) and average sample size (N =

[13] 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.

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.

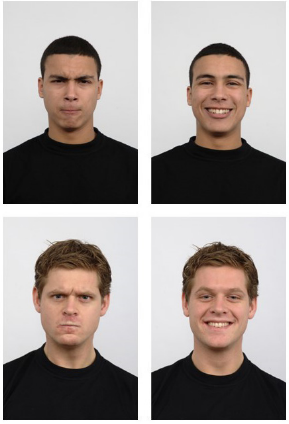
Materials and Procedure

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 recognize 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 3.1 for example stimuli). The task 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 randomized 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 recognize 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 familiarized 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 minutes and was programmed using Inquisit (*Inquisit 6*, 2022).

Figure 3.1

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).

Confirmatory Analyses

Emotion Recognition Task.

The response time needed to recognize emotional expressions served as the main dependent variable. As preregistered and similar to Bijlstra and colleagues (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 recognize 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.[14]

[14] To give more insight into the data and for the sake of completeness, we report additional exploratory analyses for all experiments in S3.2. Here, we have also included alternative ways of analyzing the data that were recommended by a reviewer.

Results

Confirmatory Analyses

Emotion Recognition Task. As preregistered and similar to Bijlstra and colleagues (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 Figure 3.3 and 3.4). For the sake of completeness, responses to happy emotional expressions were faster when displayed by White–Dutch than Moroccan–Dutch faces, *B* = -0.25, *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 hypothesized, 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 3.1 can be modified by reinforcement learning processes.

Experiment 3.2

In Experiment 3.2, 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 recognize emotional expressions. However, different from what was observed in Experiment 3.1, 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 3.1 ($B = 0.007$). 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 3.2.

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 and colleagues (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 minimize 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 2000 ms. 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 1750 ms in steps of 150 ms (see Figure 3.2 for an overview).

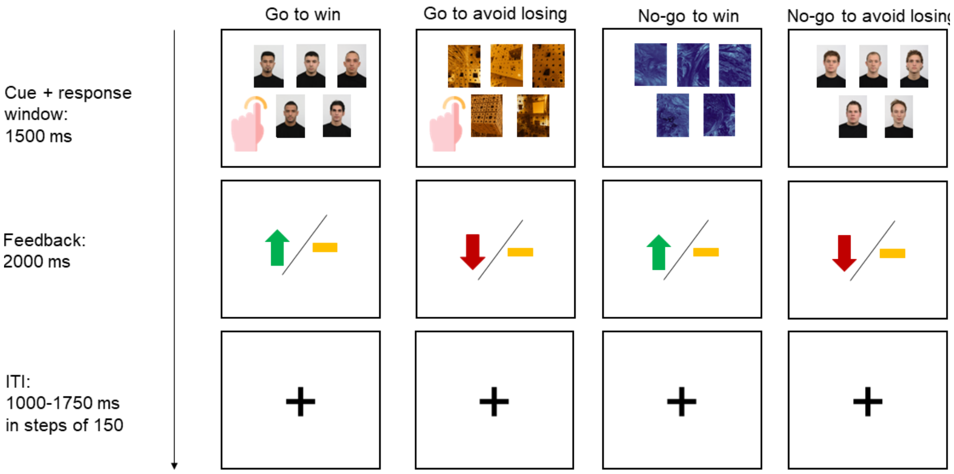
In total, the task included four categories (each including five pictures) with 60 trials per category resulting in 240 trials. After every 60 trials (15 trials per RL condition; Go-To-Win, No-Go-To-Win, Go-To-Avoid-Losing, No-Go-To-Avoid-Losing), participants had a 20 second break. The trials within the blocks were randomized. Before starting the task, participants took part in 10 practice trials per condition to get familiarized 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 3.1.

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 randomized 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 minutes and participants were paid according to their performance. The task was programmed using PsychoPy (Peirce et al., 2019).

Figure 3.2

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 3.1.

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.[15] Moreover, a random intercept of stimulus was included.

Results

Confirmatory Analyses

Emotion Recognition Task.

As preregistered and similar to Bijlstra and colleagues (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 Figure 3.3 and 3.4).

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 3.1.

[15] 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.

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 3.5). 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 3.3).

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 3.3).

Discussion

In Experiment 3.2, 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, signaling 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 3.2, but to counterbalance the order of response mapping between participants in Experiment 3.3 instead of within participants.

Experiment 3.3

Experiment 3.3 consisted of only one experimental block (144 trials) to test the influence of reinforcement learning on emotion recognition. Based on the results of Experiment 3.2, we changed our hypothesis for the emotion recognition task. We expected a main effect of expression valence on participants' response time needed to recognize 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 3.2 (B = 0.02). We decided to collect 72 participants, as the power analysis for Experiment 3.2 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 3.2. 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 minutes), 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 3.2.

Results

Confirmatory Analyses

Emotion Recognition Task.

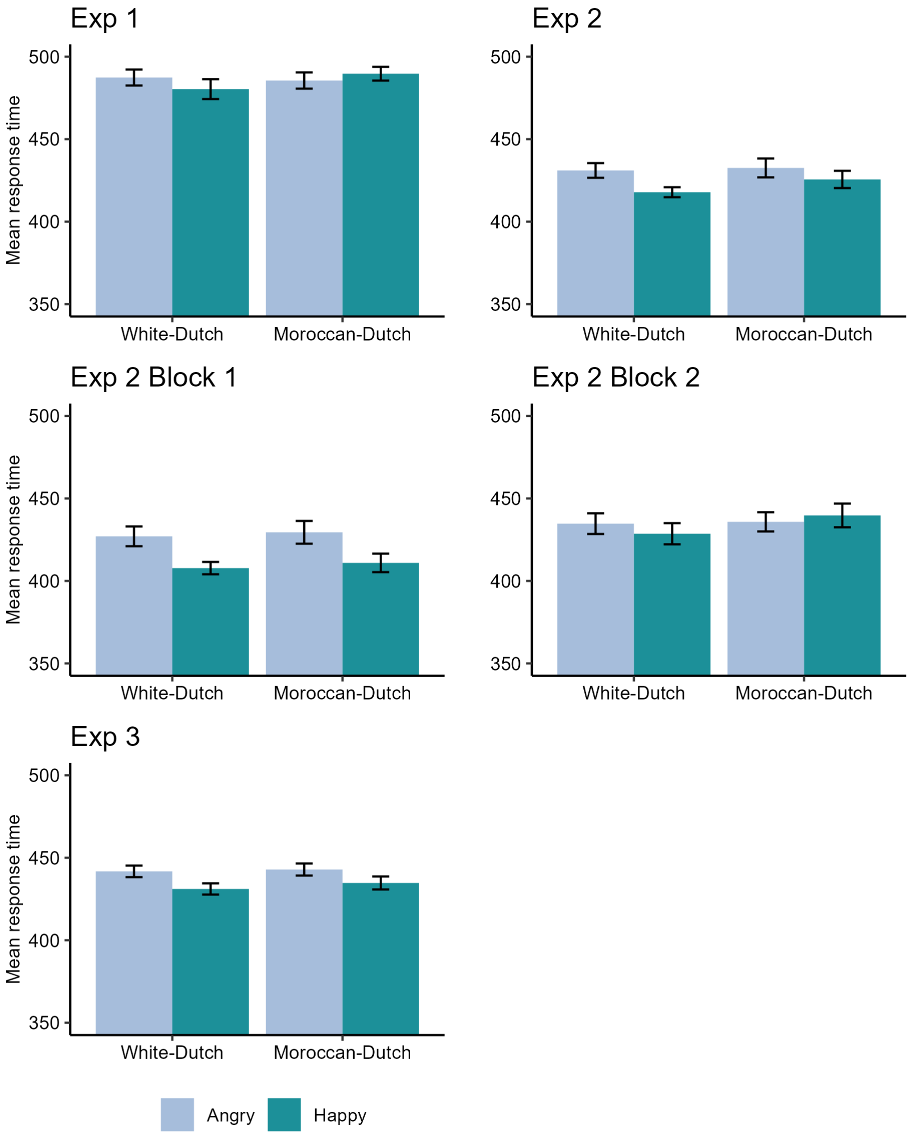
As preregistered and similar to Experiments 3.1 and 3.2, we excluded incorrect trials (7.19%) and response times below 200 ms (0.44%). In line with our preregistered hypothesis, there was a significant main effect of expression valence on response time, *B* = 0.01, *SE* = 0.004, *F*(1,81.88) = 6.50, *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 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 3.2, 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 Figure 3.3 and 3.4).[16]

[16] For completeness, and despite being severely underpowered for this analysis, we compared Experiment 3.1 with Experiments 3.2–3.3 (see S3.2). 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 3.1 and Experiments 3.2–3.3. Given the required sample size (see Footnote 13), this finding is not unexpected.

Figure 3.3

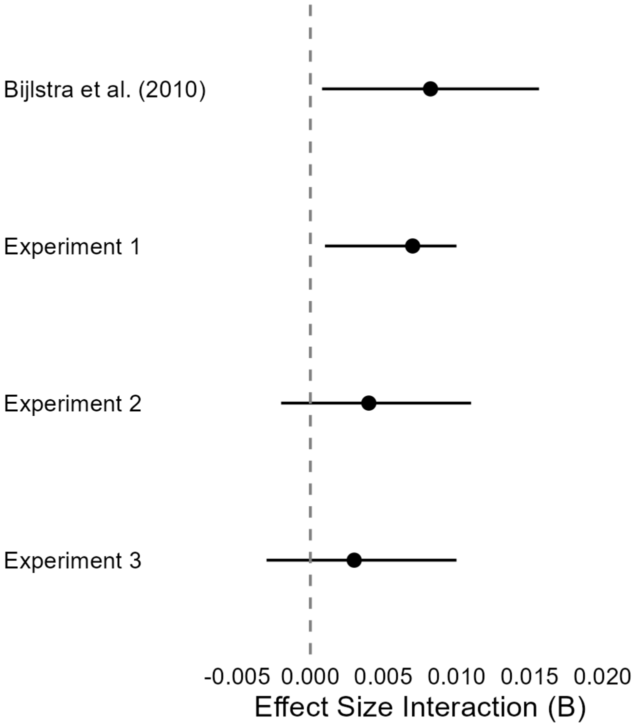
Results Emotion Recognition Task



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

Figure 3.4

Forest Plot Emotion Recognition Task



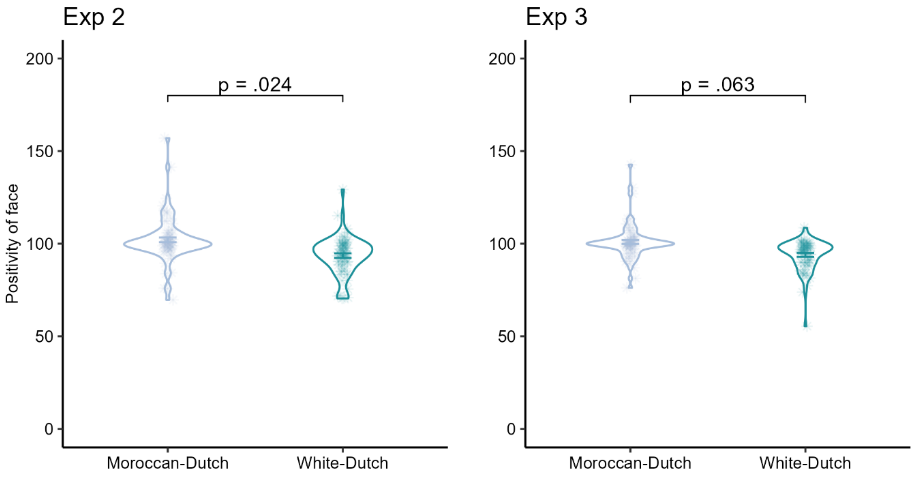
Note. Effect sizes of the interaction between social category and expression valence on response time for each experiment, including the original experiment by Bijlstra and colleagues (2010). A significant interaction effect was observed in Bijlstra and colleagues (2010) and Experiment 3.1, but not in Experiments 3.2 or 3.3. Error bars reflect confidence intervals around the effect size.

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 3.5). Although not significant, Moroccan–Dutch faces (*M* = 100.92, *SD* = 9.07) were numerically more positively evaluated than White–Dutch faces (*M* = 93.92, *SD* = 9.07).

Figure 3.5

Results Explicit Evaluation Task



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

Discussion

In Experiment 3.3 (and in line with results from Experiment 3.2), we found the preregistered main effect of valence of the expression on participants' response time needed to recognize emotional expressions, signaling 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, Kleverwal, et al., 2019; Craig, Koch, et al., 2017; Craig & Lipp, 2017, 2018a; 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 recognize evaluative consistent emotional expressions (Hugenberg & Bodenhausen, 2003). Second, Lindeberg and colleagues (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 3.2, 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 computerized 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 3.2 that the reinforcement learning task is initially effective, but after a short pause during which we swap the keys to recognize 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 generalizes 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 generalizes 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 recognizing 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 recognizing 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, Duzel, et al., 2014; Guitart-Masip et al., 2012), 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 and colleagues (2025) investigated its impact on food choice, and van Lent and colleagues (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 generalize 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 3.2–3.3, 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 S3.2). Finally, it is unknown whether the absence of a 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 behavior.

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.

Acknowledgements

We thank Thijs Verwijmeren from Radboud University for helping with programming Experiments 3.2–3.3. We thank Erik van den Berge from Radboud University for helping with color adjustments to the fractal stimuli in Experiments 3.2–3.3. 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.

Supplemental materials

S3.1: Overview Table Social Category Moderation of the HFA

Table 3.1 *Social Category Moderation of the HFA Across Key Studies*

Authors and Experiment	Participants	Moderating Category on HFA (larger HFA > smaller HFA)
Becker et al. (2007) Experiment 2	*N* = 38	Gender (female > male)
Becker et al. (2007) Experiment 4	*N* = 21	Gender (female > male)
Bijlstra et al. (2010) Experiment 1	*N* = 42	Ethnicity (White–Dutch > Moroccan–Dutch)
Bijlstra et al. (2010) Experiment 2	*N* = 78	Gender (female > male)
Bijlstra, Kleverwal, et al. (2019) Experiment 1	*N* = 60	Age (young male > old male)
Craig and Lipp (2017) Experiment 1	*N* = 30	Gender (female > male)
Craig and Lipp (2018a) Experiment 1	*N* = 32	Age (young male > old male)
Craig and Lipp (2018a) Experiment 2	*N* = 28	Age (young male > old male)
Craig and Lipp (2018b) Experiment 1a	*N* = 35	Gender (female > male)
Craig and Lipp (2018b) Experiment 1a	*N* = 35	Ethnicity (White targets > Black targets)
Craig and Lipp (2018b) Experiment 1b	*N* = 66	Gender (female > male)
Craig, Koch, et al. (2017) Experiment 1	*N* = 29	Ethnicity (White males > Black males)
Craig, Koch, et al. (2017) Experiment 1	*N* = 29	Gender (female > male)

S3.1: Overview Table Social Category Moderation of the HFA

Table 3.1 *Social Category Moderation of the HFA Across Key Studies*

Authors and Experiment	Participants	Moderating Category on HFA (larger HFA > smaller HFA)
Hugenberg (2005) Experiment 1	*N* = 20	Ethnicity (White males > Black males)
Hugenberg (2005) Experiment 2	*N* = 40	Ethnicity (White males > Black males)
Hugenberg and Sczesny (2006) Experiment 1	*N* = 80	Gender (female > male)
Hugenberg and Sczesny (2006) Experiment 2	*N* = 77	Gender (female > male)
Lipp et al. (2015) Experiment 1	*N* = 29	Ethnicity (White males > Black males)
Lipp et al. (2015) Experiment 2	*N* = 26	Ethnicity (White males > Black males)
Stebbins and Vanous (2015) Experiment 1	*N* = 45	Gender (female > male)

Note. This table presents a moderation effect of different social categories on the HFA.

S3.2: Additional Exploratory Analyses

Number of Errors in the Emotion Categorization Task

The number of errors participants made when categorizing angry and happy expressions served as the dependent variable (1 = wrong, 0 = correct). To determine differences in the number of errors when categorizing happy and angry expressions between White–Dutch and Moroccan–Dutch faces, we conducted a binomial generalized 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.

Experiment 3.1.

There was a significant interaction between social category and expression valence on the number of errors, *B* = 0.19, *SE* = 0.06, χ^2(14) = 7.79, *p* = .005, 95% CI [0.04, 0.33]. There were more errors in responses to happy emotional expressions (*M* = 0.09, *SD* = 0.05) than in responses to angry emotional expressions (*M* = 0.06, *SD* = 0.06) when displayed by Moroccan–Dutch faces, *B* = -0.44, *SE* = 0.15, *p* = .003. No difference was found between the number of errors to happy (*M* = 0.08, *SD* = 0.07) and angry (*M* = 0.09, *SD* = 0.07) emotional expressions displayed by White–Dutch faces, *B* = 0.32, *SE* = 0.19, *p* = .089. For the sake of completeness, there were more errors in responses to angry emotional expressions when displayed by White–Dutch than Moroccan–Dutch faces, *B* = 0.42, *SE* = 0.16, *p* = .01, there were more errors in responses to happy emotional expressions when displayed by Moroccan–Dutch than White–Dutch faces, *B* = -0.34, *SE* = 0.16, *p* = .031. There was no significant main effect of social category on number of errors, *B* = 0.02, *SE* = 0.05, χ^2(14) = 0.09, *p* = .764, 95% CI [-0.11, 0.14], and no significant main effect of expression valence on number of errors, *B* =-0.03, *SE* = 0.06, χ^2(14) = 0.4, *p* = .521, 95% CI [-0.16, 0.08] (see Table 3.2).

Experiment 3.2.

There was no significant interaction between social category and expression valence on the number of errors, *B* = 0.05, *SE* = 0.06, χ^2(14) = 0.76, *p* = .383, 95% CI [-0.06, 0.19], no significant main effect of social category on the number of errors, *B* =-0.005, *SE* = 0.06, χ^2(14) = 0.01, *p* = .934, 95% CI [-0.12, 0.12], and no significant main effect of expression valence on the number of errors, *B* =-0.09, *SE* = 0.06, χ^2(14) = 2.30, *p* = .129, 95% CI [-0.22, 0.03] (see Table 3.2).

Experiment 3.3.

There was no significant interaction between social category and expression valence on number of errors, *B* = 0.1, *SE* = 0.05, χ^2(14) = 2.87, *p* = .09, 95% CI [-0.01, 0.21], no significant main effect of social category on number of errors, *B* =-0.002, *SE* = 0.05, χ^2(14) = 0.001, *p* = .971, 95% CI [-0.11, 0.11], and no significant main effect of expression valence on number of errors, *B* =-0.13, *SE* = 0.07, χ^2(14) = 3.27, *p* = .07, 95% CI [-0.27,0.01] (see Table 3.2).

Table 3.2
Mean and Standard Deviations (in Parentheses) for Proportion of Errors

	Experiment 3.1		Experiment 3.2		Experiment 3.3	
	Happy	Angry	Happy	Angry	Happy	Angry
Moroccan–Dutch	0.09 (0.05)	0.06 (0.06)	0.09 (0.06)	0.07 (0.05)	0.08 (0.07)	0.07 (0.08)
White–Dutch	0.08 (0.07)	0.09 (0.07)	0.08 (0.05)	0.08 (0.06)	0.07 (0.09)	0.08 (0.09)

Reinforcement Learning Go/No-go Task

We ran four separate intercept only models (one for each condition) to identify whether participants performed above chance level. This model included a random intercept of participant. *P*-values were determined with 95% confidence intervals using likelihood profiling.

Experiment 3.2.

In all RL conditions, participants performed significantly above chance level. That is, participants performed significantly above chance level in the Go-To-Win RL Condition (*M* = 0.77, *SD* = 0.28), *B* = 1.87, *SE* = 0.30, *p* < .05, 95% CI [1.27, 2.49], Go-To-Avoid-Losing RL Condition (*M* = 0.78, *SD* = 0.14), *B* = 1.43, *SE* = 0.11, *p* < .05, 95% CI [1.21, 1.64], No-Go-To-Win RL Condition (*M* = 0.68, *SD* = 0.27), *B* = 1.04, *SE* = 0.26, *p* < .05, 95% CI [0.53, 1.55] and No-Go-To-Avoid-Losing RL condition (*M* = 0.73, *SD* = 0.16), *B* = 1.14, *SE* = 0.10, *p* < .05, 95% CI [0.94, 1.34]. This indicates that participants learned the optimal response in each RL condition above chance level.

Experiment 3.3.

In all RL conditions, participants performed significantly above chance level. That is, participants performed significantly above chance level in the Go-To-Win RL Condition (*M* = 0.78, *SD* = 0.21), *B* = 1.74, *SE* = 0.19, *p* < .05, 95% CI [1.38, 2.13], Go-To-Avoid RL Condition (*M* = 0.81, *SD* = 0.12), *B* = 1.60, *SE* = 0.09, *p* < .05, 95% CI [1.43, 1.79], No-Go-To-Win RL Condition (*M* = 0.71, *SD* = 0.23), *B* = 1.34, *SE* = 0.22, *p* < .05, 95% CI [0.91, 1.78], and No-Go-To-Avoid-Losing RL condition (*M* = 0.76, *SD* = 0.13), *B* = 1.23, *SE* = 0.08, *p* < .05, 95% CI [1.06, 1.40]. This indicates that participants learned the optimal response in each RL condition above chance level.

Differences Between Trained and Untrained Faces in RL GNG Task

To investigate whether there are differences in emotion recognition between trained and untrained faces in the RL GNG Task, we conducted a linear mixed model. This model included the within-participant factors social category (White–Dutch/Moroccan–Dutch), expression valence (positive/negative), and training (trained/untrained). Moreover, this model included a random intercept of stimulus and participant as well as random slopes for social category, expression valence, and training for participant and random slope for social category, expression valence, and training for participant and random slope for training for stimulus. The response time needed to categorize emotional expressions served as the main dependent variable.

Experiment 3.2.
There was no significant three-way interaction between social category, expression valence, and training on response time, *B* = 0.006, *SE* = 0.003, *F*(1,43.63) = 3.19, *p* = .081. We do not find evidence that the interaction between social category and expression valence is different for trained versus untrained faces.

Experiment 3.3.
There was no significant three-way interaction between social category, expression valence, and training on response time, *B* =-0.002, *SE* = 0.003, *F*(1,41.02) = 0.58, *p* = .452. We do not find evidence that the interaction between social category and expression valence is different for trained versus untrained faces.

Although the three-way interaction for Experiment 3.2 was not significant, we conducted follow-up analyses. These analyses show that for trained faces (5 faces), there was a significant interaction between social category and expression valence on response time, *B* = 0.05, *SE* = 0.019, *p* = .007. For untrained faces (13 faces), there was no significant interaction between social category and expression valence on response time, *B* = 0.006, *SE* = 0.018, *p* = .734. However, a major limitation is the low number of trials per cell involving trained faces (i.e., 10 trials per cell), whereas the untrained faces have 26 trials per cell. Since our power analysis was based on 36 trials per cell, the current number of 10 trials per cell is insufficient. Consequently, we have serious doubts about the robustness of this finding and prefer not to draw any conclusions from the data at this stage. This doubt is further reinforced by the fact that we clearly do not observe a three-way interaction in Experiment 3.3.

Comparing Emotion Recognition Between Experiments

For completeness, and although we are underpowered for this analysis, we analyzed all three datasets together, in which we compared Experiment 3.1 versus Experiments 3.2 and 3.3 combined. There was no significant three-way interaction between social category, expression valence and experiment on response time, *B* = 0.003, *SE* = 0.004, *F*(2,49.54) = 0.41, *p* = .664. We do not find evidence that the interaction between social category and expression valence is different for Experiment 3.1 versus Experiments 3.2 and 3.3.

Moreover, we combined the datasets of Experiments 3.2–3.3 and found a significant main effect of expression valence on response time, *B* = 0.01, *SE* = 0.003, *F*(1,89.43) = 11.62, *p* < .001, 95% CI [0.01, 0.02]. Responses to happy emotional expressions were faster than responses to angry emotional expressions when displayed by White–Dutch faces, *B* = 0.03, *SE* = 0.008, *p* < .001. For Moroccan–Dutch faces, this difference was not significant, *B* = 0.01, *SE* = 0.009, *p* = .108.

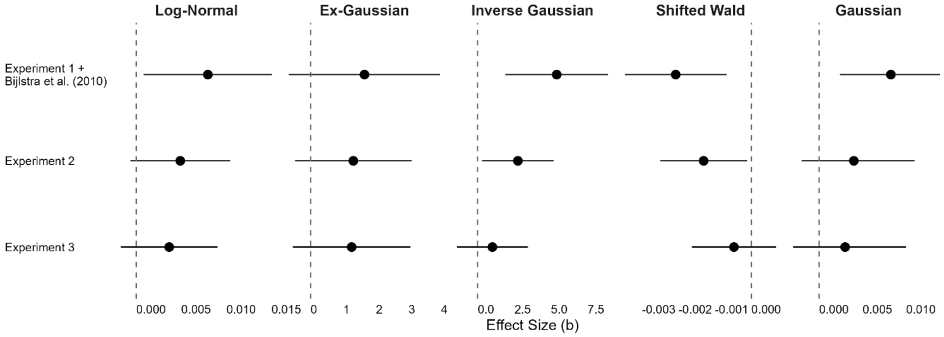
Finally, we analyzed data of block 1 of Experiment 3.2 and Experiment 3.3 and found a significant main effect of expression valence on response time, *B* = 0.015, *SE* = 0.004, *F*(1,105.19) = 16.71, *p* < .001, 95% CI [0.01, 0.02]. Responses to happy emotional expressions were faster than responses to angry emotional expressions when displayed by White–Dutch, *B* = 0.04, *SE* = 0.01, *p* < .001, and Moroccan–Dutch faces, *B* = 0.03, *SE* = 0.009, *p* = .016.

Robustness Checks

In Figure 3.1 Supplemental Materials, we have plotted alternative approaches to analyze the response time data as robustness check. These analyses were executed on top of our preregistered confirmatory analyses to assess whether the (absence of) the interaction between social category and expression valence on response time is robust across different analytic strategies. Figure 3.1 Supplemental Materials displays the effect sizes including confidence intervals for each analytic strategy (on the y-axis) per experiment. We combined the data from the original study (Bijlstra et al., 2010) and Experiment 3.1 to increase statistical power. In this combined dataset, four out of five analysis strategies revealed a significant interaction effect, demonstrating robustness of the interaction. These findings align with our hypothesis, as we expected an interaction effect. Experiment 3.2 yielded mixed results: Three out of five analyses indicated a significant interaction effect, despite our expectation of no such effect. This is not entirely surprising, as the interaction pattern re-emerges in Block 2 of Experiment 3.2. Importantly, in Experiment 3.3—our most rigorous test of the effects of the RL GNG training—all five analyses consistently found no significant interaction effect. This provides robust evidence supporting the absence of the effect.

Figure 3.1 Supplemental Materials

Forest Plot Robustness Checks Emotion Recognition Task



Note. Effect sizes of the interaction between social category and expression valence on response time for different analyses strategies for each experiment, including the original experiment by Bijlstra and colleagues (2010). Error bars reflect confidence intervals around the effect size.

CHAPTER 4

Instrumental learning shapes consideration sets and preferences

This chapter is based on:

van Lent, T., Veling, H., Holland, R. W., Bijleveld, E., & Bijlstra, G. (2025). Who do you consider? The role of individual-based and group-based values in constructing consideration sets [Manuscript submitted for publication].

Abstract

How do discriminatory decisions come about? To arrive at a decision, people create a set of options they consider relevant to the decision, such as whom to hire, which is called the consideration set. Across four preregistered experiments (N_{total} = 1600, UK participants), we systematically investigated whether (1) individuals from advantaged versus marginalized social groups (group-based values) and (2) individuals with higher competency (individual-based values) are more likely to enter the consideration set, even when values are not relevant for decisions. In Experiments 4.2–4.4, although group-based values did successfully transfer to evaluations (Experiment 4.2) and preferences (Experiment 4.4), they did not influence consideration sets. In Experiments 4.1–4.3, however, we robustly show that the higher the individual-based value of a person, the more likely the person is to be considered. We conclude that, even when the value is irrelevant to the decision at hand, individual-based values influence consideration sets.

Keywords: reinforcement learning, value-based decision making, prejudice, stereotypes, discrimination

Open Science Practices: Open Data, Open Materials, Preregistered

Introduction

Even though self-reported levels of prejudice have decreased in recent years (Charlesworth & Banaji, 2019, 2022), discrimination in decision making remains widespread. For example, recent research indicates unfair and unequal treatment in hiring (Quillian & Lee, 2023), housing (Auspurg et al., 2019), and judicial decisions (Galvan et al., 2024). Although people tend to endorse equal treatment, discrimination continues to exist in their actions (Dovidio & Gaertner, 2000). To better understand the roots of biased decision making, it is essential to gain insight into the decision-making processes that lead to discriminatory outcomes. Here, we systematically examine how biases influence the *pre-decision phase*; that is, we investigate who people consider before they make a decision.

When making a decision, such as whom to promote in an organization, people often cannot evaluate all available choice options due to time constraints and cognitive limitations (Krajbich et al., 2010; Oeberst & Imhoff, 2023; Simon, 1955), and therefore only evaluate a few options from all possibilities. That is, before they make a decision, people construct a *consideration set* out of all possible choice options (Howard & Sheth, 1969). Such a consideration set is a collection of choice options stored in working memory. Being included in the consideration set is a prerequisite for selection (Morris et al., 2021). As a result, if a choice option is not considered, this choice option will automatically not be chosen. Given that people cannot consider every possible choice option due to time and cognitive limitations, how do people construct this consideration set?

Recent research demonstrates that the construction of this consideration set depends on the value associated with the different choice options. Choice options with higher value (i.e., that are initially preferred more) have a higher probability of being considered than options with lower value (Morris et al., 2021), suggesting that biases affect the options people consider choosing from. Morris and colleagues (2021) demonstrate this through both correlational and experimental designs. Importantly, this holds even when the value of the options is irrelevant to the decision at hand. For example, this work showed dishes with a high subjective value have a higher probability of being considered, even in situations where this dish is not appropriate (e.g., a pizza enters the consideration set even when someone just had dental surgery and cannot chew; see also Posavac et al., 1997). Taken together, this suggests that positively valued choice options do come to mind more readily than negatively valued choice options.

In the current studies, we investigate consideration set construction in social decision making to determine whether this perspective can aid in understanding discrimination in decision making. That is, when making decisions about people (e.g., whom to give a promotion to), some people may be associated with higher value (i.e., are preferred more) than others based on an irrelevant criterion such as their social group membership or attractiveness. For example, people generally prefer ingroup members over outgroup members (Ratner et al., 2014), a preference reflected in brain regions linked to reward and value processing (Van Bavel et al., 2008). Similarly, people generally prefer attractive people over unattractive people (Dion et al., 1972; Griffin & Langlois, 2006). Based on earlier studies (Morris et al., 2021; Posavac et al., 1997), we predict that individuals from more negatively valued social groups have a smaller chance to enter consideration sets than individuals from more positively valued social groups, and for that reason alone, are already less likely to be selected.

When constructing a consideration set, value can be derived from several sources. First, as described above, value can be derived from someone's social category or group membership (i.e., group-based values; Fiske & Neuberg, 1990). Second, when deriving value, people can also individuate or subtype. Instead of relying on prejudicial or stereotypical information, people then derive value from the individual level (i.e., individual-based values), e.g., someone's characteristics or behaviors. People may use both social category and individual information to construct their consideration set, with various factors determining the value of each. For example, highly diagnostic individuating information decreases reliance on stereotypical information (Rubinstein et al., 2018), and willingness and ableness to devote cognitive resources to individuate decreases reliance on categorical information (Fiske et al., 2018). Here, we investigate the interplay between group-based and individual-based values on the consideration set. For example, with equal individual-based values, do people consider individuals from the advantaged social groups more than individuals from marginalized social groups?

The Present Research

We investigate in four experiments[17] whether individual-based and group-based values influence the construction of one's consideration set. In Experiment 4.1, we conduct a conceptual replication of Morris and colleagues (2021, Experiments 4–6), to investigate whether their findings generalize to decision making about people. Here, we investigate whether individuals with higher individual-based values have a higher probability of being considered than those with lower individual-based values. In Experiments 4.2–4.3, we combine both individual-based and group-based values and investigate how the combination of these values influences the construction of one's consideration set. Finally, in Experiment 4.4, we single out group-based values and investigate whether individuals with higher group-based values have a higher probability of being considered than those with lower group-based values.

[17] We conducted five experiments but report only four. We prematurely stopped data collection for Experiment 4.0 and, therefore, do not report its results (see S4.1).

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). All data, analysis code, and research materials are available via:
https://osf.io/m3z7g/?view_only=dbc4ba2321e34777840f061fa3ffe249.
Data were analyzed using R (v4.3.1; R Core Team, 2023) and the packages tidyverse (v2.0.0; Wickham et al., 2019), cowplot (v1.1.3; Wilke, 2024), Rmisc (v1.5.1; Hope, 2022), lme4 (v1.1.34; Bates et al., 2015), afex (v1.3.0; Singmann et al., 2023), pbkrtest (v0.5.2; Halekoh & Højsgaard, 2014), emmeans (v.1.8.9; Van Lenth, 2023) and parallel (v4.3.1.; R Core Team, 2023). All experiments were preregistered
(Experiment 4.1: https://osf.io/tpmjw/?view_only=037091fb0dc240c58a154222c08d07c1,
Experiment 4.2: https://osf.io/6zp7c/?view_only=207105080689412e9f3c90d3c6f332ee,
Experiment 4.3: https://osf.io/48rgw/?view_only=d8b54b2ed75b4b1b801cac6325cdd772,
Experiment 4.4: https://osf.io/et2x7/?view_only=5a9ee5d9a2a145ab8c03e7110c086742).
This research conforms to the light track procedure by the Ethics Committee Social Sciences of Radboud University (ECSW-LT-2023-2-1-43965, ECSW-LT-2024-2-4-23310 and ECSW-LT-2024-11-7-21105).

Experiment 4.1

In Experiment 4.1, we investigate whether individuals with higher individual-based values have a higher probability of being considered than those with lower individual-based values.

Method

Sample Size Justification

The sample size of 400 was determined using G*Power (Faul et al., 2009). We decided to use OR = 1.15 (α = .05, power = .80) as the theoretical minimum effect size of interest, which is considered a small effect size.[18] Thus, we decided to continue data collection until we recruited 400 participants who met all pre-registered inclusion criteria (see S4.2; we applied this procedure to all experiments). Participants were rewarded £1,50 for participating and could earn more based on their performance (up to £0,5).

Participants

We recruited 501 participants and after exclusions (see S4.2), the sample size was 400 participants (M_{age} = 39.11, SD_{age} = 12.25, 18–64 years old, 208 women, 186 men, 3 non-binary, 1 other, 2 skipped this question). In all experiments, we recruited UK participants via Prolific.

[18] We acknowledge that the G*Power power analysis does not match the mixed models analysis used. It served as a rough estimate, with formal power analyses planned based on Experiment 4.1 for Experiments 4.2–4.3.

Materials and Procedure

Consideration Set Task.

To investigate the influences of individual-based values on the consideration set, we adapted the experimental procedure from Morris and colleagues (2021). Our experiment consisted of an instruction, a learning, and a decision-making stage. First, in the instruction stage, participants provided consent. Next, we instructed participants to imagine themselves as a teacher who had to remember the performance of their eight students. Participants were informed that each student was represented by an avatar and a name. These avatars were created using https://avatarmaker.net/create-avatar and were matched on attractiveness to avoid confounds. For the same purpose, unfamiliar names of equal length and complexity were generated using ChatGPT (OpenAI, 2023; 'Hastir', 'Linpon', 'Kailas', 'Marlit', 'Kimnol', 'Milgen', 'Plorin', 'Taknir').

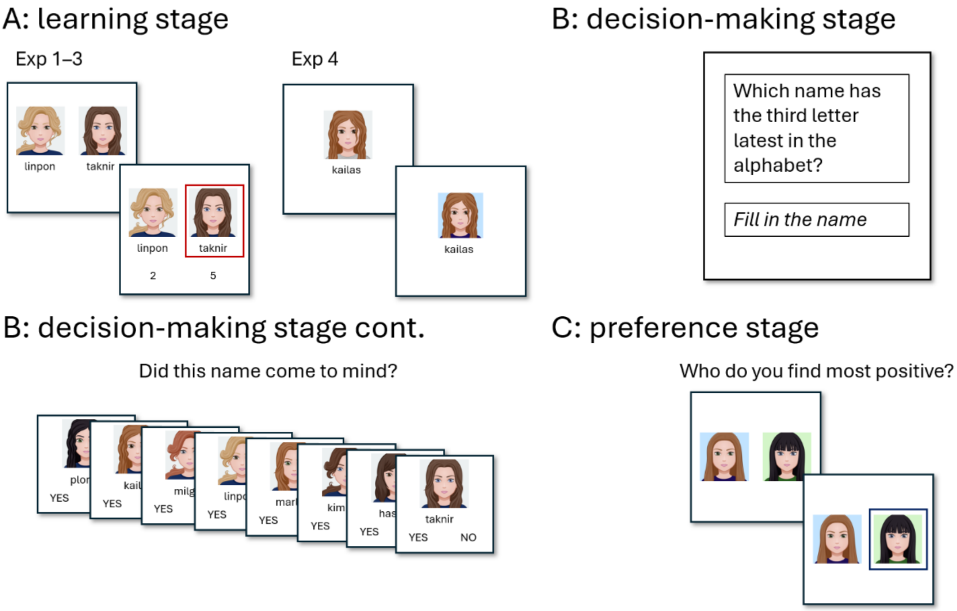
Second, in the learning stage, participants familiarized themselves with the students and their values. Participants were asked to make consecutive choices between two students and select the student with the highest performance level. On each trial, participants were presented with two students and asked to choose one of them by pressing 'W' or 'P'. After participants selected a student, the points associated with both students were revealed. The performance level was indicated by points ranging from 1 to 8. The better the performance, the more points. Points of their selected students were added up, and the total points earned were visible to participants. After completing, the points would be converted to a monetary bonus. The learning stage consisted of 56 trials. In the first 19 trials, participants could ask for a hint. By clicking the hint, both the students' points appeared. The combination of name, avatar, and performance level was randomized per participant.

Next, during the decision-making stage, participants were asked a neutral question about the eight students to enable them to construct a consideration set. That is, participants were asked to come up with a name whose third letter comes late in the alphabet, with better answers earning more money. There was no relationship between the values in the learning stage and the decision-making stage, which was emphasized to the participants. Note that this procedure allows us to assess whether a person with a higher individual value has a higher probability to be considered, even under the condition where this (performance) value is irrelevant to the question at hand. Participants had 25 seconds to choose a name. Next, participants were asked to report which names they considered while answering the question to measure their consideration set. Here, a picture of each student was presented consecutively together with their name in random order, and participants indicated whether this student came to mind as a candidate for the decision. This was the main dependent variable. If participants use values to generate choice options, they are more likely to consider names with high value from the learning stage despite these values being irrelevant in the decision-making stage.

Finally, participants were asked to report their age and gender. The task lasted approximately 13 minutes (see Figure 4.1 for an overview of the consideration set task).

Figure 4.1

Overview of the Consideration Set Task



Note. (A) Learning stage in which participants are asked to make consecutive choices between two students and select the student with the highest performance level (left panel) or to categorize which student belongs to which group (right panel). (B) Decision-making stage in which participants are first asked a neutral question about the eight students to enable them to construct a consideration set, and then to indicate which students they considered. (C) Preference stage in which participants are asked to decide which of the two students appears most positive to them at that moment (only in Experiments 4.3–4.4).

Confirmatory Analyses

For all experiments, we used the same analysis strategy. That is, to determine whether the student’s presence in the consideration set (yes or no) is influenced by individual-based values (Experiments 4.1–4.3) and/or group-based values (Experiments 4.2–4.4), we conducted binomial generalized linear mixed models with a maximal random-effects structure (Barr et al., 2013; see S4.3 for exact model specifications).[19]

[19] For Experiments 4.2–4.3, we corrected an incomplete preregistered random slope structure to ensure maximal models and report them.

Results & Discussion

Confirmatory Analyses

As preregistered, there was a significant positive linear effect of individual-based values on presence in the consideration set, $B = 0.18$, $SE = 0.03$, $\chi^2(10) = 17.86$, $p < .001$, 95% CI [0.13, 0.24], OR = 1.2. With each point increase in individual-based value, students were 1.2 times more likely to be considered (see Figure 4.2). Replicating Morris and colleagues (2021), and in line with our preregistered hypothesis, we found evidence that individual-based values positively influence presence in the consideration set.

Experiment 4.2

In Experiment 4.2, we investigate whether individuals with more positive group-based values have a higher probability of being considered than those with more negative group-based values. Secondly, we aimed to replicate that individuals with higher individual-based values have a higher probability of being considered than those with lower individual-based values.

Method

Sample Size Justification

The sample size was determined using a simulation-based power analysis with data from Experiment 4.1. A total of 400 participants allowed us to detect a main effect of individual-based values as small as $B = 0.7$ ($\alpha = .05$, power = .80). Participants were rewarded £1,20 for participating and could earn more based on their performance (up to £0,5).

Participants

We recruited 458 participants and after exclusions (see S4.2), the sample size was 400 participants ($M_{age} = 40.46$, $SD_{age} = 11.68$, 18–64 years old, 188 women, 206 men, 2 non-binary, 1 other, 3 skipped this question).

Materials and Procedure

Consideration Set Task.
To investigate how the combination of individual-based and group-based values influences the consideration set, we used the same paradigm as in Experiment 4.1 and added group-based values. Again, participants were asked to remember the performance of eight students. This time, there were four students each in the green and blue groups. In both groups, the performance levels ranged from 1 to 4.

In the instructions, before the learning and decision-making stages, participants were told that one group came from a lower-educated, poorer, and more criminal society and the other from a higher-educated, more affluent, and non-criminal society (counterbalanced as green or blue; see S4.4 for group descriptions based on Schultner et al., 2024), inducing group-based values. Group descriptions were adjusted to the UK context and pretested on valence. In this pretest, 40 participants (M_{age} = 39.45, SD_{age} = 12,68, 21–64 years old, 25 women, 14 men, 1 non-binary) read the group descriptions and subsequently evaluated four members of each group on a 200-point scale (0 = *very negative*, 200 = *very positive*). Members from the positive group (*M* = 135.14, *SD* = 24.93*)* were significantly more positively evaluated than members from the negative group (*M* = 119.37, *SD* = 24.93), *B* = 7.78, *SE* = 2.76, *F*(1,38.44) = 7.96, *p* = .008, 95% CI [-13.26,-2.31]. The remainder of Experiment 4.2 was identical to Experiment 4.1.

Results

Confirmatory Analyses

As preregistered, there was a significant positive linear effect of individual-based values on presence in the consideration set, *B* = 0.33, *SE* = 0.06, χ^2(24) = 14.2, *p* < .001, 95% CI [0.22, 0.44]. Contrary to our hypothesis, there was no significant difference between the positive (*M* = 0.36, *SD* = 0.23) and negative (*M* = 0.38, *SD* = 0.23) group on presence in the consideration set, *B* =-0.07, *SE* = 0.06, χ^2(24) = 1.31, *p* = .253, 95% CI [-0.19, 0.05]. For the sake of completeness, although not hypothesized, there was a significant interaction between group-based values and individual-based values on presence in the consideration set, *B* = 0.08, *SE* = 0.04, χ^2(24) = 3.95, *p* = .047, 95% CI [0.00, 0.15]. For the positive group, the data pattern shows a stronger positive effect, *B* = 0.41*, SE* = 0.07*, p* <.001, than for the negative group, *B* = 0.26*, SE* = 0.06*, p* < .001 (see Figure 4.2).

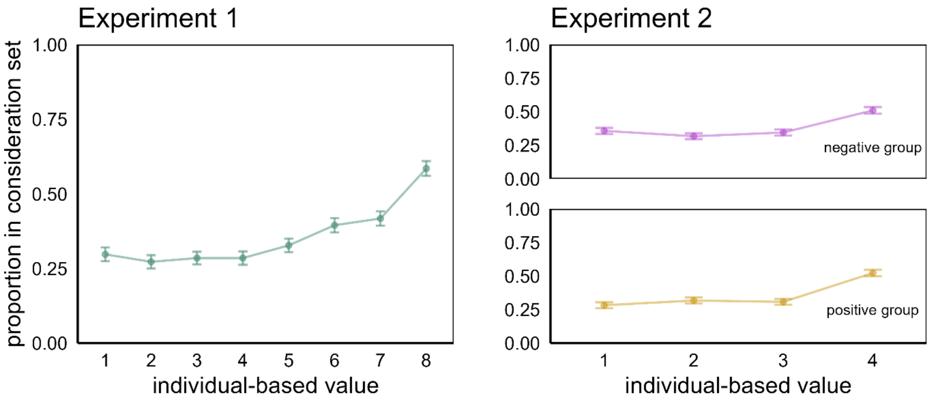
Exploratory Analyses

We explored whether receiving hints in the first 19 trials affected potential group-based values effects. That is, using the hint button revealed both students' individual-based values. In doing so, hints may have made the task less ambiguous and encouraged individuation: Participants learned the performance of each individual rather than relying on group membership information when considering individuals. To examine this potential confound, we analyzed the 42 participants who did not make use of this functionality.

Here, there was a significant interaction between group-based values and individual-based values on presence in the consideration set, *B* = 0.29, *SE* = 0.12, χ^2(15) = 4.80, *p* = 0.028, 95% CI [-0.60, 0.005]. Although not significant, for the positive group, the data seem to show a positive trend, *B* = 0.32*, SE* = 0.18*, p* = .075, whereas for the negative group, the data seem to show a negative trend, *B* =-0.25*, SE* = 0.17*, p* = .141. For the sake of completeness, although being underpowered for this analysis, there was no significant effect of individual-based values on presence in the consideration set, *B* = 0.03, *SE* = 0.13, χ^2(15) = 0.07, *p* = .796, 95% CI [-0.25, 0.35], and there was no significant difference between the positive (*M* = 0.41, *SD* = 0.25) and negative group (*M* = 0.38, *SD* = 0.25) on presence in the consideration set, *B* = 0.12, *SE* = 0.20, $\chi^2(15) = 0.33$, *p* = .565, 95% CI [-0.64, 0.36] (see S4.5 for a figure).

Figure 4.2

Results Consideration Set Experiments 4.1–4.2



Note. The mean proportion considered per within-participant condition is visualized. Error bars reflect within-participant standard error around those means.

Discussion

In line with Experiment 4.1 and our preregistered hypothesis, we found consistent evidence that individual-based values positively influence the consideration set. However, contrary to our preregistered hypothesis, we did not find evidence that group-based values influence the consideration set. Exploratory analyses indicate an interaction pattern between group-based and individual-based values among participants who did not use hints. For the positive group, the data seem to indicate that individuals with higher individual-based values are more likely to be considered, whereas for the negative group, the data seem to indicate that those lower individual-based values are more likely to be considered. An explanation for this interaction may be congruency: Negative group and value 1 are congruent, while positive group and value 4 are congruent. Congruent information may facilitate easier memory retrieval because it is, for instance, experienced as more fluent and, therefore, has a higher value (Gigerenzer & Gaissmaier, 2011; Hertwig et al., 2008). We tested this idea in a confirmatory manner in Experiment 4.3 by removing the hint function.

Experiment 4.3

Experiment 4.3 is a replication of Experiment 4.2. Here, we expect an interaction between group-based and individual-based values on the probability of being considered. In this, we expect a positive linear effect of individual-based values for the positive group and for the negative group we do not expect this pattern. Moreover, we expect that students from the negative group with individual value 1 have a higher probability of being considered than students from the positive group with individual value 1.

Method

Sample Size Justification

The sample size was determined using a simulation-based power analysis with data of Experiment 4.2 (of the 42 participants who did not use any hints). A total of 400 participants allowed us to detect an interaction effect of group-based values and individual-based values as small as *B* =-1.75 (α = .05, power = .80). Payment was identical to Experiment 4.2.

Participants

We recruited 446 participants and after exclusions (see S4.2), the sample size was 400 participants (M_{age} = 39.32, SD_{age} = 11.15, 18–63 years old, 204 women, 187 men, 6 non-binary, 3 skipped this question).

Materials and Procedure

Consideration Set Task.

We employed the same paradigm as in Experiment 4.2 but removed the hint function during the learning stage. Moreover, we added a preference stage at the end of the experiment to explore whether group-based values influence participants' preferences for *new* individuals they did not learn about from those groups. In nine different blue-green choice pairs, participants were instructed to repeatedly and rapidly (within 1500ms) decide which of the two students appeared most positive to them at that moment.

Results

Confirmatory Analyses

Contrary to our hypothesis, there was no significant interaction between group-based and individual-based values on presence in the consideration set, *B* =-0.03, *SE* = 0.04, χ^2(29) = 0.36, *p* = .547, 95% CI [-0.06, 0.10]. There was no significant difference between the positive (*M* = 0.37, *SD* = 0.23) and negative (*M* = 0.38, *SD* = 0.23) group on presence in the consideration set, *B* =-0.02, *SE* = 0.06, χ^2(29) = 0.10, *p* = .748, 95% CI [-0.10, 0.14], but there was again a significant positive linear effect of individual-based values on presence in the consideration set, *B* = 0.38, *SE* = 0.06, χ^2(29) = 15.22, *p* < .001, 95% CI [0.26, 0.50] (see Figure 4.3).

Moreover, contrary to our hypothesis, there was no difference between students with value 1 from the negative group (*M* = 0.35, *SD* = 0.46) and positive (*M* = 0.33, *SD* = 0.44) group in presence in the consideration set, *B* =-0.04, *SE* = 0.19, *p* = .815.

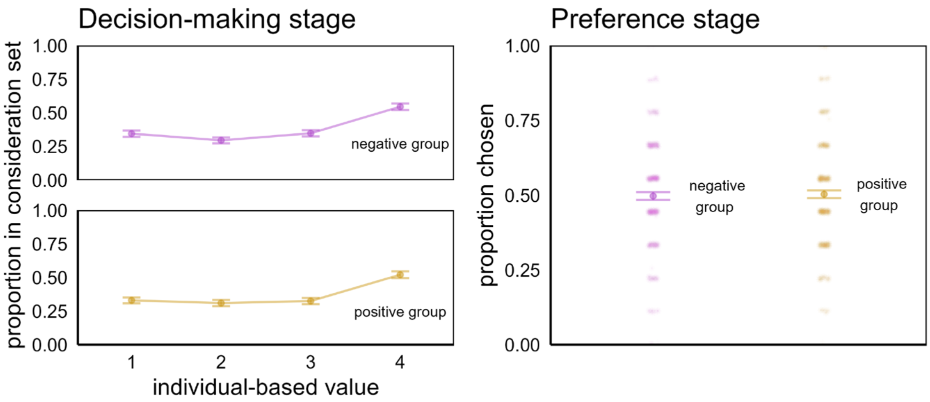
Exploratory Analyses

Preference Stage.

An intercept-only binomial generalized linear mixed model with preference for face from the positive group (1 = positive group, 0 = negative group) as dependent variable and a random intercept varying across participants, showed that the probability of choosing a face from the positive group (*M* = 0.5, *SD* = 0.19) did not significantly differ from chance level (*M* = 0.5), *B* = 0.01, *SE* = 0.04, 95% CI [-0.06, 0.09] (see Figure 4.3). Significance was determined using confidence intervals.

Figure 4.3

Results Experiment 4.3



Note. The mean proportion considered (left panel) and mean proportion chosen (right panel) per within-participant condition is visualized. Error bars reflect within-participant standard error around those means.

Discussion

Contrary to our preregistered hypothesis, we did not find evidence for an interaction between group-based values and individual-based values on the probability of being considered. However, we again found evidence that individual values positively influence the consideration set, showing the robustness of this effect. Thus, we do not find evidence in favor of congruency.

It seems the case that group-based values were not incorporated when constructing consideration sets. This may be because the group membership manipulation might not have been relevant enough (see S4.6 for supporting analyses). We observed that group-based values affect choices in the learning stage at first, but these effects weaken as participants learn the individual-based values. That is, at first, participants choose students from the positive group at a rate higher than chance level. However, as the trials continue and participants learn about students' individual-based values, this preference diminishes over time. This indicates that the group manipulation might not have been relevant enough: Only individual-based values were consequential and, therefore, received more attention than group-based values. This observation is in line with previous work indicating that learning by experience (here, individual-based values) tends to take precedence over learning by description (here, group-based values; Erev et al., 2017; Lejarraga & Gonzalez, 2011). To push the relevance of the group-based values, we changed the learning stage from learning about both individual-based and group-based values to learning about group-based values only.

Experiment 4.4

In Experiment 4.4, we investigate whether people with more positive group-based values have a higher probability of being considered than those with more negative group-based values.

Method

Sample Size Justification

The sample size was determined using G*Power (Faul et al., 2009). A total of 400 participants allowed us to detect a main effect as small as OR = 1.15 ($\alpha = .05$, power = .80). Payment was identical to Experiment 4.1.

Participants

We recruited 531 participants and after exclusions (see S4.2), the sample size was 400 participants (M_{age} = 35.93, SD_{age} = 11.02, 18–64 years old, 209 women, 188 men, 3 skipped this question).

Materials and Procedure

Consideration Set Task.
To investigate the influences of group-based values on the consideration set, we used a similar experiment as Experiment 4.3 but only introduced group-based values. Again, there were four students each in the green and blue groups, and participants were told that one group was more negative and the other more positive (counterbalanced as green or blue; see S4.4 for group descriptions based on Schultner et al., 2024), inducing group-based values.

In the learning stage, participants did not learn about individual-based values, but were asked to categorize which student belonged to which group. Each trial started with presenting one student with a grey background and t-shirt and during presentation, participants could choose whether this student belonged to the green or blue group. After choosing the group, they learned to which group the student belonged by showing them the student with background and shirt in the group color. If they chose the correct group, they received 1 point, and if they chose the incorrect group, they received 0 points. The remainder of the experiment was identical to Experiment 4.3.

Results

Confirmatory Analyses

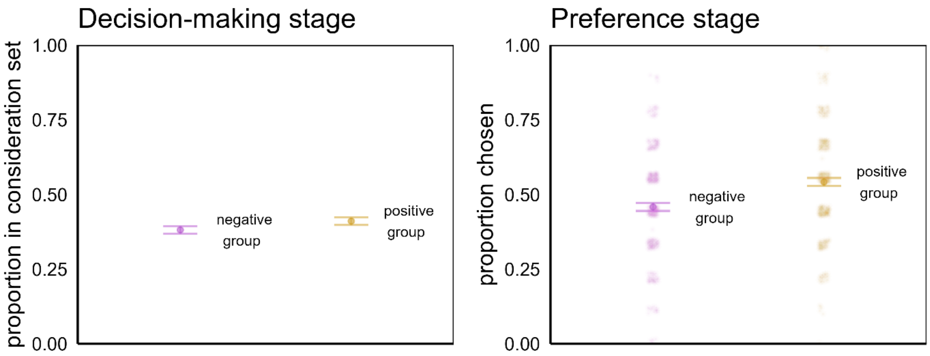
Contrary to our hypothesis, there was no significant difference between the positive (*M* = 0.41, *SD* = 0.25) and negative (*M* = 0.38, *SD* = 0.25) group on presence in the consideration set, *B* = 0.08, *SE* = 0.05, $\chi^2(10) = 2.07$, *p* = .15, 95% CI [-0.18, 0.02], OR = 1.08 (see Figure 4.4).

Exploratory Analyses

Preference Stage.
The same intercept-only binomial generalized linear mixed model as in Experiment 4.3 showed that here the probability of choosing a face from the positive group (*M* = 0.54, *SD* = 0.19) differed significantly from chance (*M* = 0.5), *B* = 0.17, *SE* = 0.04, 95% CI [0.10, 0.25] (see Figure 4.4).

Figure 4.4

Results Experiment 4.4



Note. The mean proportion considered (left panel) and mean proportion chosen (right panel) per within-participant condition is visualized. Error bars reflect within-participant standard error around those means.

Discussion

In Experiment 4.4, contrary to our preregistered hypothesis, we found no evidence that group-based values influence presence in the consideration set, although group-based values did influence preferences. Here, participants preferred *new* individuals from the positive group over individuals from the negative group.

General Discussion

To better understand the roots of biased decision making, the present research focused on identifying *where* biases exert influence within the decision-making process. We specifically focused on one step in the decision-making process: The pre-decision phase. We investigated whether, besides someone's competence, group membership plays a role in whether people are considered. Overall, our findings consistently show that the higher the individual-based value, the higher the probability of an individual being present in the consideration set. Notably, the individual-based value effect occurs even when the value is irrelevant to the decision at hand. Specifically, we did not ask participants who they considered for a value-relevant question (e.g., who do you want to hire?). Instead, we asked a value-irrelevant question (i.e., which name's third letter comes later in the alphabet?), and even then, we observed this positive effect of individual-based values. Contrary to expectations, we consistently find no evidence for an influence of group-based values on the consideration set. Our strongest test for group-based values is in Experiment 4.4, and even there, when we eliminate the possibility of being overshadowed by individual-based values, we find no effects of group-based values on the consideration set.

Interestingly, although group-based values did not influence the consideration set in Experiment 4.4, group-based values did influence whom participants found more positive. Here, participants preferred *new* individuals from the positive group over the negative group, indicating that our group membership manipulation was successful. In line with this, in the pretest of Experiment 4.2, individuals from the positive group were evaluated more positively than those from the negative group. Thus, importantly, even though group-based values did successfully transfer to preferences and evaluations, they did not influence consideration sets.

There are several explanations for the absence of group-based value effects on consideration sets. First, it could be that the effect size of group-based values on consideration sets is smaller than anticipated, resulting in a null-effect. Future research could use a stronger manipulation of group-based values. For example, by implementing a minimal group paradigm (Otten, 2016; Tajfel et al., 1971; van Lent et al., 2024) or by investigating real-world groups (Goette et al., 2012; Mousa, 2020). Second, it could be that our experimental set-up might have stimulated individuation. By introducing only eight students, people may have the cognitive capacity to individuate, which potentially leads to less strong group-based value effects (Rubinstein et al., 2018; van Lent et al., 2025; Wheeler & Fiske, 2005). Future research could attempt to counteract individuation by increasing the number of possible choice options. Thirdly, it could not be value itself influencing the consideration set, but another process. One candidate could be attention, since attention is related to better encoding and retrieval from memory (Chun & Turk-Browne, 2007). In Experiments 4.1–4.3, optimal task performance requires paying attention to and learning highly valued names well. Contrary, in Experiment 4.4, there is no incentive to pay more attention to one group than the other, possibly resulting in the absence of group effects. Future research could investigate alternative processes related to consideration sets.

Implications

This research shows that individual-based values influence consideration sets, even when irrelevant to the decision at hand. This could affect real-life decisions, where you might consider some people over others, despite it not being relevant or appropriate. This raises questions about which types of values drive consideration sets in real-life decisions. For example, in promotion decisions, could it be that certain individuals come to mind more quickly based on attractiveness or warmth, despite these individuals not being the most competent? Future research could investigate the contexts in which values are dysfunctional in influencing consideration sets and may result in biases.

Limitations

One limitation is that we investigated only one phase of the decision-making process, gaining insights into the role of values in the pre-decision phase, but not in actual choices. Moreover, these experiments focused on a teacher-student relationship, making it unknown whether the results generalize to other contexts.

Conclusion

In conclusion, the current research robustly showed that individual-based values positively affect presence in the consideration set, even when the acquired value is irrelevant to the decision at hand. We do not find evidence for an influence of group-based values on presence in the consideration set, although people did successfully transfer group-based values to preferences and evaluations.

Acknowledgements

We thank Jerom Götz, Ismene Klostermann, Katharina Max, Amelie Scherm, Indy Slagmeulen, and Cristian Ungurean for their help with the projects. We thank Hubert Voogd for programming the experiments and Floor Burghoorn for assisting with the simulation-based power analyses. We thank the Focus Theme Sustainability of the Faculty of Social Sciences at Radboud University for supporting this research financially.

Supplemental materials

S4.1: Experiment 4.0

Experiment 4.0 is, in many aspects, similar to Experiment 4.1. However, in Experiment 0, we employed different exclusion criteria than in Experiments 4.1–4.4. That is, we preregistered, in line with Morris and colleagues (2021), to "exclude participants who failed to give a name within the time limit in Stage 2". The time limit used was 17 seconds. We conducted a data quality check after collecting 75 participants and discovered that we had to exclude 54.67% of all participants based on this exclusion criterion. To not waste resources and because we had ethical concerns about excluding so many participants, we decided to stop data collection and go back to the drawing board. After consulting the first author of the paper in which the original paradigm was described (Morris et al., 2021), we learned that there was no theoretical reason to exclude participants who failed to give a name within the time limit in Stage 2. Rather, this exclusion criterion was installed as a quality check. For the next experiments, we did not use this exclusion criterion, and to ensure data quality in a different way, namely, to only allow participants who had a 100% approval rate on Prolific. Experiment 4.0 was preregistered at the Open Science Framework: https://osf.io/fr8dk/?view_only=5793bca9e4cf4fb187cb66ebcdecc5cb.

S4.2: Preregistered Exclusion Criteria

Experiment 4.1

We excluded participants (1) who did not complete the experiment, (2) who chose the better alternative in the learning stage on fewer than 70% of the trials, (3) who failed a decision-making stage comprehension check, (4) or who wrote things down physically during the experiment (measured by self-report at the end).

Experiments 4.2–4.3

We excluded participants (1) who did not complete the experiment, (2) who failed a decision-making stage comprehension check, or (3) who wrote things down physically during the experiment (measured by self-report at the end).

Experiment 4.4

We excluded participants (1) who did not complete the experiment, (2) who had less than 55% of the trials correct in the last 28 trials of the learning stage, (3) who failed a decision-making stage comprehension check, or (4) who used aids during the experiment (such as pencil and paper, measured by self-report at the end).

S4.3: Model Specifics

For Experiment 4.1, this model included the within-participant linear predictor individual value (1–8) and random intercepts of participant, target name, and avatar as well as random slopes for individual value.

For Experiments 4.2–4.3, this model included the within-participant linear predictor individual value (1–4) and the within-participant factor group value (positive vs. negative) and their interaction. Moreover, this model included a random intercept of participant including random slopes for individual value and group value, a random intercept of target name including a random slope for individual value, group value and their interaction, and a random intercept for avatar including a random slope for individual value, group value and their interaction. Additionally, in Experiment 4.3, to determine whether the target from the negative group with individual value 1 has a higher probability of being considered than the target from the positive group with individual value 1, we used the same model but we included individual value as within-participant factor.

For Experiment 4.4, this model included the within-participant factor group value (positive vs. negative) and random intercepts of participant, target name, and avatar as well as random slopes for group value.

S4.4: Group Descriptions

Experiments 4.2–4.3

Members of Group Blue/Green live in a more affluent society, where crime is low, and most people have a high education and good jobs. People from Group Blue/Green are often perceived to be polite, peaceful, and trustworthy, and they are proud of their success.

Group Green/Blue lives in a society that is economically poor, lower educated, with a high rate of unemployment and serious crimes such as shoplifting and drug dealing. People from Group Green/Blue are often perceived to be hostile, untrustworthy, and ignorant.

Experiment 4.4

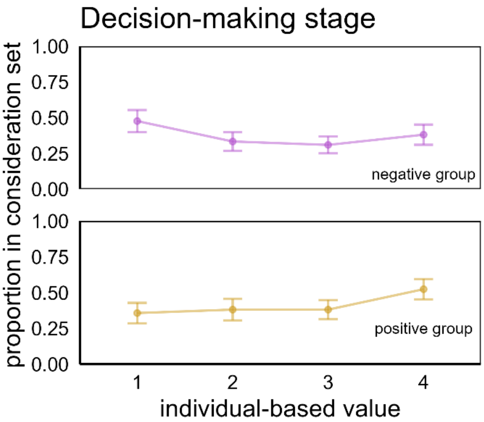
Members of Group Blue/Green live in a more affluent society, where crime is low, and most people have a high education and good jobs. People from Group Blue/Green are often perceived to be polite, peaceful, and trustworthy, and proud of their success.

Group Green/Blue lives in a society that is economically poor, lower educated, with a high rate of unemployment and serious crimes such as shoplifting and drug dealing. People from Group Green/Blue are often perceived to be hostile, untrustworthy, and ignorant.

S4.5: Figure Experiment 4.2 Decision-making Stage Without Hints

Figure 4.1 Supplemental Materials

Results Consideration Set Experiment 4.2 Subset Without Hints



Note. The mean proportion considered per within-participant condition is visualized. Error bars reflect within-participant standard error around those means.

S4.6: Additional Exploratory Analyses

Initial Influences of Group-Based Values on The Learning Stage

To explore whether group-based values initially influence choices during the learning stage, we analyzed a subset of the learning stage data from Experiments 4.2–4.3. This subset only included trial 1 where participants were required to select between a student from the positive group and a student from the negative group. Significance was determined using confidence intervals.

Experiment 4.2.

An intercept-only binomial generalized linear mixed model with choice for face from the positive group (1 = positive group, 0 = negative group) as dependent variable and a random intercept varying across participants, showed that the probability of choosing a face from the positive group (*M* = 0.6, *SD* = 0.49) significantly differed from chance level (*M* = 0.5), *B* = 0.40, *SE* = 0.14, 95% CI [0.14, 0.71].

Experiment 4.3.

The same intercept-only binomial generalized linear mixed model as in Experiment 4.2 showed that the probability of choosing a face from the positive group (*M* = 0.71, *SD* = 0.45) significantly differed from chance level (*M* = 0.5), *B* = 0.91, *SE* = 0.14, 95% CI [0.66, 1.21].

Taken together, this suggests that group-based values initially influenced the choice for a student from the positive group over a student from the negative group.

Influences of Group-Based Values on The Learning Stage over Trials

To explore the influence of group-based values on choices during the learning stage over trials, we analyzed a subset of the learning stage data from Experiments 4.2–4.3. This subset included only choice pairs (32 trials) where participants were required to select between a student from the positive group and a student from the negative group.

Experiment 4.2.

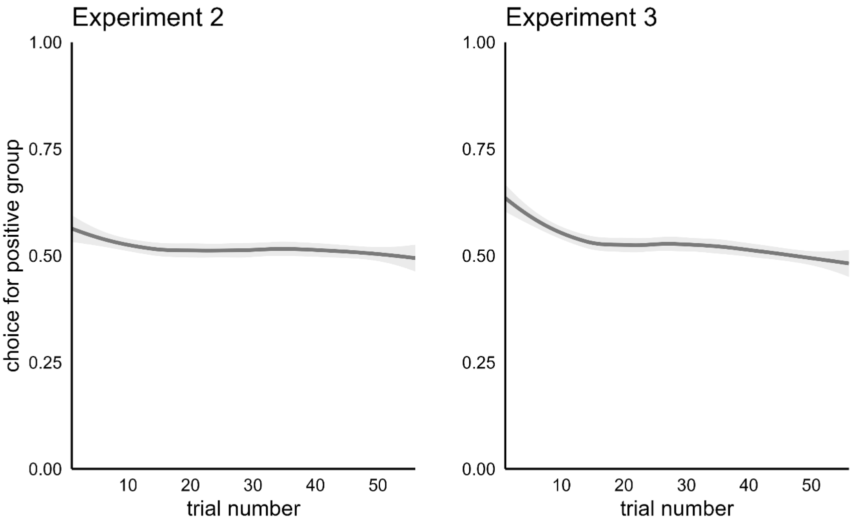
A binomial generalized linear mixed model with choice for a face from the positive group (1 = positive group, 0 = negative group) as dependent variable, trial number as predictor and a random intercept varying across participants, showed that the probability of choosing a face from the positive group significantly decreased as the trial numbers increased, *B* =-0.003, *SE* = 0.001, $\chi^2(2) = 6.19$, *p* = .013, 95% CI [-0.005,-0.0004] (see Figure 4.2 Supplemental Materials).

Experiment 4.3.

The same binomial generalized linear mixed model as in Experiment 4.2 showed that the probability of choosing a face from the positive group significantly decreased as the trial numbers increased, *B* =-0.007, *SE* = 0.001, $\chi^2(2) = 42.91$, *p* < .001, 95% CI [-0.01,-0.005] (see Figure 4.2 Supplemental Materials).

Taken together, this suggests that group-based values initially influenced the choice for a student, but this influence diminishes over time as participants gain more insight into individual-based values.

Figure 4.2 Supplemental Materials



Results Group-Based Value Influences on the Learning Stage

Note. Shaded area reflects 95% confidence level around the smoothing line.

CHAPTER 5

General discussion

The main aim of this dissertation was to gain a better understanding of the instrumental learning processes that contribute to how people evaluate individuals and social categories. This was investigated in four research questions across three empirical chapters, focusing on the consequences of instrumental learning for social evaluations and behavior. Table 5.1 provides an overview of the brief conclusions for each research question. The order of the research questions in this table does not match the order of the empirical chapters, as the research questions are arranged according to their importance. The main aim concerns the consequences of instrumental learning. This dissertation also addresses a question about the instrumental learning process itself, which is covered in Chapter 2 but listed here as research question 4. This general discussion describes the insights gained from each research question, followed by implications, a general reflection on the research's strengths and limitations, ideas for future research, and a conclusion.

Table 5.1

Overview of Brief Conclusions for Each Research Question

Research Questions	Ch. 2	Ch. 3	Ch. 4	Brief Conclusions
1. What is the unique contribution of inactions over and above reward and punishment in shaping evaluations and behavior?	X	X		Inactions matter at the level of individuals, they resulted in less positive evaluations and preferences than actions, even beyond the influence of punishment signals. At the social category level, there was no evidence that punishment-avoidant inactions influence emotion recognition of ingroup individuals.
2. Does instrumental learning generalize to evaluations and behaviors toward others from the same social category?		X	X	Action–reward learning about outgroup individuals seemed to generalize to emotion recognition of new individuals from the same outgroup. However, there was no evidence that inaction punishment-avoidant learning about ingroup individuals generalizes to emotion recognition of new individuals from the same ingroup. Moreover, instrumental learning of social category-based evaluations of nonexistent social categories generalized to preferences for unknown others from the same social category.
3. Does instrumental learning about social category-based and individual-based evaluations influence the construction of consideration sets?			X	More positive individual-based evaluations increased the probability that someone is included in the consideration set. There was no evidence that social category-based evaluations influence inclusion in the consideration set.
4. Does the target's social category influence (in) action learning?	X			There was no evidence that the social category of the target influences (in)action learning.

Insights per Research Question

RQ1: The Unique Contribution of Inactions on Evaluations and Behavior

In Chapters 2 and 3, I investigated the unique contribution of inactions over and above reward and punishment in shaping evaluations and behavior toward individuals and social categories. In line with the preregistered hypotheses, the results of Chapter 2 showed that inactions have a unique contribution to evaluations and preferences for individuals. In general, the results of three experiments (Experiments 2.2–2.4, $N_{total} = 180$) showed that performing inactions leads to less positive evaluations than performing actions. When breaking it down by consequences (reward and punishment avoidance), results showed that when being rewarded for (in)actions, inactions lead to less positive evaluations and preferences than actions. Similarly, when avoiding a punishment with (in)actions, inactions lead to more negative evaluations and preferences than actions. The latter finding is particularly interesting and insightful, as these results cannot be explained by mere co-occurrence of punishment signals with individuals. That is, there was no evidence for a difference in learning between avoiding a punishment with inactions compared to avoiding a punishment with actions. This suggests that participants were exposed to similar individual-punishment relations in both learning conditions. This shows that inactions matter even beyond the influence of punishment signals in shaping evaluations and preferences. Generally, the results of Chapter 2 showed that people evaluated individuals most positively when acting had previously led to the attainment of rewards (i.e., rewarded actions); people evaluated individuals most negatively when *not* acting had previously led to the avoidance of punishment (i.e., punishment-avoidant inactions). Taken together, inactions matter; consequently, *not* acting negatively influences how people feel about others.

In contrast to initial expectations and findings at the individual level, Experiments 3.2–3.3 ($N_{total} = 144$) in Chapter 3 provided no evidence that consequential inactions (i.e., punishment-avoidant inactions) influence emotion recognition of ingroup members at the social category level. This suggests that it is hard to negatively influence evaluations of ingroup members with consequential inactions. Thus, while punishment-avoidant inactions negatively influenced evaluations and preferences at the individual level (Chapter 2), this dissertation showed no evidence for the negative effects of punishment-avoidant inactions at the social category level (Chapter 3). Therefore, it is unclear whether the negative effects of inactions also apply to social categories in general. By examining ingroup individuals at the social category level, it might be that the learned evaluations are very strong and positive (Sherman, 1996; Zajonc, 1968; Zebrowitz et al., 2008), or the learning does not generalize beyond the ingroup individuals one has just learned about. Alternatively, individuals might be highly motivated to maintain a positive view of ingroup individuals (Tajfel & Turner, 1979; see also Tajfel, 1982), or a combination of factors, resulting in the absence of effects for punishment-avoidant inactions at the social category level. Thus, *not* acting toward ingroup members to avoid punishments does not seem to influence how people feel about ingroup members.

In sum, in Chapter 2, I provide evidence that there is a unique contribution of inactions in shaping evaluations and behavior at the individual level. Here, as expected, inactions lead to less positive evaluations than actions, even beyond the influence of punishment signals. However, unexpectedly, in Chapter 3, no evidence was found for the impact of punishment-avoidant inactions at the social category level. This could potentially be attributed to the use of an ingroup as the social category. Future research could investigate the influences of inactions at the social category level by examining a different social category than the ingroup. Taken together, these results imply that *not* engaging in interactions with individuals influences how people feel about others whereas *not* engaging with ingroup members to avoid punishments does not seem to influence how people feel about ingroup members.

RQ2: Generalization of Instrumental Learning

In Chapters 3 and 4, I investigated whether instrumental learning generalizes to evaluations and behavior toward others from the same social category using three outcome measures: Emotion recognition, evaluations, and preferences.

Emotion Recognition*.*

In Chapter 3 (Experiments 3.2–3.3, N_{total} = 144), I demonstrated that learning to act to Moroccan–Dutch faces to obtain rewards and learning to not act to White–Dutch faces to avoid punishments influences emotion recognition at the social category level. First, I conducted a replication (Experiment 3.1, *N* = 40), where I replicated that emotion recognition is influenced by social category membership of the face (Bijlstra et al., 2010; see also Becker et al., 2007; Bijlstra, Kleverwal, et al., 2019; Craig, Koch, et al., 2017; Craig & Lipp, 2017, 2018a; Hugenberg, 2005; Hugenberg & Sczesny, 2006; Lipp et al., 2015). Dutch, German, or Belgian participants were faster at recognizing happiness as happiness than anger as anger for White–Dutch faces (ingroup), while no difference was found for Moroccan–Dutch faces (outgroup). In other words, there was a Happy Face Advantage (HFA) for White–Dutch but not for Moroccan–Dutch faces. Next, in Experiments 3.2–3.3 (N_{total} = 144), after the instrumental learning task, the commonly observed moderation effect of the social category of the face on the HFA (as seen in Experiment 3.1 and previous studies) was no longer present. Instead, there was 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).

Importantly, results of Chapter 3 showed that the effects of the instrumental learning task generalize to Moroccan–Dutch faces in two ways. First, it generalized to different facial expressions. That is, people learned about neutral emotional expressions in the instrumental learning task, and these learning effects translated to different response times for happy and angry emotional expressions. Second, the learning also generalized to unfamiliar faces that were not present in the instrumental learning task. That is, people learned about five faces in the instrumental learning task, and these learning effects translated to new and unfamiliar faces (13 faces) from the same social category when recognizing emotional expressions. Thus, participants did not learn about these faces, but still, it influenced emotion recognition. Note that this only applies to Moroccan–Dutch faces, as I did not find a descriptively different pattern in emotion recognition for White–Dutch faces as a function of instrumental learning, for possible reasons discussed above.

Evaluations.

In Chapter 3, I found mixed results regarding the effects of the instrumental learning task on evaluations. After the instrumental learning task, participants were asked to evaluate the 18 Moroccan–Dutch and 18 White–Dutch faces. While results of Experiment 3.2 (N = 72) showed that evaluations of Moroccan–Dutch faces were more positive than evaluations of White–Dutch faces after participating in the instrumental learning task, results of Experiment 3.3 (N = 72) did not show a difference. Overall, these results suggest that at least Moroccan–Dutch faces were not evaluated more negatively, in contrast to the findings of Traast and colleagues (2025), who demonstrated anti-Moroccan explicit attitudes. Since a control condition or pre-measure of evaluations is missing, the results are not as informative; it is unknown whether the difference in evaluations can be attributed to other factors, such as social desirability.

Preferences.

In Chapter 4 (Experiment 4.4, N = 400), I demonstrated that instrumental learning of social category-based evaluations of nonexistent social categories generalizes to preferences. Here, participants learned about four faces in the instrumental learning task, and the learning effects translated to preferences about new and unfamiliar faces (3 faces) from the same social category. In short, participants preferred new and unfamiliar faces from advantaged social category over new and unfamiliar faces from disadvantaged social category.

In sum, does instrumental learning generalize? Yes, partly. These results imply that, on a broader level, (1) engaging in interactions with positive consequences with outgroup members positively generalizes to evaluations and behavior toward new outgroup members, while (2) there was no evidence that *not* engaging in interactions to avoid negative consequences with ingroup members generalizes to behavior toward new ingroup members. More closely to the data, these results show that learning which members belong to an advantaged social category generalizes to preferences for new members from that same advantaged social category.

RQ3: The Influence of Social Category and Individual Evaluations on Consideration Sets

To gain a deeper understanding of the roots of biased decision making, in Chapter 4 (Experiments 4.1–4.4, N_{total} = 1600), I specifically focused on one step in the decision-making process and aim to identify whether biases influence the construction of the consideration set. Overall, the findings consistently showed that more positive individual-based evaluations increase the probability that someone is included in the consideration set. Notably, this effect occurred even when the individual-based evaluation was irrelevant to the decision at hand (see also Morris et al., 2021). In doing so, I extended previous research by Morris and colleagues (2021) on consideration sets to social decision-making. Contrary to expectations, I consistently found no evidence that social category-based evaluations influence inclusion in the consideration set. In this, I did not find a difference between students from advantaged versus disadvantaged social categories in terms of whether they are considered or not.

There are several potential explanations for the absence of social category-based evaluation effects on consideration sets. First, the effect size of social category-based evaluations on consideration sets could be smaller than anticipated. Second, it could be that the consideration set task stimulated individuation. By introducing eight students, people may be able to individuate, potentially leading to less strong influences of social category-based evaluations (Rubinstein et al., 2018; Wheeler & Fiske, 2005). Finally, it could be that another psychological process is more prominent in influencing the consideration set than evaluations. A candidate process could be attention, as attention is related to better encoding and retrieval from memory (Chun & Turk-Browne, 2007).

In sum, the findings of Chapter 4 robustly show that individual-based evaluations influence presence in the consideration set, even when the evaluation is irrelevant to the decision at hand. Moreover, no evidence was found for an influence of social category-based evaluations on presence in the consideration set. Taken together, these results imply that in real-life decisions, some people may be considered over others, even when this is not relevant or appropriate.

RQ4: The Influence of the Social Category of the Target on Learning

In Chapter 2, I investigated how instrumental learning works when learning about a variety of people. That is, in this chapter, I explore (1) whether the action–valence asymmetry in learning is present when learning about individuals and (2) whether there are differences in the action–valence asymmetry in learning between ingroup (White–Dutch) and outgroup (Moroccan–Dutch) faces. In doing so, I extend research on the action–valence asymmetry in learning to the domain of social learning. First, I replicated the action–valence asymmetry in learning about fractals (Experiment 2.1, N = 60, Guitart-Masip et al., 2012). Next, the findings of three experiments in Chapter 2 (Experiments 2.2–2.4, N_{total} = 180), provided evidence that the action–valence asymmetry in learning transfers to individuals and to both ingroup and outgroup faces to the same degree: For both groups this asymmetry is present, and there are no differences in learning between ingroup and outgroup faces.

I expected influences of social category-based information on instrumental learning. For example, recent research shows that stereotypical information or ethnicity influences people's preferences toward individual group members in an instrumental learning task (Schultner et al., 2024; Traast et al., 2024). Consistently, one could anticipate that it is more difficult to learn the most optimal action for Moroccan–Dutch faces due to the 'own-race identification bias' or outgroup homogeneity, in which people find it more difficult to distinguish individual faces within outgroups than ingroups (Devine & Malpass, 1985;

Kawakami et al., 2014; Platz & Hosch, 1988). Moreover, one could anticipate more initial actions (go responses) toward White–Dutch faces due to ingroup favoritism (in general, people are more positive about ingroup than outgroup members, Tajfel, 1969; Tajfel et al., 1971). Positivity may be reward-predictive, and therefore, due to the Pavlovian bias, this may prepare actions. Yet, the findings do not show any signs of sensitivity to group membership in learning actions or inactions. One potential explanation for this null result is that the instrumental learning task stimulated individuation. By coupling each individual to one specific instrumental learning condition, people learned specific outcome–action contingencies about one specific individual. Therefore, generalization to social category was less valuable–i.e., it might lead to lower task performance and consequently less monetary bonus. Similarly, previous research has shown that individuation potentially reduces the activation of group stereotypes or social category-based information (Rubinstein et al., 2018; Wheeler & Fiske, 2005), which in turn may decrease the probability of finding effects of social category on learning.

In sum, the findings of Chapter 2 show no evidence that the social category of the target influences (in)action learning. This could be due to individuation. To further explore potential influences of social category on (in)action learning, future research could adapt the paradigm to the social category level, aiming to overcome individuation. For example, by introducing multiple faces from the same social category in each instrumental learning condition.

Implications

Understanding Contact as a Form of Instrumental Learning

This dissertation may help to understand how positive and negative consequences of social interactions, or the absence of such interactions, influence prejudice and discrimination. Therefore, on a more distal level, this dissertation may have implications for the effectiveness of intergroup contact. First, on a positive note, this research is in line with a large body of research suggesting that contact with positive consequences is important for better intergroup relations (Allport, 1954; Paluck et al., 2019; Paolini et al., 2024; Pettigrew & Tropp, 2006). For example, a landmark contact study found that intergroup contact on a sports team (between members of different castes in India) generally had positive effects on prejudice-related outcomes (Lowe, 2021). From the perspective of this dissertation, why would contact here lead to positive effects on prejudice? Contact can be interpreted as rewarded actions, i.e., actions with positive consequences. As demonstrated in this dissertation, the combination of action and reward had the most positive effects on individual evaluations and preferences, and also suggested positive effects on evaluations and behavior at the social category level. Thus, not only do rewarded actions result in positive effects at the individual level, but they also seem to positively generalize to the social category level. This generalization is critical for intergroup contact to be effective in reducing prejudice and discrimination. Moreover, in theory (Boureau & Dayan, 2011; Ereira et al., 2021; Guitart-Masip, Duzel, et al., 2014), contact with positive consequences should function as a flywheel, driving continuous progress. That is, positive contact with someone may cause more positive evaluations. This, in turn, may lead to the prediction of future rewards when encountering that person again. Because of the Pavlovian bias, this reward-predictive 'stimulus' may prepare action, resulting in future contact. Indeed, various lines of research show that future interactions with someone are more likely if previous interactions suggest that future interactions will be rewarding (Hackel, Kogon, et al., 2022; Lott & Lott, 1974; Montoya & Horton, 2004; Newcomb, 1953; Sunnafrank, 1986; Taylor et al., 1969). As a function of generalization, this flywheel may also work for individuals you have not encountered before. Taken together, these findings provide indications of how contact leads to less prejudice, namely, when actions yield rewarding consequences.

Second, this research provides insights into why (1) negative contact and (2) the subsequent absence of contact is particularly harmful at the individual level. In the inaction punishment-avoidant learning condition, participants are asked to learn inactions to avoid a punishment. Learning inactions also imply that participants first perform actions and subsequently receive punishment for those actions, which is inherent to trial-and-error learning. Therefore, in this learning process, two forces are at play that potentially influence evaluations: (1) punished actions (where people experience the punishment), and (2) subsequent inactions to avoid the punishment. How do both of these forces relate to contact?

First, although many field studies on contact focus on positive contact, recent work suggests that contact with negative consequences has detrimental effects for prejudice (Allidina & Cunningham, 2021; Hayward et al., 2017; Paolini et al., 2024). From the perspective of this dissertation, why would negative contact lead to negative effects on prejudice? Negative contact can be interpreted as punished actions. As demonstrated in this dissertation, when initial actions were punished (as was the case in the punishment-avoidant inaction condition), this caused the least positive evaluations and preference (even more so when initial inactions were punished; in the punishment-avoidant action condition) for both Moroccan–Dutch and White–Dutch individuals. Thus, I speculate that there is a strong, distinct aversion to being punished after actions: Similar punishments may have a bigger influence when it follows an action (e.g., greeting a member from a prejudiced social category) than inaction (e.g., ignoring a member from a prejudiced social category). Taken together, these findings provide an indication of how contact leads to more prejudice, namely when actions have negative consequences.[20]

Second, contact literature typically focuses on the presence of social interactions. Consequently, little is known about the effects of the *absence* of social interactions on prejudice. From the perspective of this dissertation, why would the absence of contact lead to negative effects on prejudice? Absence of contact can be interpreted as inactions. Inactions led to less positive evaluations and preferences than actions, even beyond the influence of punishment signals (less positive evaluations in the punishment-avoidant inaction than punishment-avoidant action condition). To the best of my knowledge, there

[20] Please note that learning about ingroup members may differ from learning about outgroup members, as Chapter 3 shows that punishing actions for ingroup faces does not seem to change social category evaluations.

is no field research on the effects of the absence of intergroup contact on prejudice. Related research on inactions within the prejudice domain, however, shows that refraining from speaking up against prejudice amplifies negative intergroup attitudes (Szekeres et al., 2023). This finding gives indications for the detrimental effects of inactions on prejudice: Not acting led to more prejudice. A potential avenue for future research could be to investigate the consequences of inactions in the field and determine whether the negative effects of not having contact also occur in real-world settings.

In theory (e.g., Guitart-Masip, Duzel, et al., 2014), similar to positive contact, contact with negative consequences, along with the absence of contact should also function as a flywheel, hindering progress. Negative contact with someone may cause more negative evaluations. Moreover, mere inactions may also be responsible for more negative evaluations. Both, in turn, may lead to the prediction of future punishments when encountering that person again. Due to the Pavlovian bias, this punishment-predictive ‘stimulus’ may prepare inaction, resulting in *no* future contact.

Implications for Behavior Change in General

Although somewhat distant from the data, the findings of the current dissertation have implications for behavior change in general. Behavior change practitioners are constantly engaging in learning or unlearning others’ behaviors. The action–valence asymmetry in learning (Chapter 2; see also Guitart-Masip et al., 2012) provides a useful framework on how to approach this more successfully. When the goal is to promote a new action, reinforcing the behavior through positive outcomes (i.e., action–reward condition) is the most effective. For example, consider a workplace diversity initiative that encourages employees to actively engage with colleagues they do not yet know. In this context, individuals can be rewarded with enthusiasm or appreciation when they take the initiative to talk to a stranger. Conversely, when the aim is to suppress undesired behavior, applying negative consequences following the action (i.e., punishment-avoidant inaction condition, in which actions are punished) is the most effective. For example, an employee who makes a disrespectful comment may receive a formal reprimand. Consequently, this employee will stop making disrespectful comments (i.e., an inaction) to avoid the punishment (i.e., formal reprimand). Instead of experiencing punishment directly, the anticipation of punishment or observing someone else being punished might also be sufficient to suppress undesired behavior. Thus, the action–valence asymmetry in learning demonstrates that it is most effective to teach desired actions through rewards, and to discourage undesired actions by leveraging the avoidance of punishment—that is, by creating situations where performing the undesired behavior leads to punishment, thereby reinforcing inaction.

Strengths, Limitations and Future Research

This dissertation has several strengths. All chapters include a replication, all experiments are preregistered, and all new fundamental insights in this dissertation are based on multiple experiments and, at times, multiple measures for the same psychological construct. This together contributes to solid, rigorous, and trustworthy science.

The main limitation is that, although the research has high internal validity due to controlled lab and online experiments, these experiments are removed from real-life social interactions or contact. This limits translation of the findings to everyday social interactions—thereby limiting ecological validity. Given that we already observe substantial effects on evaluations in these experimental tasks in the laboratory, what might the effect of real-world contact be, where the consequences are greater? And how could one bring the processes studied in these experimental tasks closer to real-world contact research? I hope that the experimental findings on instrumental learning will inspire practice-oriented researchers when investigating real-world social interactions or designing interventions to combat prejudice. For example, this can be done on a personal level by rewarding actions, and/or on a systematic level by creating environments in which actions are rewarded (see Chater & Loewenstein, 2023, for an exposition on personal versus systemic behavior change strategies).

Another limitation is that, regarding the influences of instrumental learning on inclusion in the consideration set, by focusing solely on consequential actions, I only investigated half of the story. That is, I did not investigate the influences of consequential inactions on consideration sets. In doing so, there is no complete picture of the impact of actions, inactions, and their consequences on inclusion in the consideration set. Since Chapter 2 clearly indicates the unique negative contribution of inactions for individual evaluations and preferences, future research could investigate the influence of consequential inactions on inclusion in consideration sets. If someone refrains from acting toward an individual, what would that do with this individual's probability of being considered? If indeed inactions negatively shape evaluations, one could expect that this individual has a lower probability of being included in the consideration set.

Results of this dissertation show that the effects of instrumental learning generalize to emotion recognition for Moroccan–Dutch faces (Chapter 3), as well as to preferences for new members from the same nonexistent social category (Chapter 4). However, the boundary conditions regarding when and why instrumental learning generalizes to the social category level remain unclear. For example, what is the minimum number of people with whom a person needs to experience reward through their actions for generalization? What happens when information about the social category is slightly mixed? Do effects generalize to different sub-categories from the same social category, such as different genders? Future research could investigate the boundary conditions when instrumental learning transfers to evaluations or behavior toward new members from the same social category.

Finally, results of this dissertation clearly indicate that inactions influence individual evaluations and preferences. This is in line with earlier work showing that mere inactions, without any external consequences, are sufficient to influence preferences for choice options (e.g., food options; Chen et al., 2019; Veling et al., 2022). However, this is not consistent with other research on inactions, such as the 'feature-positive' effect (Fazio et al., 1982), in which actions have a stronger effect on attitudes than inactions. Here, Fazio and colleagues (1982) show that making an action to classify a neutral cartoon as funny or unfunny has a greater impact on attitudes than not making an action to classify a neutral cartoon as funny or unfunny. This raises the question of *when* inactions are diagnostic in shaping evaluations? Perhaps the relevance or meaningfulness of the stimulus toward which someone does not act (e.g., food and faces versus cartoons) influences whether this inaction is diagnostic. Future research could further investigate the boundary conditions of the influence of inactions and the psychological mechanisms underlying this influence.

Conclusion

What did we learn about the instrumental learning processes that contribute to how people evaluate individuals and social categories? In an extensive set of studies, my dissertation highlights the negative effects of inactions. Inactions negatively shape individual evaluations and preferences, even beyond the influence of punishments. At the social category level, there is no evidence for this negative effect of inactions. This could be because I only investigated an ingroup as the social category. Moreover, my dissertation highlights the importance of actions with positive consequences. These actions positively influence individual evaluations, preferences, and probabilities of the individual being included in the consideration set. Additionally, my findings suggest that these actions positively influence evaluations, preferences, and emotion recognition at the level of the social category. Finally, regarding instrumental learning itself, I do not find evidence that the social category of the target influences the learning. Individuation is discussed as one of the potential explanations for the absence of this evidence. To conclude, I hope this dissertation inspires future work by both researchers and practitioners who aim to understand and reduce prejudice and discrimination through instrumental learning processes.

CHAPTER 6

Miscellaneous

References

Research data management, transparency, and privacy

English summary

Dutch summary | Nederlandse samenvatting

Acknowledgements | Dankwoord

About the author

List of academic publications

References

Aberson, C. L., & Gaffney, A. M. (2009). An integrated threat model of explicit and implicit attitudes. *European Journal of Social Psychology*, *39*(5), 808–830. https://doi.org/10.1002/ejsp.582

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. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670–682. https://doi.org/10.1038/nrn3800

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. (2025). A learning and memory account of impression formation and updating. *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-025-00445-x

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

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562. https://doi.org/10.1177/0956797617723724

Andriessen, I., Hoegen Dijkhof, J., van der Torre, A., van den Berg, E., Pulles, I., Iedema, J., & de Voogd-Hamelink, M. (2020). *Ervaren discriminatie in Nederland II.* Sociaal en Cultureel Planbureau. https://www.scp.nl/publicaties/publicaties/2020/04/02/ervaren-discriminatie-in-erland-ii

Auspurg, K., Schneck, A., & Hinz, T. (2019). Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing markets. *Journal of Ethnic and Migration Studies*, *45*(1), 95–114. https://doi.org/10.1080/1369183X.2018.1489223

Bamford, L. E., Klassen, N. R., & Karl, J. M. (2020). Faster recognition of graspable targets defined by orientation in a visual search task. *Experimental Brain Research*, *238*(4), 905–916. https://doi.org/10.1007/s00221-020-05769-z

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

Berkman, E. T. (2018). Value-based choice: An integrative, neuroscience-informed model of health goals. *Psychology & Health*, *33*(1), 40–57. https://doi.org/10.1080/08870446.2017.1316847

Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-control as value-based choice. *Current Directions in Psychological Science*, *26*(5), 422–428. https://doi.org/10.1177/0963721417704394

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. (2019). 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. (2019). 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

Boureau, Y. L., & Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, *36*(1), 74–97. org/10.1038/npp.2010.151

Bovenkerk, F., & Fokkema, T. (2016). Crime among young Moroccan men in the Netherlands: Does their regional origin matter? *European Journal of Criminology*, *13*(3), 352–371. https://doi.org/10.1177/1477370815623566

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*(11), 681–684. https://doi.org/10.1037/h0040090

Brown, P., & van Eijk, N. (2021). Cultural processes shaping stop-and-check practices and interaction dynamics in a large Dutch city: Police vulnerabilities, thought styles and rituals. *The British Journal of Criminology*, *61*(3), 690–709. https://doi.org/10.1093/bjc/azaa083

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401–423. https://doi.org/10.1037/0033-2909.115.3.401

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*(1), 5–17. https://doi.org/10.1037/0022-3514.65.1.5

Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., & Frank, M. J. (2013). Frontal theta overrides Pavlovian learning biases. *Journal of Neuroscience*, *33*(19), 8541–8548. https://doi.org/10.1523/JNEUROSCI.5754-12.2013

Centraal Bureau voor de Statistiek. (2024). *3.1 Ervaren discriminatie.* https://www.cbs.nl/nl-nl/longread/rapportages/2024/ervaren-discriminatie-in-nederland/3-1-ervaren-discriminatie

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

Charlesworth, T. E. S., & Banaji, M. R. (2022). Patterns of implicit and explicit attitudes: IV. Change and stability from 2007 to 2020. *Psychological Science*, *33*(9), 1347–1371. https://doi.org/10.1177/09567976221084257

Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, *46*, e147. https://doi.org/10.1017/S0140525X22002023

Chen, Z., Holland, R. W., Quandt, J., Dijksterhuis, A., & Veling, H. (2019). When mere action versus inaction leads to robust preference change. *Journal of Personality and Social Psychology*, *117*(4), 721–740. https://doi.org/10.1037/pspa0000158

Chotpitayasunondh, V., & Douglas, K. M. (2018). The effects of "phubbing" on social interaction. *Journal of Applied Social Psychology*, *48*(6), 304–316. https://doi.org/10.1111/jasp.12506

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dolan, R. J., & Düzel, E. (2013). Structural integrity of the substantia nigra and subthalamic nucleus predicts flexibility of instrumental learning in older-age individuals. *Neurobiology of Aging*, *34*(10), 2261–2270. https://doi.org/10.1016/j.neurobiolaging.2013.03.030

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005

Collins, A. G. E., & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review*, *121*(3), 337–366. https://doi.org/10.1037/a0037015

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. (2017). 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. (2018a). 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., & Lipp, O. V. (2018b). The influence of multiple social categories on emotion perception. *Journal of Experimental Social Psychology*, *75*, 27–35. https://doi.org/10.1016/j.jesp.2017.11.002

Craig, B. M., Zhang, J., & Lipp, O. V. (2017). 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

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(2), 230–241. https://doi.org/10.1017/S1138741600006491

Del Valle, S. Y., Hyman, J. M., Hethcote, H. W., & Eubank, S. G. (2007). Mixing patterns between age groups in social networks. *Social Networks*, *29*(4), 539–554. https://doi.org/10.1016/j.socnet.2007.04.005

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*(4), 951–978. https://doi.org/10.1037/0033-295X.112.4.951

Devine, P. G., & Malpass, R. S. (1985). Orienting strategies in differential face recognition. *Personality and Social Psychology Bulletin*, *11*(1), 33–40. https://doi.org/10.1177/0146167285111003

Di Lemma, L. C. G., & Field, M. (2017). Cue avoidance training and inhibitory control training for the reduction of alcohol consumption: A comparison of effectiveness and investigation of their mechanisms of action. *Psychopharmacology*, *234*(16), 2489–2498. https://doi.org/10.1007/s00213-017-4639-0

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. https://doi.org/10.1037/h0033731

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*(4), 315–319. https://doi.org/10.1111/1467-9280.00262

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

Dowle, M., & Srinivasan, A. (2023). *data.table: Extension of `data.frame’* (Version 1.14.8) [R package]. https://CRAN.R-project.org/package=data.table

Driscoll, R. L., Clancy, E. M., & Fenske, M. J. (2021). Motor-response execution versus inhibition alters social-emotional evaluations of specific individuals. *Acta Psychologica*, *215*, 103290. https://doi.org/10.1016/j.actpsy.2021.103290

Dubinsky, J. M., & Hamid, A. A. (2024). The neuroscience of active learning and direct instruction. *Neuroscience & Biobehavioral Reviews*, *163*, 105737. https://doi.org/10.1016/j.neubiorev.2024.105737

Eberly, H. W., Carbine, K. A., LeCheminant, J. D., & Larson, M. J. (2022). Testing the relationship between inhibitory control and soda consumption: An event-related potential (ERP) study. *Appetite*, *173*, 105994. https://doi.org/10.1016/j.appet.2022.105994

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

Elliot, A. J. (2008). Approach and avoidance motivation. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 3–14). Psychology Press.

English, J. (2024). The "content" of intergroup contact: Lessons from the Denton Women's Interracial Fellowship. *Politics, Groups, and Identities*, *13*(2), 368–386. https://doi.org/10.1080/21565503.2024.2336979

Ereira, S., Pujol, M., Guitart-Masip, M., Dolan, R. J., & Kurth-Nelson, Z. (2021). Overcoming Pavlovian bias in semantic space. *Scientific Reports*, *11*(1), 3416. https://doi.org/10.1038/s41598-021-82889-8

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review*, *124*(4), 369–409. https://doi.org/10.1037/rev0000062

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). Academic Press. https://doi.org/10.1016/S0065-2601(08)60318-4

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, *87*(3), 293–311. https://doi.org/10.1037/0022-3514.87.3.293

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Fazio, R. H., Sherman, S. J., & Herr, P. M. (1982). The feature-positive effect in the self-perception process: Does not doing matter as much as doing? *Journal of Personality and Social Psychology*, *42*(3), 404–411. https://doi.org/10.1037/0022-3514.42.3.404

Fenske, M. J., Raymond, J. E., Kessler, K., Westoby, N., & Tipper, S. P. (2005). Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychological Science*, *16*(10), 753–758. https://doi.org/10.1111/j.1467-9280.2005.01609.x

Fiske, S. T., Lin, M., & Neuberg, S. L. (2018). The continuum model: Ten years later. In S. T. Fiske, *Social cognition* (1st ed., pp. 41–75). Routledge. https://doi.org/10.4324/9781315187280

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). Academic Press. https://doi.org/10.1016/S0065-2601(08)60317-2

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage Publications.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51–72. https://doi.org/10.1162/0898929052880093

Galvan, M. J., Alvarez, G. M., Cipolli, W., Cooley, E., Muscatell, K. A., & Payne, B. K. (2024). Is discrimination widespread or concentrated? Evaluating the distribution of anti-black discrimination in judicial, hiring, and housing decisions. *Personality and Social Psychology Bulletin*, *0*(0). https://doi.org/10.1177/01461672241288929

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Glickman, S. E., & Schiff, B. B. (1967). A biological theory of reinforcement. *Psychological Review*, *74*(2), 81–109.

Gluth, S., Rieskamp, J., & Büchel, C. (2012). Deciding when to decide: Time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *Journal of Neuroscience*, *32*(31), 10686–10698. https://doi.org/10.1523/JNEUROSCI.0727-12.2012

Goette, L., Huffman, D., & Meier, S. (2012). The impact of social ties on group interactions: Evidence from minimal groups and randomly assigned real groups. *American Economic Journal: Microeconomics*, *4*(1), 101–115. https://doi.org/10.1257/mic.4.1.101

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Grady, K. E. (1981). Sex bias in research design. *Psychology of Women Quarterly*, *5*(4), 628–636. https://doi.org/10.1111/j.1471-6402.1981.tb00601.x

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

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*(1), 419–445. https://doi.org/10.1146/annurev-psych-010419-050837

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition*, *24*(2), 187–206. https://doi.org/10.1521/soco.2006.24.2.187

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., Economides, M., Huys, Q. J. M., Frank, M. J., Chowdhury, R., Duzel, E., Dayan, P., & Dolan, R. J. (2014). Differential, but not opponent, effects of L-DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology*, *231*(5), 955–966. https://doi.org/10.1007/s00213-013-3313-4

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., Berg, J. J., Lindström, B. R., & Amodio, D. M. (2019). Model-based and model-free social cognition: Investigating the role of habit in social attitude formation and choice. *Frontiers in Psychology*, *10*, 2592. https://doi.org/10.3389/fpsyg.2019.02592

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., 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

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

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—The R Package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–32. https://doi.org/10.18637/jss.v059.i09

Harrell, F. E. (2023). *Hmisc: Harrell miscellaneous* (Version 4.6-0) [R package]. https://CRAN.R-project.org/package=Hmisc

Hascher, J., Desai, N., & Krajbich, I. (2021). Incentivized and non-incentivized liking ratings outperform willingness-to-pay in predicting choice. *Judgment and Decision Making*, *16*(6), 1464–1484. https://doi.org/10.1017/S1930297500008500

Hayward, L. E., Tropp, L. R., Hornsey, M. J., & Barlow, F. K. (2017). Toward a comprehensive understanding of intergroup contact: Descriptions and mediators of positive and negative contact among majority and minority groups. *Personality and Social Psychology Bulletin*, *43*(3), 347–364. https://doi.org/10.1177/0146167216685291

Hehman, E., & Neel, R. (2024). Prejudice model 1.0: A predictive model of prejudice. *Psychological Review*, *131*(5), 1235–1265. https://doi.org/10.1037/rev0000470

Hershberger, W. A. (1986). An approach through the looking-glass. *Animal Learning & Behavior*, *14*(4), 443–451. https://doi.org/10.3758/BF03200092

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1191–1206. https://doi.org/10.1037/a0013025

Heyes, C. M. (1994). Social learning in animals: Categories and mechanisms. *Biological Reviews*, *69*(2), 207–231.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

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

Howard, J. A., & Sheth, J. N. (1969). *The theory of buyer behavior*. Wiley.

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

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Computational Biology*, *7*(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028 *Inquisit 6*. (2022). [Computer software]. Retrieved from https://www.millisecond.com.

Insko, C. A., & Wilson, M. (1977). Interpersonal attraction as a function of social interaction. *Journal of Personality and Social Psychology*, *35*(12), 903–911. https://doi.org/10.1037/0022-3514.35.12.903

Johannes, N., Buijzen, M., & Veling, H. (2021). Beyond inhibitory control training: Inactions and actions influence smartphone app use through changes in explicit liking. *Journal of Experimental Psychology: General*, *150*(3), 431–445. https://doi.org/10.1037/xge0000888

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., Voss, H. U., Ballon, D. J., & Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *Journal of Neuroscience*, *31*(37), 13039–13045. https://doi.org/10.1523/JNEUROSCI.2972-11.2011

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

Katzman, P. L., & Hartley, C. A. (2020). The value of choice facilitates subsequent memory across development. *Cognition*, *199*, 104239. https://doi.org/10.1016/j.cognition.2020.104239

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

Kawakami, K., Steele, J. R., Cifa, C., Phills, C. E., & Dovidio, J. F. (2008). Approaching math increases math=me and math=pleasant. *Journal of Experimental Social Psychology*, *44*(3), 818–825. https://doi.org/10.1016/j.jesp.2007.07.009

Kawakami, K., Williams, A., Sidhu, D., Choma, B. L., Rodriguez-Bailón, R., Cañadas, E., Chung, D., & Hugenberg, K. (2014). An eye for the I: Preferential attention to the eyes of ingroup members. *Journal of Personality and Social Psychology*, *107*(1), 1–20. https://doi.org/10.1037/a0036838

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the nature of automatically triggered approach–avoidance behavior. *Emotion Review*, *5*(3), 280–284. https://doi.org/10.1177/1754073913477501

Krieglmeyer, R., Deutsch, R., De Houwer, J., & De Raedt, R. (2010). Being moved: Valence activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science*, *21*(4), 607–613. https://doi.org/10.1177/0956797610365131

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

Lancee, B. (2019). Ethnic discrimination in hiring: Comparing groups across contexts. Results from a cross-national field experiment. *Journal of Ethnic and Migration Studies,* *47*(6), 1181–1200. https://doi.org/10.1080/1369183X.2019.1622744

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

Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes*, *116*(2), 286–295. https://doi.org/10.1016/j.obhdp.2011.05.001

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168. https://doi.org/10.1002/ejsp.2079

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

Linn, S., Lawley, S. D., Karamched, B. R., Kilpatrick, Z. P., & Josić, K. (2024). Fast decisions reflect biases; slow decisions do not. *Physical Review E*, *110*(2), 024305. https://doi.org/10.1103/PhysRevE.110.024305

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*, 107950. https://doi.org/10.1016/j.appet.2025.107950

Lott, A. J., & Lott, B. E. (1974). The role of reward in the formation of positive interpersonal attitudes. In T. L. Huston (Ed.), *Foundations of interpersonal attraction* (pp. 171–192). Academic Press. https://doi.org/10.1016/B978-0-12-362950-0.50014-8

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, *111*(6), 1807–1844. https://doi.org/10.1257/aer.20191780

Lowe, M. (2025). Has intergroup contact delivered? *Annual Review of Economics, 17.* https://doi.org/10.1146/annurev-economics-081324-091109

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

Lucas, B. J., Berry, Z., Giurge, L. M., & Chugh, D. (2021). A longer shortlist increases the consideration of female candidates in male-dominant domains. *Nature Human Behaviour*, *5*, 736–742. https://doi.org/10.1038/s41562-020-01033-0

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Markman, A. B., & Brendl, C. M. (2005). Constraining theories of embodied cognition. *Psychological Science*, *16*(1), 6–10. https://doi.org/10.1111/j.0956-7976.2005.00772.x

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

Miller, N. E. (1959). Liberalization of basic S-R concepts: Extensions to conflict behavior and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* (Vol. 2, pp. 196–292). McGraw-Hill.

Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P., & Robinson, O. J. (2017). Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biological Psychiatry*, *82*(7), 532–539. https://doi.org/10.1016/j.biopsych.2017.01.017

Montoya, R. M., & Horton, R. S. (2004). On the importance of cognitive evaluation as a determinant of interpersonal attraction. *Journal of Personality and Social Psychology*, *86*(5), 696–712. https://doi.org/10.1037/0022-3514.86.5.696

Moran, T., Nudler, Y., & Bar-Anan, Y. (2023). Evaluative conditioning: Past, present, and future. *Annual Review of Psychology*, *74*(1), 245–269. https://doi.org/10.1146/annurev-psych-032420-031815

Morris, A., Phillips, J., Huang, K., & Cushman, F. (2021). Generating options and choosing between them depend on distinct forms of value representation. *Psychological Science*, *32*(11), 1731–1746. https://doi.org/10.1177/09567976211005702

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science*, *369*(6505), 866–870. https://doi.org/10.1126/science.abb3153

Moutoussis, M., Bullmore, E. T., Goodyer, I. M., Fonagy, P., Jones, P. B., Dolan, R. J., & Dayan, P. (2018). Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. *PLoS Computational Biology*, *14*(12), e1006679. https://doi.org/10.1371/journal.pcbi.1006679

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

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review*, *60*(6), 393–404. https://doi.org/10.1037/h0063098

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of biases. *Perspectives on Psychological Science*, *18*(6), 1464–1487. https://doi.org/10.1177/17456916221148147

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

OpenAI. (2023). *ChatGPT* [Computer software]. https://chat.openai.com/chat

Otten, S. (2016). The Minimal Group Paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology*, *11*, 85–89. https://doi.org/10.1016/j.copsyc.2016.06.010

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, *3*(2), 129–158. https://doi.org/10.1017/bpp.2018.25

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

Paolini, S., & McIntyre, K. (2019). Bad is stronger than good for stigmatized, but not admired outgroups: Meta-analytical tests of intergroup valence asymmetry in individual-to-group generalization experiments. *Personality and Social Psychology Review*, *23*(1), 3–47. https://doi.org/10.1177/1088868317753504

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

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and blacks with approach behaviors. *Journal of Personality and Social Psychology*, *100*(2), 197–210. https://doi.org/10.1037/a0022159

Plant, E. A. (2004). Responses to interracial interactions over time. *Personality and Social Psychology Bulletin*, *30*(11), 1458–1471. https://doi.org/10.1177/0146167204264244

Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin*, *29*(6), 790–801. https://doi.org/10.1177/0146167203029006011

Platz, S. J., & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology*, *18*(11), 972–984. https://doi.org/10.1111/j.1559-1816.1988.tb01187.x

Posavac, S. S., Sanbonmatsu, D. M., & Fazio, R. H. (1997). Considering the best choice: Effects of the salience and accessibility of alternatives on attitude–decision consistency. *Journal of Personality and Social Psychology*, *72*(2), 253–261. https://doi.org/10.1037/0022-3514.72.2.253

Quillian, L., & Lee, J. J. (2023). Trends in racial and ethnic discrimination in hiring in six Western countries. *Proceedings of the National Academy of Sciences*, *120*(6), e2212875120. https://doi.org/10.1073/pnas.2212875120

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

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*(6), 897–911. https://doi.org/10.1037/a0036498

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology*, *101*(3), 557–570. https://doi.org/10.1037/a0022885

Revelle, W. (2023). *psych: Procedures for personality and psychological research* (Version 2.1.9) [R package]. https://CRAN.R-project.org/package=psych

Richter, A., Guitart-Masip, M., Barman, A., Libeau, C., Behnisch, G., Czerney, S., Schanze, D., Assmann, A., Klein, M., Düzel, E., Zenker, M., Seidenbecher, C. I., & Schott, B. H. (2014). Valenced action/inhibition learning in humans is modulated by a genetic variant linked to dopamine D2 receptor expression. *Frontiers in Systems Neuroscience*, *8*, 140. https://doi.org/10.3389/fnsys.2014.00140

Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, *75*, 54–70. https://doi.org/10.1016/j.jesp.2017.11.009

Schmitz, L., Bijleveld, E., & Veling, H. (2020). Cognitive labor shapes the desire for social and monetary compensation. *Motivation and Emotion*, *44*(6), 797–809. https://doi.org/10.1007/s11031-020-09856-0

Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., Nagar, L., Perez, J., & Poldrack, R. A. (2014). Changing value through cued approach: An automatic mechanism of behavior change. *Nature Neuroscience*, *17*(4), 625–630. https://doi.org/10.1038/nn.3673

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 prejudice through instrumental learning. *Proceedings of the National Academy of Sciences*, *121*(45), e2414518121. https://doi.org/10.1073/pnas.2414518121

Seibt, B., Neumann, R., Nussinson, R., & Strack, F. (2008). Movement direction or change in distance? Self- and object-related approach–avoidance motions. *Journal of Experimental Social Psychology*, *44*(3), 713–720. https://doi.org/10.1016/j.jesp.2007.04.013

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology*, *70*(6), 1126–1141.

Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis. *Psychological Science*, *19*(7), 717–723. https://doi.org/10.1111/j.1467-9280.2008.02147.x

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

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

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

Skinner, B. F. (1938). *The behavior of organisms: A behavioral analysis*. Oxford.

Slepian, M. L., Young, S. G., Rule, N. O., Weisbuch, M., & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition*, *30*(2), 232–240. https://doi.org/10.1521/soco.2012.30.2.232

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

Stebbins, H. E., & Vanous, J. B. (2015). The influence of stimulus sex and emotional expression on the attentional blink. *Emotion*, *15*(4), 511–521. https://doi.org/10.1037/emo0000082

Stepanikova, I. (2012). Racial-ethnic biases, time pressure, and medical decisions. *Journal of Health and Social Behavior*, *53*(3), 329–343. https://doi.org/10.1177/0022146512445807

Sunnafrank, M. (1986). Predicted outcome value during initial interactions: A reformulation of uncertainty reduction theory. *Human Communication Research*, *13*(1), 3–33. https://doi.org/10.1111/j.1468-2958.1986.tb00092.x

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

Swart, J. C., Froböse, M. I., Cook, J. L., Geurts, D. E., Frank, M. J., Cools, R., & den Ouden, H. E. (2017). Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. *eLife*, *6*, e22169. https://doi.org/10.7554/eLife.22169

Szekeres, H., Halperin, E., Kende, A., & Saguy, T. (2023). Endorsing negative intergroup attitudes to justify failure to confront prejudice. *Group Processes & Intergroup Relations*, *26*(7), 1499–1524. https://doi.org/10.1177/13684302221120488

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, *1*(1), 173–191. https://doi.org/10.1017/S0021932000023336

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

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

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.

Taylor, D. A., Altman, I., & Sorrentino, R. (1969). Interpersonal exchange as a function of rewards and costs and situational factors: Expectancy confirmation-disconfirmation. *Journal of Experimental Social Psychology*, *5*(3), 324–339. https://doi.org/10.1016/0022-1031(69)90057-2

Thijssen, L., Coenders, M., & Lancee, B. (2021). Ethnic discrimination in the dutch labor market: Differences between ethnic minority groups and the role of personal information about job applicants—evidence from a field experiment. *Journal of International Migration and Integration*, *22*(3), 1125–1150. https://doi.org/10.1007/s12134-020-00795-w

Thorndike, E. L. (1927). The Law of Effect. *The American Journal of Psychology*, *39*(1/4), 212–222. https://doi.org/10.2307/1415413

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., Doosje, B., & Amodio, D. M. (2025). Impression formation through social interaction: The effect of ethnicity in the Dutch context. *Group Processes & Intergroup Relations, 0*(0). https://doi.org/10.1177/13684302241305054

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 Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychological Science*, *19*(11), 1131–1139. https://doi.org/10.1111/j.1467-9280.2008.02214.x

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). Consequence-based approach-avoidance training: A new and improved method for changing behavior. *Psychological Science*, *29*(12), 1899–1910. https://doi.org/10.1177/0956797618796478

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 Experiment Social Psychology*, *119*, 104754. https://doi.org/10.1016/j.jesp.2025.104754

van Lent, T., Verwijmeren, T., & Bijlstra, G. (2024). Dishonest collaboration in an intergroup context. *British Journal of Social Psychology*, *63*(1), 256–272. https://doi.org/10.1111/bjso.12675

Veling, H., Becker, D., Liu, H., Quandt, J., & Holland, R. W. (2022). How go/no-go training changes behavior: A value-based decision-making perspective. *Current Opinion in Behavioral Sciences*, *47*, 101206. https://doi.org/10.1016/j.cobeha.2022.101206

Veling, H., Chen, Z., Tombrock, M. C., Verpaalen, I. A. M., Schmitz, L. I., Dijksterhuis, A., & Holland, R. W. (2017). Training impulsive choices for healthy and sustainable food. *Journal of Experimental Psychology: Applied*, *23*(2), 204–215. https://doi.org/10.1037/xap0000112

Verbruggen, F., McLaren, I. P. L., & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives on Psychological Science*, *9*(5), 497–524. https://doi.org/10.1177/1745691614526414

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

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*(1), 56–63. https://doi.org/10.1111/j.0956-7976.2005.00780.x

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

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*(4), 490–497. https://doi.org/10.1177/0956797611400615

Wilke, C. (2024). *cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2"* (Version 1.1.3) [R package]. https://wilkelab.org/cowplot/

Williams, D. R., Lawrence, J. A., Davis, B. A., & Vu, C. (2019). Understanding how discrimination can affect health. *Health Services Research*, *54*(2), 1374–1388. https://doi.org/10.1111/1475-6773.13222

Woud, M. L., Maas, J., Becker, E. S., & Rinck, M. (2013). Make the manikin move: Symbolic approach–avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology*, *25*(6), 738–744. https://doi.org/10.1080/20445911.2013.817413

Yebra, M., Galarza-Vallejo, A., Soto-Leon, V., Gonzalez-Rosa, J. J., de Berker, A. O., Bestmann, S., Oliviero, A., Kroes, M. C. W., & Strange, B. A. (2019). Action boosts episodic memory encoding in humans via engagement of a noradrenergic system. *Nature Communications*, *10*, 3534. https://doi.org/10.1038/s41467-019-11358-8

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

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2009). Interracial contexts debilitate same-race face recognition. *Journal of Experimental Social Psychology*, *45*(5), 1123–1126. https://doi.org/10.1016/j.jesp.2009.05.009

Zadro, L., & Gonsalkorale, K. (2014). Sources of ostracism: The nature and consequences of excluding and ignoring others. *Current Directions in Psychological Science*, *23*(2), 93–97. https://doi.org/10.1177/0963721413520321

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

Zhaoyang, R., Sliwinski, M. J., Martire, L. M., & Smyth, J. M. (2018). Age differences in adults' daily social interactions: An ecological momentary assessment study. *Psychology and Aging*, *33*(4), 607–618. https://doi.org/10.1037/pag0000242

Research data management, transparency, and privacy

This research was conducted in compliance with the General Data Protection Regulation (GDPR) and all applicable laws and ethical guidelines. The Ethical Committee of the faculty of Social Sciences (ECSS) has given a positive advice to conduct these studies to the Dean of the Faculty, who formally approved the conduct of these studies. Radboud University and the Behavioural Science Institute (BSI) have set strict conditions for the management of research data. Research Data Management was conducted according to the FAIR principles. All research data resulting from this dissertation were handled in accordance with the university's research data management policy (https://www.ru.nl/rdm/) and the BSI's research data management protocol (https://www.radboudnet.nl/bsi/rdm).

To enhance open science and transparent research practices, anonymized research data and relevant research materials for each empirical chapter have been made openly available at the Open Science Framework (OSF; https://osf.io). The table below summarizes the materials that are available at the OSF for each chapter. The materials for Chapters 2–3 are publicly available. Because Chapter 4 is still in the peer review process, the materials are accessible only via a private URL. These materials will be made publicly available upon publication of the paper in a peer-reviewed journal.

Ch.	Repository Link	Available Materials
2	https://osf.io/52k73/	Preregistrations Experimental scripts Stimulus materials for Experiment 2.1 R script Anonymized data
3	https://osf.io/38qdk/	Preregistrations Experimental scripts R script Anonymized data
4	https://osf.io/m3z7g/?view_only=dbc4ba2321e34777840f061fa3ffe249	Preregistrations Experimental scripts Stimulus materials R script Anonymized data

English summary

Prejudice and discrimination remain major societal problems, both globally and in the Netherlands. Prejudices are general affective evaluations toward a social category and its members. In other words, how much a person likes or dislikes a social category and its members. If someone acts upon their prejudice, this is called discrimination: Unfair treatment of members of certain social categories. Since prejudice and discrimination have major negative consequences, a crucial question is how to effectively mitigate these phenomena. The main intervention strategy to reduce prejudice is promoting contact between members of different social categories. That is, simply getting people to interact already influences prejudice. However, due to the lack of a clear understanding of how consequences of such interactions influence evaluations, it remains unclear when such interactions lead to a decrease or an increase in prejudice.

Consequences of behavior within contact can be understood as a form of instrumental learning. In instrumental learning, people learn about others through the consequences of behavior. Actions that are rewarded are repeated, and actions that are punished are not repeated. In addition to behavior, instrumental learning can influence evaluations: The rewards and punishments associated with the actions, as well as the decision to act itself, can affect what we think of others. In this dissertation, in addition to actions, the focus is specifically on the impact of absence of actions, since little is known about the effects of such inactions on evaluations.

To better understand these instrumental learning processes, I investigated the instrumental learning processes that contribute to how people evaluate individuals and social categories through three empirical research projects, using four research questions. These are: (1) What is the unique contribution of inactions over and above reward and punishment in shaping evaluations and behavior? (2) Does instrumental learning generalize to evaluations and behavior toward unknown others from the same social category? (3) Does learning information about an individual's social category and personal characteristics influence social decision-making, more specifically, whether this individual is considered in a decision? (4) Does the target's social category influence learning itself?

In Chapter 2, I examined instrumental learning and its consequences for social evaluations and behavior at the individual level. In four online experiments, participants first took part in an instrumental learning task, where they learned to act or not act in response to images of fractals (Experiment 2.1) or faces of individuals (Experiments 2.2, 2.3, and 2.4) to obtain rewards or avoid punishments. Results showed no evidence that the social category of the face (Moroccan–Dutch or White–Dutch) influences learning itself. After the learning task, I measured evaluations of the faces. Results showed that actions, inactions and their consequences (rewards vs. avoiding punishments) influenced evaluations. People evaluated faces most positively if actions had previously resulted in getting rewards for this face, while people evaluated faces most negatively if inactions had previously resulted in avoiding punishment for this face. In addition, the results showed that inactions lead to less positive impressions than actions, even beyond the effect of punishment signals. This shows that inactions play a significant role in shaping impressions. In Chapter 3, I investigated instrumental learning and its consequences for social

evaluations and behavior at the social category level, rather than at the individual level. In the first lab experiment, I replicated a previous pattern of emotion recognition, specifically an interaction effect between social category and expression valence (positive versus negative). Dutch, German or Belgian participants were faster at recognizing happiness as happiness than anger as anger for White–Dutch faces (ingroup), while there was no difference for Moroccan–Dutch faces (outgroup). Next, in two lab experiments, participants first took part in an instrumental learning task in which they learned to act to images of Moroccan–Dutch faces to obtain rewards (this learning had the most positive consequences in Chapter 2) and to not act to images of White–Dutch faces to avoid punishments (this learning had the most negative consequences in Chapter 2). After the learning task, I measured participants' recognition of angry and happy emotional expressions across White–Dutch and Moroccan–Dutch faces. The results demonstrated that instrumental learning influences emotion recognition. Instead of the commonly observed interaction effect (as seen in Experiment 3.1 and previous work) between social category and expression valence (positive versus negative), I consistently showed a main effect of expression valence on emotion recognition. In general, participants were faster at recognizing happiness as happiness than anger as anger, regardless of the social category of the face. This suggests that while emotion recognition for Moroccan–Dutch faces was impacted by the learning, emotion recognition for White–Dutch faces was not. Taken together, instrumental learning seems to influence emotion recognition of existing social categories, and thus social behavior.

In Chapter 4, I investigated instrumental learning and its consequences for social decision-making, focusing specifically on the pre-decision phase, that is, how people are considered when making decisions. In four online experiments, participants took part in the consideration set task, in which they first learned information about the social category and/or personal characteristics of individuals from nonexistent social categories. After learning, participants were asked a neutral question about the individuals. The information just learned about the individuals was not relevant in answering the question. Next, participants were asked which individuals they had considered when answering the previous question. This reflected their consideration sets: A set of individuals they had considered. Results showed that information about social categories did not influence whether the individual was considered, although it did influence evaluations (Experiment 4.2) and preferences (Experiment 4.4). However, information about personal characteristics did influence whether the individual was considered: The more positive this information, the more likely the individual was considered. This shows that even when information is irrelevant to the decision at hand, information about the individual influences the construction of consideration sets.

6

Conclusion

These findings shed light on the instrumental learning processes that contribute to how people evaluate individuals and social categories. First, they show that inactions affect social evaluations and behavior at the individual level. Inactions result in less positive evaluations than actions, even beyond the influence of punishment signals. At the social category level, there is no evidence that punishment-avoidant inactions influence emotion recognition of ingroup targets. Second, these findings suggest that action-reward learning about outgroup individuals generalizes to emotion recognition of new individuals from the same outgroup. Additionally, instrumental learning of information about nonexistent social categories generalizes to preferences for unknown others from the same social category. Third, these findings contribute to insights into social decision making. Information about personal characteristics of an individual positively influences whether that individual is considered, even when the information is irrelevant to the decision at hand. However, there was no evidence that information about an individual's social category influenced whether this individual was considered. Finally, these findings show no evidence that the target's social category influences learning itself.

Taken together, this work highlights the negative effects of inactions. Generally, inactions lead to less positive evaluations of individuals than actions. Moreover, this work highlights the importance of actions with positive consequences. Actions with positive consequences positively influence individual evaluations, behavior, and social decision-making. Additionally, these results suggest that actions with positive consequences positively influence evaluations and behavior at the social category level. Future research could bring the instrumental learning processes studied here in experimental tasks closer to real-world contact, investigate the boundary conditions when consequences of instrumental learning generalize, and contribute to theory development by examining when inactions influence evaluations. I hope the current experimental work inspires future researchers and practitioners to understand and reduce prejudice through instrumental learning.

Dutch summary | Nederlandse samenvatting

Vooroordelen en discriminatie blijven grote maatschappelijk problemen, zowel wereldwijd als in Nederland. Vooroordelen zijn algemene affectieve evaluaties ten opzichte van een sociale groep en haar leden. In andere woorden, hoe leuk of niet leuk iemand een sociale groep en haar leden vindt. Als iemand naar zijn of haar vooroordelen handelt, is er sprake van discriminatie: Oneerlijke behandeling van leden van bepaalde sociale groepen. Vooroordelen en discriminatie hebben grote negatieve gevolgen; daarom is een cruciale vraag hoe ze effectief kunnen worden beperkt. De voornaamste interventiestrategie om vooroordelen te verminderen is het bevorderen van contact tussen leden van verschillende sociale groepen. Mensen simpelweg met elkaar in contact laten komen beïnvloedt al vooroordelen. Het is echter onduidelijk hoe de gevolgen van zulke interacties evaluaties beïnvloeden. Hierdoor is het niet helder wanneer dergelijke interacties leiden tot een afname of toename in vooroordelen.

De gevolgen van gedrag binnen contact kunnen worden begrepen als een vorm van instrumenteel leren. Instrumenteel leren is een leerproces waarbij mensen leren over anderen door de gevolgen van hun gedrag. Acties die worden beloond, worden herhaald, en acties die worden gestraft, worden niet herhaald. Instrumenteel leren beïnvloedt naast gedrag ook evaluaties: De beloningen en straffen die gepaard gaan met de acties, evenals de beslissing om te handelen, kunnen beïnvloeden wat we van anderen vinden. In dit proefschrift ligt de focus, naast acties, specifiek op de impact van de afwezigheid van acties, aangezien er weinig bekend is over de effecten van dergelijke inacties op evaluaties.

Om deze instrumentele leerprocessen beter te begrijpen onderzocht ik in drie empirische onderzoeksprojecten de instrumentele leerprocessen die bijdragen aan hoe mensen individuen en sociale categorieën evalueren aan de hand van vier onderzoeksvragen. Deze zijn: (1) Wat is de unieke impact van inacties bovenop beloning en straf op evaluaties en gedrag? (2) Generaliseert instrumenteel leren naar evaluaties en gedrag ten opzichte van onbekende anderen uit dezelfde sociale groep? (3) Beïnvloedt het leren van informatie over de sociale groep en persoonlijke eigenschappen van een individu de sociale besluitvorming, meer specifiek, of dit individu wordt overwogen in een beslissing? (4) Heeft de sociale groep van het gezicht van het individu invloed op leren zelf?

In Hoofdstuk 2 onderzocht ik instrumenteel leren en de gevolgen daarvan voor sociale evaluaties en gedrag op individueel niveau. In vier online experimenten namen deelnemers eerst deel aan een instrumenteel leren taak, waarin ze leerden om wel of niet te handelen naar afbeeldingen van fractalen (Experiment 2.1) of gezichten van individuen (Experimenten 2.2, 2.3, en 2.4) om beloningen te krijgen of straffen te vermijden. De resultaten toonden geen bewijs dat de sociale groep van het gezicht van het individu (Marokkaans–Nederlands of Wit–Nederlands) het leren zelf beïnvloedt. Na de leertaak mat ik evaluaties van de gezichten. De resultaten toonden aan dat acties, inacties en hun gevolgen (belonen versus het vermijden van straffen) de evaluaties beïnvloeden. Mensen evalueerden gezichten het meest positief wanneer handelen in de leertaak leidde tot het krijgen van beloningen voor dit gezicht, terwijl mensen gezichten het meest negatief evalueerden wanneer niet handelen leidde tot het vermijden van straf voor dit gezicht. Daarnaast toonden de resultaten aan dat inacties leiden tot minder positieve evaluaties dan acties, zelfs boven het effect van strafsignalen. Dit toont aan dat inacties een belangrijke rol spelen bij het vormen van indrukken.

In Hoofdstuk 3 onderzocht ik instrumenteel leren en de gevolgen daarvan voor sociale evaluaties en gedrag op het sociale groepsniveau, in plaats van op het niveau van individuen. In het eerst lab experiment repliceerde ik een eerder patroon in emotieherkenning, namelijk een interactie-effect tussen sociale categorie en expressie valentie (positief versus negatief). Nederlandse, Duitse of Belgische deelnemers herkenden blijdschap sneller als blijdschap dan boosheid als boosheid voor Wit–Nederlandse gezichten (ingroup), terwijl er geen verschil werd gevonden voor Marokkaans–Nederlandse gezichten (outgroup). Vervolgens voerde ik twee lab experimenten uit. In deze lab experimenten namen deelnemers eerst deel aan een instrumenteel leren taak, waarin ze leerden te handelen naar afbeeldingen van Marokkaans–Nederlandse gezichten om beloningen te krijgen (dit leren had de meest positieve gevolgen in Hoofdstuk 2) en niet te handelen naar afbeeldingen van meerdere Wit–Nederlandse gezichten om straffen te voorkomen (dit leren had de meest negatieve gevolgen in Hoofdstuk 2). Na de leertaak mat ik de herkenning van boze en blije emotionele uitdrukkingen door deelnemers bij meerdere Wit–Nederlandse en Marokkaans–Nederlandse gezichten. De resultaten toonden aan dat instrumenteel leren emotieherkenning beïnvloedt. In plaats van het vaak waargenomen interactie-effect (in Experiment 3.1 en in eerder onderzoek) tussen sociale categorie en expressie valentie (positief versus negatief), toonde ik consistent een hoofdeffect van expressie valentie op emotieherkenning aan. Over het algemeen herkenden deelnemers blijdschap sneller als blijdschap dan boosheid als boosheid, ongeacht de sociale groep van het gezicht. Dit suggereert dat de emotieherkenning voor Marokkaans–Nederlandse gezichten wel beïnvloed werd door het leren, maar de emotieherkenning voor Wit–Nederlandse gezichten niet. Alles bij elkaar genomen lijkt instrumenteel leren de emotieherkenning van bestaande sociale categorieën te beïnvloeden, en daarmee sociaal gedrag.

In Hoofdstuk 4 onderzocht ik instrumenteel leren en de gevolgen daarvan voor sociale besluitvorming. Hierbij richtte ik me specifiek op de fase voorafgaand aan de beslissing, dat wil zeggen de manier waarop mensen worden overwogen bij het nemen van beslissingen. In vier online experimenten namen deelnemers deel aan de ‘consideration set task’, waarin ze eerst informatie leerden over de sociale groep en/of persoonlijke eigenschappen van individuen van niet-bestaande sociale groepen. Na het leren werd deelnemers een neutrale vraag gesteld over de individuen. De zojuist geleerde informatie over de individuen was niet relevant bij het beantwoorden van de vraag. Vervolgens werd deelnemers gevraagd welke individuen ze hadden overwogen bij het beantwoorden van de vorige vraag. Dit weerspiegelde hun overwegingsset: Een set van individuen die ze hadden overwogen. De resultaten toonden aan dat informatie over de sociale groep geen invloed had of het individu werd overwogen, hoewel deze informatie wel evaluaties (Experiment 4.2) en voorkeuren (Experiment 4.4) beïnvloedde. Informatie over persoonlijke eigenschappen had echter wel invloed op of het individu werd overwogen: Hoe positiever deze informatie, hoe groter de kans dat het individu werd overwogen.

Dit toont aan dat zelfs wanneer informatie irrelevant is voor de te nemen beslissing, informatie over het individu de samenstelling van overwegingssets beïnvloedt.

Conclusie

De bevindingen werpen licht op de instrumentele leerprocessen die bijdragen aan hoe mensen individuen en sociale categorieën evalueren. Ten eerste laten ze zien dat inacties invloed hebben op sociale evaluaties en gedrag op individueel niveau. Inacties resulteren in minder positieve evaluaties dan acties, zelfs voorbij het effect van strafsignalen. Op het niveau van de sociale groep is er geen bewijs dat strafvermijdende inacties emotieherkenning van ingroup individuen beïnvloeden. Ten tweede suggereren deze bevindingen dat actie-belonend leren over outgroup individuen generaliseert naar emotieherkenning van nieuwe individuen uit dezelfde outgroup. Daarbovenop generaliseert instrumenteel leren van informatie over niet-bestaande sociale groepen naar voorkeuren voor onbekende anderen uit dezelfde sociale groep. Ten derde dragen deze bevindingen bij aan inzichten in sociale besluitvorming. Informatie over persoonlijke eigenschappen van een individu heeft een positieve invloed op of dit individu wordt overwogen, zelfs wanneer de informatie irrelevant is voor de te nemen beslissing. Er is echter geen bewijs gevonden dat informatie over de sociale groep van een individu invloed had of dit individu wordt overwogen. Tot slot tonen deze bevindingen geen bewijs dat de sociale groep van het individu het leren zelf beïnvloedt.

Al met al benadrukt dit werk de negatieve effecten van inacties. Over het algemeen leiden inacties tot minder positieve evaluaties van individuen dan acties. Bovendien benadrukt dit werk het belang van acties met positieve consequenties. Acties met positieve consequenties beïnvloeden individuele evaluaties, gedrag en sociale besluitvorming positief. Daarnaast suggereren deze resultaten dat acties met positieve consequenties evaluaties en gedrag op het niveau van sociale groepen positief beïnvloeden. Toekomstig onderzoek zou de instrumentele leerprocessen die in deze experimentele taken worden bestudeerd dichter bij contact in de echte wereld kunnen brengen, de randvoorwaarden kunnen onderzoeken waaronder consequenties van instrumenteel leren generaliseren, en bijdragen aan theorieontwikkeling door te onderzoeken wanneer inacties evaluaties beïnvloeden. Ik hoop dat het huidige experimentele werk toekomstige onderzoekers en mensen in de praktijk inspireert om vooroordelen te begrijpen en te verminderen door middel van instrumenteel leren.

Acknowledgements | Dankwoord

Dit proefschrift is het resultaat van vele dagen op de universiteit, talloze kopjes koffie en ontelbare gesprekken. Hoewel alleen mijn naam op de voorkant vermeld staat, droegen veel mensen direct of indirect bij aan de totstandkoming van dit proefschrift. Deze mensen wil ik graag hier bedanken.

Allereerst, mijn fantastische team van begeleiders: Gijs, Harm, Rob, en Erik. Jullie zijn stuk voor stuk intelligente, competente, lieve en meedenkende begeleiders. Jullie eerlijkheid over eigen tegenvallers leerde mij met nuchterheid naar academische hobbels te kijken. Door jullie prettige begeleiding kijk ik terug op een mooi, fijn en leerzaam promotietraject waarin ik nooit voelde dat ik alleen op een eilandje aan het werk was.

Gijs, onze samenwerking begon in 2017 toen ik voor je ging werken als onderzoeksassistent. De kans die je me toen bood, is kenmerkend voor jou; ook tijdens mijn promotietraject bood je me allerlei kansen—van meedenken over een aflevering van Het Klokhuis, het meeschrijven van verschillende beursaanvragen, tot het geven van een praatje voor het autismenetwerk. Dit verbreedde mijn horizon en ervaring, en ik ben je daar erg dankbaar voor. Daarnaast leerde je me dat vriendelijkheid vooropstaat, ongeacht wie er tegenover je zit: van een eerstejaarsstudent tot een onaardige high-profile editor. “Kill them with kindness” lijkt wel jouw motto. Ik ervaarde onze samenwerking als erg plezierig, zo hadden we veel lol over klassiek kantoorjargon zoals ‘in de wandelgangen uitkruisen’ en ‘inkoppen’. Je enthousiasme voor onderzoek werkt aanstekelijk. Voor jou is alles ‘gaaf’ en dat enthousiasme heeft mij zeker beïnvloed. Ik ben blij dat onze samenwerking er nog niet geheel op zit en dat er nog projecten lopen.

Harm, een van de eerste dingen die je tegen me zei toen ik aan dit promotietraject begon was dat we ‘slow science’ zouden beoefenen. Deze focus op kwaliteit gaf mij veel rust. Gedurende het promotietraject leerde je me de juiste vragen te stellen (‘Wat is nou precies de onderzoeksvraag?’), nog concreter te worden en niet bang te zijn om toe te geven als je iets niet helemaal begrijpt. Ik genoot van al onze gesprekken over een hele waaier aan onderwerpen zoals: onderzoek, gezondheid, politiek, feminisme en goede restaurants in Nijmegen. Tijdens een van onze laatste gesprekken op de Radboud vroeg je me wie ik nou een inspirerend persoon in de wetenschap vond. Hoewel ik daar toen geen antwoord op formuleerde, denk ik nu dat jij onder andere zo’n persoon bent.

Rob, in een eerdere fase van het promotietraject was je wat meer op de achtergrond, maar toen Harm van baan wisselde raakte je dichter betrokken bij het project. Ik genoot van onze vaste meetings op vrijdag. Het hoogtepunt voor mij was onze gezamenlijke trip naar Krakau en gelijk door naar Bath voor twee congressen op rij. Tijdens deze trip kwam jouw ontspannen en positieve houding, die ik zo inspirerend vind, duidelijk tot uiting. Zo schreef je je per ongeluk in voor het verkeerde congres en vergat je je creditcard waardoor ik je nootjes (want er moeten altijd nootjes zijn!) in het vliegtuig voorschoot. Bovendien hielp je me contact te leggen met Russ voor mijn onderzoeksbezoek. Ik ben erg dankbaar dat je je netwerk met mij deelde.

Erik, ondanks dat je wat meer vanaf de zijlijn betrokken was bij mijn promotietraject, was je begeleiding van grote waarde. Ik volgde niet altijd direct je schrijfadvies kloppend op, maar je was geduldig in je feedback. Je hebt een groot oog voor detail. Dit komt niet alleen tot uiting in schrijven, maar ook bijvoorbeeld in data visualisatie en het feit dat je een soort professionele studio in je huis hebt om colleges op te nemen. Ik genoot van onze afspraken samen en van je open blik op mijn promotieonderzoek. Je dacht altijd goed mee, en kwam met goed advies voor bijvoorbeeld keuze voor journals, manieren om het onderzoek online aan te prijzen, of het inhuren van een tekenaar om een onderzoeksopzet te visualiseren.

Moreover, I would like to thank the members of the manuscript committee (**Karin Roelofs, David Amodio,** and **Kerry Kawakami**) and opposition for their willingsness to review my dissertation and attend my defense. Thank you for the time and attention you have put into this.

Ik wil ook graag iedereen bedanken die het onderzoek mogelijk heeft gemaakt. Allereerst **Hubert**, jij was mijn rots in de branding wat betreft het programmeren van experimenten. Naast je hulp met programmeren, was je ook een prettige collega en hadden we allerlei leuke gesprekken zoals over de Efteling, carnaval en restauranttips in Groesbeek. Daarnaast wil ik alle deelnemers, alle onderzoeksassistenten, alle studenten, **Ronny, Anna** en het **BSI management**; **Meta**, **Debby**, **Sander**, **Julie**, **Diana** en **Sonia**, bedanken.

Mijn paranimfen, **Joël** en **Jard**, ik wil jullie bedanken voor de vriendschap en ondersteuning. **Joël**, met jou voelde ik meteen een klik, zowel persoonlijk als werk gerelateerd. Je houdt me bij de tijd wat betreft online cultuur en vindt me delulu als ik niet weet wat delulu betekent. Samen naar de ASPO was een professionele highlight, maar daarnaast heb ik ook genoten van onze biertjes op het terras en uitstapjes naar de film of het theater. Je bent niet bang om je uit te spreken als er ergens onrecht is. Dit is een grote inspiratiebron voor mij. **Jard**, de band die zussen hebben blijft erg speciaal. Ook al woon je vrij ver weg (te ver als je het mij vraagt!), ben ik blij dat we erg in contact staan. Ik heb genoten van alle onze museum-, restaurant- en actieve uitjes in al die jaren. Ik kijk uit naar alle avonturen die de toekomst ons gaat brengen. Ik blijf me erover verbazen hoe slim en wat voor een doorzetter jij bent! Ik ben blij dat jullie beide aan mijn zijde staan tijdens de verdediging.

Mijn collega's van de Behavior Change & Well-Being/Social & Cultural Psychology Group. Ik kijk terug op een enorm fijne tijd waarin ik omgeven was door fijne en open-minded mensen, en ik wil graag iedereen van de afdeling en de labgroepen (Behavior Regulation and Change & Radboud Social Cognition Lab) hiervoor bedanken. In het bijzonder wil ik graag een paar collega's extra bedanken. **Martijn**, dankjewel voor al je advies—van lid worden van een vakbond, tot je hulp bij salarisonderhandelingen. Ik heb genoten van onze gesprekken over wijn, gedragsverandering en obscure stonerbands. Een hoogtepunt voor mij was ons auto uitje naar de Dag van het Gedrag (de file op de terugweg en het feit dat ik ons toen verkeerde navigeerde vergeten we maar even). **Thijs**, ons avontuur begon in 2017, toen we samen in zee gingen voor mijn masterscriptie. Voor dit project zijn we ook samen naar Lowlands gegaan om data te verzamelen met jouw en mijn vrienden; een ervaring waar ik erg positief op terugkijk. Je bent een vrolijke verschijning en als jij er bent is een werkborrel meteen een stuk leuker. Naast dat je veel vrolijkheid brengt, ben je ook een erg competente wetenschapper en heb je me in mijn promotietraject goed geholpen met het programmeren van één van de experimenten. **Cor**, dankjewel voor de gezellige (en soms serieuze) gesprekken. Daniela, ik had me geen betere overbuurvrouw kunnen wensen. Dankjewel voor het zijn van een rolmodel, met je eerlijkheid, gezelligheid en enthousiasme. Johan, jouw warmte en eerlijkheid zorgen ervoor dat iedereen zich snel thuis voelt op de afdeling. Mariet, Mariëm, Madelon en alle andere medewerkers van het secretariaat, dankjewel voor jullie ondersteuning bij allerlei (al dan niet bureaucratische) zaken. Speciaal bedankt voor het helpen bij het financiële labyrint van de universiteit.

I want to thank my officemates. Max, in my first year of the PhD, I was a little lonely in the room, but luckily you came along. We have had endless conversations on common interests such as psychology and stats, but also on not so common interests such as musicals and metal bands. It was inspiring to see how you approach science. Kim Lien, het was een eer jouw paranimf te zijn, dankjewel voor al je crazy ideeën die de werkdag een stuk minder eentonig maakte; zoals het opnemen van een dansvideo over je proefschrift. Leo, thank you for listening to all my questions about philosophy and trying to answer them in 'normal' language, it was great to be your officemate. Eric, it was fun to talk to you about various topics, from investing to value-based decision making. Thank you for teaching me how to play to poker, resulting in me winning the poker game at writing week (gentle reminder: I still didn't get my money).

I also want to thank all other PhDs, each of whom contributed uniquely to the department. Iris, het was een eer jouw paranimf te zijn. We hebben enorm veel gedeelde werkinteresses en ik genoot van onze gesprekken over de connectie tussen onderzoek en praktijk. Ik bewonder jouw werkethos. Julian, you are extremely helpful, a very competent researcher and I enjoyed the (too few!) beers we had together. Ted, it was an honor to be your paranymph, and especially in the first half of my PhD I would like to thank you for sharing your perspective on my research. You are a fast learner and skilled researcher. Rob, ook al kennen we elkaar al lang en hebben onze paden zich vaak gekruist, ben ik je tijdens mijn tijd op de afdeling beter gaan leren kennen. Dankjewel voor het luisterend oor en de gezellige tijd. Dannis, it was great to be your 'finish PhD' buddy at the end of both our PhDs. Thank you for the motivation and the games of ping pong. Farnaz, you pointed out all kinds of peculiarities in Dutch culture. Thank you for broadening my perspective. It was special to be present when you received Dutch citizenship. Tiara, your presence was calming and I loved exchanging ideas with you. Daniel, thank you for hosting multiple PhD get togethers. Sari, Afreen, Erin, Matteo, Ege, Charlotte, Joëlle, Tess, Martijn, Milou, and Theresa, thank you for the good times and for giving me a different perspective.

Ook buiten de afdeling heb ik met ontzettend fijne mensen samengewerkt.

Het ASPO bestuur, bedankt voor de kennismaking met psychologie op andere universiteiten dan de Radboud en voor de gezellige borrels en diners. Hedy Greijdanus, dankjewel voor de hele prettige samenwerking bij de organisatie van Social Animal.

Floor Burghoorn, dankjewel voor onze eindeloze chat gesprekken op mattermost en voor je advies over statistiek. Op het einde was mattermost compleet uitgestorven, maar zat ik daar nog steeds met jou te chatten. Ik hoop dat we elkaar in de toekomst blijven vinden. Gijsje, ik heb genoten van jouw vrolijke aanwezigheid in het RSCL en je inspireert me je eigen gevoel te volgen. Anouk, samen met jou DRIVE geven was een prettige ervaring. Hoe jij naar het werkveld van gedragsverandering kijkt is een inspiratiebron voor mij. **Ilse**, de paar keer per jaar dat jij de Radboud Universiteit met je aanwezigheid kwam verblijden, waren altijd hartstikke gezellig. Ik heb genoten van onze gesprekken over wetenschap en de cultuur aan universiteiten, maar bovenal van mijn bezoekje aan jou in Berlijn. Ik hoop dat je me binnenkort in Utrecht komt opzoeken.

I feel very grateful for the opportunity to visit **Russ Fazio** and **Suraiya Allidina** at Ohio State University, USA. I immensely enjoyed talking about research with you, this was very inspiring. You were extremely warm and welcoming, thank you for that.

Dan mijn vrienden en familie.

Lieve **Eline**, mijn vriendin van de studie, wat was het mooi geweest als we deze mijlpaal samen konden vieren, zoals we dat bij zoveel mijlpalen deden. Het leven is een stuk minder leuk zonder je, ik mis je.

Lieve **Kim**, mijn vriendin van de studie, ondanks dat je nooit je samenvattingen met me wilde delen tijdens de studie is het me toch gelukt om te promoveren. Vooral het afgelopen jaar was je af en toe mijn persoonlijke adviseur, en dat was een enorme hulp. Bovenal ben je natuurlijk een goede vriendin en gaat er geen week voorbij zonder je te zien. Je bent lief, avontuurlijk, slim en open-minded. Ik voel me extreem bevoorrecht met jou in mijn leven en ik hoop op nog vele avonturen samen.

Mijn lieve Eindhoven vrienden, **Zoe G., Nine, Zoe D., Lisanne, Wouter, Davy**, en **Niels** (aka The Wolfpack), dankjewel voor jullie Brabantse nuchterheid, gezelligheid en alle avonturen die we samen hebben meegemaakt. Ik koester deze herinneringen. **Lisanne**, dankjewel voor de introductie in Nederlandse volksmuziek. Ik hoop dat we nog vaak van links naar rechts gaan bij de Snollebollekes. Mijn oudste vriendin, **Zoe G.**, wij zijn al vrienden vanaf dat we vier zijn en dat is me heel dierbaar. Jouw creativiteit inspireert mij enorm. Een zilveren vriendschap en op naar goud!

Lieve **Irene**, mijn vriendin van de studie, we hebben knotsgekke dingen samen meegemaakt en je bent al die jaren een grote support geweest. Zo wacht je al jaren om met een spandoek in de academiezaal te kunnen zitten tijdens mijn verdediging (ik hoop van harte dat dit niet echt gaat gebeuren). Dankjewel voor je luisterend oor, je zachtaardigheid en je erg heftige mixdrankjes.

COVID bracht veel ellende, maar ook nieuwe vriendschappen. Zoals met **Renske**, toch leuk dat ik bij deze cursus kan promoveren! Grapjes terzijde; dankjewel voor je vriendschap. Ik genoot van je enthousiasme, de gezellige etentjes en het ontdekken van nieuwe artiesten samen. En ook met **Iris**, doordat we tegelijk aan onze PhD zijn begonnen zaten we in hetzelfde schuitje. Dankjewel voor alle gezellige momenten samen, met als hoogtepunt ons georganiseerde Space B weekendje. Jouw relativeringsvermogen is een bron van inspiratie!

Lieve **Fenna**, mijn middelbare school vriendin, dankjewel voor alle interessante gesprekken over politiek en onze gedeelde liefde voor port en dessertwijn. Ik ken niemand die zo lief is als jij. Als jij er bent is het altijd gezellig; op naar nog vele jaren!

Mijn lieve oud-huisgenoten, **Pim, Casper** en **Dio** (aka koffie uni; al ben ik de enige die nog koffie op de uni drinkt). Hoewel jullie gruwelen van het doen van een PhD, hebben jullie mij door de jaren heen goed gesteund. Dankjewel voor het koffie drinken, padellen en altijd maar weer friet eten.

Britt, ondanks dat je soms grappend twijfelt aan mijn onderzoeksvaardigheden ('jij maakt vast je grafieken in Excel') ben je in de laatste fase van mijn PhD een erg grote steun geweest. Dankjewel voor alle gezellige (al dan niet sarcastische) gesprekken.

Larissa, soms moet je naar de andere kant van de wereld reizen, om een Nederlander tegen te komen met wie het heel erg klikt en met wie je gestoorde avonturen beleeft. Dankjewel dat we af en toe opduiken in elkaars leven.

Madou, jouw enthousiasme en 'work hard play hard' mentaliteit blijven een bron van inspiratie (en soms verbazing).

Oud collega talents van de BCG, **Dané, Kyron, Marenne, Sanne, Sjors** en **Sterre**, dankjewel voor onze regelmatige etentjes samen en het mij bij de tijd houden van wat er zich afspeelt omtrent gedragsverandering buiten de universiteit. Ook dankjewel aan **Kevin, Luuk, Gert, Max**, en **Tanneke** voor jullie zeer waardevolle begeleiding bij de start van mijn professionele leven.

Vrienden van Floor, **Mandy, Giuseppe, Chloe, Boy, Evi** en **Rick**, dankjulliewel voor alle mooie momenten die we samen hebben beleefd.

Roy, bedankt voor de gezamenlijke muzikale uitspattingen op Boom.

Johnny, dankjewel voor het bieden van een welkome sportieve afleiding. Naast sporten was het ook leuk om met je te kletsen over onzin en psychologie.

Alle lieve mensen van Nijmegen feestteam, zoals **Fieke, Wouter, Sjacky, Maarten, Roos, Joey, Jan, Bjorn, Peter** (een bezoekje aan state of trance is inderdaad de beste voorbereiding voor een sollicitatie voor een postdoc), en **Jidde**. Bedankt voor alle mooie feestjes en borrels samen: van oud & nieuw, tot Down The Rabbit Hole tot de vierdaagse, ik heb genoten!

Mijn lieve ouders, **Maarten** en **Mariëlle**, jullie zijn mijn fundament. Ik voel me trots jullie dochter te zijn.

Mijn grootouders, **Edith, Harry, Noor** en **Paul**, van wie niemand helaas dit moment meemaakt, dankjulliewel voor het geloven in mij. Ik weet zeker dat jullie dit prachtig hadden gevonden. Ook bedankt aan de gehele **van Lenten familie**, van bezoekjes in Amerika tot kerstmis, het was gezellig!

Mijn schoonouders, **Paul** en **Tanja**, dankjewel voor jullie onvoorwaardelijke steun.

Tot slot, lieve **Floor**, sinds onze ontmoeting zijn we onafscheidelijk. Jouw optimisme en onuitputtelijke bron van enthousiasme maken dat ik zin heb in elke dag. En hoewel je de afgelopen jaren helemaal gek werd van mijn eindeloze geklets over werk, en me wekelijks een nieuw 'label' gaf, wil je gelukkig nog steeds met me zijn. Het maakt niet uit waar we zijn, als jij er bent, dan ben ik thuis. Volgens ons, zelfbenoemde relatie-experts, is er altijd één ding dat mensen bindt in een relatie, en in ons geval is dat, denk ik, het openstaan voor en 'ja' zeggen tegen nieuwe ervaringen en ideeën, en juist daarom verheug ik me op alles wat we samen gaan beleven. Ik verheug me ook op hoe slecht je gaat op het aantal komma's in de voorgaande zin.

Tjits
September 2025

About the author

Tjits van Lent was born on August 16th, 1994, in Amsterdam. After finishing primary education at school ‘De Opbouw’ in Eindhoven, she went on to the Lorentz Casimir Lyceum in Eindhoven. In 2012, she started studying Psychology at the Radboud University, Nijmegen. During her study she participated in the honours programme, founded a student wine association, and was active in several committees, such as the introduction committee of Psychology. As part of her study, she went a semester abroad to Stellenbosch University, South Africa. In 2018, she graduated cum laude from the two-year Research Master Behavioural Science, focusing on how the people we collaborate with can influence our honesty. She then worked at a behavior change company in Nijmegen.

In 2020, she started to work on her PhD project on the role of instrumental learning in prejudice, supervised by Gijs Bijlstra, Harm Veling, Rob Holland and Erik Bijleveld at Radboud University Nijmegen. During her PhD, she was awarded a seedcorn grant (2024) and the Christine Morhmann stipend (2024). The latter allowed her to visit Russ Fazio at Ohio State University, USA. In addition to her research, Tjits was a board member of ASPO (Associatie van Sociaal Psychologische Onderzoekers) for two years. Currently, Tjits works as a postdoctoral researcher at Utrecht University.

List of academic publications

van Lent, T., Veling, H., Holland, R. W., Bijleveld, E., & Bijlstra, G. (2025). Reshaping the happy face advantage with reinforcement learning. *Cognition and Emotion*, 1–17. https://doi.org/10.1080/02699931.2025.2568553.

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

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

van Lent, T., Verwijmeren, T., & Bijlstra, G. (2024). Dishonest collaboration in an intergroup context. *British Journal of Social Psychology, 63*(1), 256–272. https://doi.org/10.1111/bjso.12675

Bijlstra, G., Kleverwal, D., **van Lent, T.,** & Holland, R.W. (2019). 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

Radboud University

IN·DEI·NOMINE·FELICITER

Behavioural
Science
Institute