Where Does Taste-Based Discrimination Come From?

3

As the last chapter has revealed, the reason why a decision-maker makes use of statistical discrimination is easily comprehensible. If a decision situation underlies uncertainty, he has to assess the probabilities of possible scenarios with some degree of vagueness. In this process, group memberships of providers can serve as a proxy for these probabilities.¹ So, statistical discrimination is a tool so as to better handle uncertainty and in this way commonly applied. As Lippert-Rasmussen (2014) states: “[A]ll of us engage in statistical discrimination in that we treat people differently on the basis of explicit or implicit statistical generalizations pertaining to the group to which they belong; native speakers speak more slowly when talking to nonnative speakers (which, generally speaking, is quite nice and facilitates understanding); women walking home at night respond differently to an approaching lone stranger if this person is male than if she is a female; racial minority members are more alert to signs of racial bias when speaking to a majority member than when speaking to another minority member. Indeed, acting in a social world without relying on statistical information about socially salient groups seems impossible.” (p. 80)

But why do we have certain tastes (and distastes) for other people? Already Becker (1971) said that the causes of taste-based discrimination have to be sought in psychology (and sociology) and that he merely analysed the economic consequences of it. Therefore, in this chapter we consult psychological and evolutionary

¹ Which beliefs are rational to be held and used in case of statistical discrimination and which not will be discussed in chapter 4.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-658-34569-3_3) contains supplementary material, which is available to authorized users.

biological concepts so as to find proximate and ultimate explanations for taste-based discrimination.² This is important out of two reasons: First, it reveals how our tastes are structured and thereby whether they are fixed or dependent on external aspects such as social context and culture. Second, there is a discussion about whether such tastes and therefore preferences for certain people/groups actually exist which brings us to the question of whether and how they could have evolved.

The chapter is structured as follows: First, we introduce the idea of ingroup favouritism and discuss how it is linked to taste-based discrimination. Second, we analyse how we can delimitate taste-based discrimination from statistical discrimination and thereby ask whether the former truly exists. Third, we investigate ultimate explanations for taste-based discrimination and in so doing present the evolution of agent-relative social preferences.

3.1 A Taste for the Ingroup

We know from chapter 2 that a taste-based discriminator prefers certain people or groups to others and because of that treats these people or groups better than others. To put it differently, the preference ordering of a taste-based discriminator is not agent-neutral but agent-relative. In this chapter, we are mainly interested in what we called strong taste-based discrimination. For repetition, we defined strong taste-based discrimination as follows: The decision-maker is willing to bear costs in order to choose the alternative whose characteristics are provided by the preferred person. In formal terms, under the assumption that $I = \{1, 2\}$ and $M = \{1, 2\}$, where characteristics 1 are preferred to characteristics 2 and provider 2 is preferred to provider 1:

$$x_1, x_2 \in \mathcal{X} : u(x_1) > u(x_2)$$

$$\wedge x_1^1, x_2^2 \in X : u(x_1^1) \leq u(x_2^2)$$

However, this definition is limited to a provider situation, meaning where the provider of an alternative's characteristics is relevant. This perspective on an interaction process is no longer sufficient. We have to expand it to situations where not the *provider* of certain characteristics but the *receiver* of these characteristics

² In section 4.3, we will discuss the sociological implications and consequences of taste-based discrimination.

is relevant.³ One major difference between these two situations is that while we excluded that the decision-maker himself can be a provider, he very well can be a receiver.

Therefore, in this chapter, we first define taste-based discrimination in a receiver situation. Then, we examine what determines how altruistic we behave towards others. In order to do that we introduce ingroup favouritism and social identity theory. Next, we investigate whether ingroup favouritism stems from ingroup love, outgroup derogation, or both. Finally, we demonstrate that not all tastes have to stem from an ingroup-outgroup context, yet, social identity is often still intertwined with them when we look more closely.

3.1.1 Defining Taste-Based Discrimination in a Receiver Situation

When we introduced agent-neutrality and agent-relativity, we have already encountered a choice set where the receiver and not the provider of certain characteristics is relevant. There, we discussed an example where a decision-maker has a choice set X with the following three alternatives: x_1 = “the decision-maker gets \$100”; x_2 = “person 2 gets \$100”; x_3 = “person 3 gets \$100”. Additionally, we assumed that the decision-maker, person 2, and person 3 would be all equally happy to get \$100, provided that there is no further information that tells us differently. We now want to adjust this notation so as to make it more applicable. Instead of having three characteristics (1 = “the decision-maker gets \$100”; 2 = “person 2 gets \$100”; and 3 = “person 3 gets \$100”), we only use one (1 = “receiver gets \$100”). The identity of the receiver who gets the \$100 is indicated by m (or $\mathcal{M}_a$ if we consider group memberships), which in this case could be the decision-maker (DM), person 2 (P2), or person 3 (P3). Applying our new notation, $X = \{x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ}\}$. Note that the little circle ($^\circ$) marks that DM, P2, and P3 are receivers and not providers of the alternative’s characteristics.

How do we differentiate weak and strong taste-based discrimination in a receiver situation? We have to distinguish two cases. Case number one involves that the decision-maker is not a possible receiver. In such a situation, there for example is case of weak taste-based discrimination if the decision-maker is indifferent between the characteristics of his alternatives but still prefers one alternative to another. For example, $I = \{1, 2\}$, where 1 = “receiver gets a \$100 note” and

³ For example, tip givers or effective altruists that we discussed in previous chapters act in a receiver situation.

2 = “receiver gets two \$50 notes”. We presuppose that the decision-maker is indifferent between x_1 and x_2 in a choice set $\mathcal{X}$, where the receivers’ identity is unspecified. Now, given further knowledge about the receivers’ identity leads to a preference of one alternative over the other in a choice set $X = \{x_1^{P1^\circ}, x_2^{P2^\circ}\}$, there is weak taste-based discrimination.⁴ In formal terms:

$$x_1, x_2 \in \mathcal{X} : u(x_1) = u(x_2)$$

$$\wedge x_1^{P1^\circ}, x_2^{P2^\circ} \in X : \left[u(x_1^{P1^\circ}) > u(x_2^{P2^\circ}) \right] \dot{\vee} \left[u(x_1^{P1^\circ}) < u(x_2^{P2^\circ}) \right]$$

We assume that strong taste-based discrimination is inexistent in a situation where the decision-maker is not a possible receiver. The reason for this is that since the decision-maker is not a possible receiver, he cannot bear any costs in the first place, which is a requirement for strong taste-based discrimination.

This assumption might face the following objection: Let’s say there are two possible receivers of \$100 called Barbara and Ben. The decision-maker knows that if Barbara gets \$100, she will give him back \$20. In contrast, he also knows that Ben will keep all the money. As a consequence, if the decision-maker still decides that Ben gets \$100 due to agent-relative preferences, he would bear costs and thus display strong taste-based discrimination. However, this is a fallacy because in this example, the characteristics of the two alternatives are not the same. While the characteristics of the alternative where Ben is the receiver are “receiver gets \$100”, those of the alternative where Barbara is the receiver are “receiver gets \$100 and gives decision-maker \$20 back”. Therefore, if he decides to give Barbara \$100, he becomes a receiver as well which enables him to bear costs and display strong taste-based discrimination.

Let’s continue with case number two: The decision-maker is one of the possible receivers. Here, the setup is more complicated and needs several steps. By way of illustration, we use the same example as at the beginning of this subchapter. Our choice set X consists of three alternatives that always have the same characteristics i ($I = \{1\}$) but differ regarding the identity of the receiver ($M = \{DM^\circ, P2^\circ, P3^\circ\}$). So, $X = \{x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ}\}$. Moreover, the characteristics 1 = “receiver gets \$100”, ($1 \in I$).

Now, as a first step, we have to clarify whether the decision-maker would want to receive the alternative’s characteristics 1 or not in a hypothetical isolated

⁴ We exclude the possibility that further information might lead to statistical discrimination or the knowledge that one receiver does actually not want to receive characteristics i .

decision situation. An isolated decision situation implies that there is only one possible receiver. In this way, the decision-to-be-taken can only affect the outcome of that receiver (which is the decision-maker in our case). We do this as follows: We add a second element to the set I . Thus, I newly consists of 1 and 2 ($I = \{1, 2\}$). This second element of I constitutes the negation of the first one. As a result, 2 = “receiver does not get \$100”, ($2 \in I$). From here, we build a second choice set $\mathbb{X}$ that has two elements: $\mathbb{X} = \{x_1^{DM^\circ}, x_2^{DM^\circ}\}$. A preference ordering on this choice set $\mathbb{X}$ indicates whether the decision-maker would rather receive characteristics 1 or not (and thus receive characteristics 2) given he is the only possible receiver. In case of our example, we assume that the decision-maker prefers $x_1^{DM^\circ}$ to $x_2^{DM^\circ}$, leading to the following formulation:

$$x_1^{DM^\circ}, x_2^{DM^\circ} \in \mathbb{X} : u(x_1^{DM^\circ}) > u(x_2^{DM^\circ})$$

In a second step, we examine the preference orderings of the other receivers (person 2 and person 3) regarding the choice set $X = \{x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ}\}$. We do that out of the perspective of the decision-maker and thus use the decision-maker’s assumptions about the utility function of person 2 ($u_{DM}^{P2^\circ}$) and person 3 ($u_{DM}^{P3^\circ}$). Moreover, we assume that the decision-maker’s assumptions about others’ utility functions are always correct and therefore $u_{DM}^{P2^\circ} = u_{P2}$ and $u_{DM}^{P3^\circ} = u_{P3}$, which is why we directly use u_{P2} respectively u_{P3} in the formulations.⁵ Now, let’s say that both person 2 and person 3 prefer the alternative where they themselves get \$100 and otherwise are indifferent as indicated by the following preferences:

$$x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : u_{P2}(x_1^{P2^\circ}) > u_{P2}(x_1^{DM^\circ}) = u_{P2}(x_1^{P3^\circ})$$

$$x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : u_{P3}(x_1^{P3^\circ}) > u_{P3}(x_1^{DM^\circ}) = u_{P3}(x_1^{P2^\circ})$$

Note that so long as there is no further information that tells us differently, we infer from such preferences that person 2 and person 3 are equally happy to receive characteristics 1. In turn, this implies that an agent-neutral decision-maker has reason to give characteristics 1 to any of the two.

Building on this pre-setup, we can now define weak and strong taste-based discrimination in a decision situation where the decision-maker is a possible receiver

⁵ If u is not further specified, it describes the decision-makers utility function (which would be u_{DM}).

and all alternatives involve the same characteristics. We start with weak taste-based discrimination. Since the decision-maker prefers $x_1^{DM^\circ}$ to $x_2^{DM^\circ}$ within choice set $\mathbb{X}$, we know that he generally prefers getting \$100 to not getting \$100. Next, we assume that $x_1^{DM^\circ}$ is also the most preferred alternative within choice set X , which implies that the decision-maker has egoistic preferences. In this dissertation, provided that there are no strategic reasons to do differently, such preferences involve that their holder (a) always chooses the same alternative in a choice set with all possible receivers as in his isolated choice set and if this is not possible (b) least likely chooses that alternative in a choice set with all possible receivers which is lesser preferred in his isolated choice set.⁶ Now, given he has weakly agent-neutral preferences, he is indifferent between $x_1^{P2^\circ}$ and $x_1^{P3^\circ}$. Accordingly, if a decision-maker is not indifferent between these two alternatives, he displays weak taste-based discrimination, as can be seen in the following formulation:

$$\begin{aligned}
 & x_1^{DM^\circ}, x_2^{DM^\circ} \in \mathbb{X} : u(x_1^{DM^\circ}) > u(x_2^{DM^\circ}) \\
 & \wedge x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : u_{P2}(x_1^{P2^\circ}) > u_{P2}(x_1^{DM^\circ}) = u_{P2}(x_1^{P3^\circ}) \\
 & \wedge x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : u_{P3}(x_1^{P3^\circ}) > u_{P3}(x_1^{DM^\circ}) = u_{P3}(x_1^{P2^\circ}) \\
 & \wedge x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : [u(x_1^{DM^\circ}) > u(x_1^{P2^\circ}) > u(x_1^{P3^\circ})] \\
 & \dot{\vee} [u(x_1^{DM^\circ}) > u(x_1^{P3^\circ}) > u(x_1^{P2^\circ})]
 \end{aligned}$$

We notice that there are two ingredients of weak taste-based discrimination in a situation where the decision-maker himself is a possible receiver: agent-relative preferences and egoistic preferences. The former state that the decision-maker treats receivers differently. The latter guarantee that the decision-maker is not willing to bear costs in order to choose an alternative in the choice set with all possible receivers that differs from the preferred one in his isolated choice set.⁷

⁶ In fact, if there are strategic reasons to do differently, the alternatives' characteristics of the respective receivers differ from each other (maybe only in a statistically discriminatory sense). We will discuss such strategic reasons in section 3.2.

⁷ We could actually differentiate between a weak and a strong type of egoistic preferences. Let's say that there are two best alternatives in a decision-maker's choice set. We call them alternative 1 and alternative 2. While choosing alternative 1 is also favourable for the other

This is different in case of strong taste-based discrimination. Here, the decision-maker is willing to bear costs in order to choose the alternative whose characteristics are received by the preferred person. As a consequence, a strong taste-based discriminator cannot have egoistic preferences but needs to have social preferences. Such preferences enable altruistic and/or antisocial behaviour. Fehr (2015) defines altruistic behaviour as follows: “If a person acts in a way that is costly for herself but provides a benefit [disbenefit] to someone else, the person’s behavior is altruistic [antisocial]. The actor is not motivated by direct or indirect future material benefits associated with the act, but she may still experience a psychological benefit. She may feel better because she engaged in the altruistic [antisocial] act, but according to this definition, that does not prevent it from being altruistic [antisocial].” (p. 78) The definition for antisocial behaviour was added in brackets. Yet, note that from now on, we will not always mention the antisocial manifestations of social preferences as well since we mainly concentrate on altruistic behaviour.

Let’s technically illustrate this definition. We shrink the above example where a decision-maker has to decide who of three people gets \$100 to a two-person setup. We again call these two receivers “DM” for decision-maker and “P2” for person 2. So, $I = \{1, 2\}$, where 1 = “receiver gets \$100” and 2 = “receiver does not get \$100”, $M = \{DM^\circ, P2^\circ\}$, the actual choice set $X = \{x_1^{DM^\circ}, x_1^{P2^\circ}\}$, and the hypothetical isolated choice set $\mathbb{X} = \{x_1^{DM^\circ}, x_2^{DM^\circ}\}$. Furthermore, we make the following two assumptions: (1) In the isolated decision situation, the decision-maker prefers getting \$100 to not getting \$100. (2) If the decision regarding choice set X were up to person 2, he would prefer that person 2 (he himself) gets \$100 to the other alternative. In such a situation, the decision-maker has altruistic preferences and as a result behaves altruistically if there are the following preference orderings:

$$x_1^{DM^\circ}, x_2^{DM^\circ} \in \mathbb{X} : u(x_1^{DM^\circ}) > u(x_2^{DM^\circ})$$

persons involved in the decision situation, choosing alternative 2 is not. Now, in case of strong egoistic preferences, the decision-maker is indifferent between alternative 1 and alternative 2. Therefore, someone with strong egoistic preferences only cares about himself. In contrast, in case of weak egoistic preferences, the decision-maker prefers alternative 1 to alternative 2 (altruistic manifestation) or vice versa (antisocial manifestation). Therefore, someone with weak egoistic preferences first cares about himself and then, if possible, also considers others. Yet, this differentiation of egoistic preferences is not of importance for this dissertation, which is why we do not use it.

$$\wedge x_1^{DM^\circ}, x_1^{P2^\circ} \in X : u_{P2}(x_1^{DM^\circ}) < u_{P2}(x_1^{P2^\circ})$$

$$\wedge x_1^{DM^\circ}, x_1^{P2^\circ} \in X : u(x_1^{DM^\circ}) \leq u(x_1^{P2^\circ})$$

This means that the decision-maker basically prefers getting \$100 to not getting \$100. However, if getting \$100 implies that person 2, who wants to get \$100, does not get \$100, the decision-maker rather relinquishes the \$100 and gives them to person 2 or is indifferent between those two alternatives. To put it differently, the decision-maker acts in a way that is costly for himself but provides a benefit to someone else which precisely is Fehr's definition of altruistic behaviour.⁸

Now, let's get to strong taste-based discrimination. We use the same setup as above, add a third receiver ($M = \{DM^\circ, P2^\circ, P3^\circ\}$), and assume that the decision-maker prefers $P2$ to $P3$. There is strong taste-based discrimination in such a situation if the following requirements are fulfilled: (1) In a hypothetical isolated choice set $\mathbb{X}$, the decision-maker prefers characteristics 1 to characteristics 2. (2) If the decision regarding choice set X were up to person 2, he would prefer that person 2 (he himself) gets characteristics 1 to the other alternatives. The same applies to person 3. (3a) The decision-maker prefers the alternative where $P2$ is the receiver of characteristic 1 to the alternative where he himself is the receiver of characteristic 1 or is indifferent between these two alternatives. Moreover, the decision-maker prefers the alternative where he himself is the receiver of characteristics 1 to the alternative where $P3$ is the receiver of characteristics 1. As a result, the decision-maker prefers $P2$ to $P3$ and is only willing to bear costs in order to choose the alternative whose characteristics are received by $P2$. (3b) The decision-maker prefers the alternative where $P2$ is the receiver of characteristic 1 to the alternative where he himself is the receiver of characteristic 1. Moreover, the decision-maker prefers the alternative where he himself

⁸ In other words, the decision-maker not only considers his isolated choice set but also how the other person would decide in the actual choice set and then, so as to attain a better outcome for the other person, chooses an alternative which deviates from his preferences regarding the isolated choice set. Now, it is possible that the alternative that the other person prefers in the actual choice set also depends on which alternative the decision-maker prefers in the actual choice set. In such a case, both the decision-maker and the other person depend their preferences regarding the actual choice set on the other's preferences regarding the actual choice set. So, the decision-maker needs to know the other person's preferences so as to form his preferences. But then again, the other person needs to know the decision-maker's preferences so as to form his preferences. This could go back and forth endlessly where no one ever attains a preference ordering concerning the actual choice set. In this dissertation, we exclude such cases.

is the receiver of characteristics 1 to the alternative where $P3$ is the receiver of characteristics 1 or is indifferent between these two alternatives. As a result, the decision-maker prefers $P2$ to $P3$ and is either only willing or more willing to bear costs in order to choose the alternative whose characteristics are received by $P2$. (3c) The decision-maker prefers the alternative where $P2$ or $P3$ is the receiver of characteristic 1 to the alternative where he himself is the receiver of characteristic 1. Moreover, the decision-maker prefers the alternative where $P2$ is the receiver of characteristics 1 to the alternative where $P3$ is the receiver of characteristics 1. As a result, the decision-maker prefers $P2$ to $P3$ and is more willing to bear costs in order to choose the alternative whose characteristics are received by $P2$ than by $P3$. In formal terms:

$$\begin{aligned}
 & x_1^{DM^\circ}, x_2^{DM^\circ} \in \mathbb{X} : u(x_1^{DM^\circ}) > u(x_2^{DM^\circ}) \\
 & \wedge x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : u_{P2}(x_1^{P2^\circ}) > u_{P2}(x_1^{DM^\circ}) = u_{P2}(x_1^{P3^\circ}) \\
 & \wedge x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : u_{P3}(x_1^{P3^\circ}) > u_{P3}(x_1^{DM^\circ}) = u_{P3}(x_1^{P2^\circ}) \\
 & \wedge x_1^{DM^\circ}, x_1^{P2^\circ}, x_1^{P3^\circ} \in X : \left[u(x_1^{P2^\circ}) \geq u(x_1^{DM^\circ}) > u(x_1^{P3^\circ}) \right] \\
 & \dot{\vee} \left[u(x_1^{P2^\circ}) > u(x_1^{DM^\circ}) \geq u(x_1^{P3^\circ}) \right] \dot{\vee} \left[u(x_1^{P2^\circ}) > u(x_1^{P3^\circ}) > u(x_1^{DM^\circ}) \right]
 \end{aligned}$$

As we see, strong taste-based discrimination in a receiver situation is a combination of agent-relativity and altruistic (and/or antisocial) preferences.⁹

After these technical definitions, let's discuss a study of Batson et al. (1981) that beautifully reveals strong taste-based discrimination. As part of an experiment, a student called Elaine had to perform a memory task. While she was doing so, participants had to observe her via a video control.¹⁰ It was said that the study is about the effect of aversive conditions on performance. This is why during the test Elaine got random electric shocks. These shocks certainly were uncomfortable but not dangerous. The experimenters told participants that Elaine does not know who is observing her and that they would not meet her in person.

⁹ In contrast, in a provider situation agent-relativity can be sufficient for strong taste-based discrimination, meaning that strong taste-based discriminators with egoistic preferences are possible in a provider situation.

¹⁰ All participants were female. Thus, the generalisability of the experiment is limited.

However, they concealed that the video control is actually a videotape and that Elaine is an actress who only acted like getting electric shocks.

Two further details about the experimental setup: (1) Participants were told that it was up to Elaine how many trials she wants to perform, with a minimum of two and a maximum of ten. Yet, regardless of how many trials Elaine does, every participant only had to observe two trials of her.¹¹ During the experiment, they learned that she agreed to do all ten trials. (2) Before the experiment began, subjects were split into two groups. One group was told that Elaine shared values and interests that were compatible with those they had stated in a previous questionnaire. The other group was told that Elaine shared values and interests that were incompatible with those they had stated in a previous questionnaire.

Now, as the experiment started, it was highly discernible that the electric shocks are very unpleasant to Elaine. Because of her strong reactions the experimenter interrupted after the second trial and got Elaine a glass of water. While she was gone, the observer had to complete a brief questionnaire regarding her impression on Elaine and whether seeing her suffering causes distress and/or concern. Then, the experimenter returned and Elaine explained why she responded so strongly to the shocks: As a child, she had a horse accident, where she fell onto an electric fence. This traumatic experience made her overly sensitive to electric shocks. The experimenter proposed to Elaine that she could quit the experiment. However, Elaine declined because she knew that the experiment was of great importance. Next, the experimenter hit upon another idea: The observer could continue for her. Being both relieved and reluctant, Elaine approved to check this option. Half a minute later, another experimenter stepped into the room of the observer and asked her if she is willing to take over for Elaine. In case of yes, she would have to complete the remaining eight sessions. In case of no, she only had to answer some questions about her impression on Elaine. After that she could leave. Of course, the experimenter stressed that there was no obligation to step in for Elaine. After the participant made her choice she again had to fill in some questionnaires (and did not get any electric shocks).

If we extract the choice sets given in this experiment and think about possible preference orderings on these choice sets, we attain the following setup. The decision-maker has two alternatives: Either she herself gets electro shocks or Elaine gets electro shocks. Moreover, there are two versions of Elaine: a likeable Elaine (E_+) and an unlikable Elaine (E_-). So, the alternatives have the

¹¹ In fact, there were two groups of participants: the “easy escapers” and the “difficult escapers”. In contrast to the easy escapers, who had to watch only two trials, the difficult escapers had to watch all sessions. However, for our purpose, it is sufficient to only consider the easy escapers, which is why we ignore the difficult escapers.

same characteristics 1 ($1 \in I$), where 1 = “receiver gets the remaining electro shocks”, but different receivers ($M = \{DM^\circ, E_+^\circ, E_-^\circ\}$), leading to the choice set $X = \{x_1^{DM^\circ}, x_1^{E_+^\circ}, x_1^{E_-^\circ}\}$. Of course, in a hypothetical isolated choice set $\mathbb{X}$ with alternatives $x_1^{DM^\circ}$ and $x_2^{DM^\circ}$, where 2 = “receiver does not get the remaining electro shocks”, the decision-maker prefers the latter. Moreover, if the decision regarding choice set X were up to the likeable or unlikeable Elaine, she would prefer that the decision-maker gets the remaining electro shocks. And although this is solely hypothetical, we further assume that the two versions of Elaine are indifferent between which Elaine gets the remaining electro shocks. Formally spoken:

$$\begin{aligned} x_1^{DM^\circ}, x_2^{DM^\circ} &\in \mathbb{X} : u(x_1^{DM^\circ}) < u(x_2^{DM^\circ}) \\ \wedge x_1^{DM^\circ}, x_1^{E_+^\circ}, x_1^{E_-^\circ} &\in X : u_{E_+}(x_1^{DM^\circ}) > u_{E_+}(x_1^{E_+^\circ}) = u_{E_+}(x_1^{E_-^\circ}) \\ \wedge x_1^{DM^\circ}, x_1^{E_+^\circ}, x_1^{E_-^\circ} &\in X : u_{E_-}(x_1^{DM^\circ}) > u_{E_-}(x_1^{E_+^\circ}) = u_{E_-}(x_1^{E_-^\circ}) \end{aligned}$$

Let’s get to the results so as to see the decision-makers preferences on getting electric shocks herself or giving them to Elaine.

Provided that participants had agent-neutral preferences, personal characteristics of Elaine should not have influenced their behaviour. So, let us compare the two conditions. In the dissimilar one, where Elaine’s values and interests were incompatible with those of participants, 18% took over for Elaine. In contrast, in the similar condition, 91% stepped in for her. This leads to two observations. First, in both conditions there were people who helped Elaine and thus behaved altruistically. Second, the degree of similarity between the decision-maker and the person in need was of utter importance for whether the latter received help or not, which implies agent-relative preferences. The combination of these two observations leads to strong taste-based discrimination. Thus, most participants had a preference ordering like the following one:

$$x_1^{DM^\circ}, x_1^{E_+^\circ}, x_1^{E_-^\circ} \in X : u(x_1^{E_-^\circ}) > u(x_1^{DM^\circ}) > u(x_1^{E_+^\circ})$$

It might be objected that participants have always exclusively made one decision, meaning they either had the likeable or unlikeable Elaine as a second possible

receiver and not both. Thus, there is no point of reference so as to assess whether their preferences truly are agent-relative. However, participants were randomly allocated to a condition. Therefore, the condition specific subsamples should be comparable and due to that serve as a reference point.

These outcomes are not very surprising anyway. We know from daily experiences that we do not treat everyone equally and thus that our preferences are not agent-neutral. For example, closeness to a person normally enhances the willingness to help. If a good friend asks you to assist him moving, you do so. But if a far relative communicates his moving date, you might pretend to be out of town that day. The same tendency is also observable in life-and-death issues. Even though there are people who donate one of their kidneys to a stranger, they represent less than 2% of all live donations. Mostly, a family member is the donor (Bernstein, 2017). Yet, we also differentiate between people that are equally unfamiliar to us. The lost-letter-technique provides a great method to show that. Milgram et al. (1965) placed letters in a city so it seemed as if someone had lost them. The authors examined how many letters were posted and whether the posting-rate depended on the address on the letter.¹² They used four different addresses: medical research associates, personal letter, friends of the Communist Party, and friends of the Nazi Party. Roughly three-fourths of the medical research associates and the personal letters returned. As opposed to this, only one out of four letters with friends of the Communist Party or the Nazi Party as the addresses came back. Thus, finders obviously made their behaviour conditional on the receiver. And this is not only true in case of political ideology but many other characteristics such as nationality or whether the receiver has a doctor's degree (Hellmann et al., 2015).¹³

Of course, the crucial question is why we prefer certain people to others and are mainly altruistic to these people (and even antisocial to the others). The concepts of ingroup favouritism and social identity theory shed light on it.

3.1.2 Ingroup Favouritism and Social Identity Theory

When we talk about groups, there are always two meta-categories that emerge (Turner et al., 1987). Either we ourselves (saliently) belong to the group as well,

¹² Of course, in actual fact the address on the letters was always the one of their labs. They only changed to whom (name or organisation) the letter was directed.

¹³ In section 3.2 we will discuss whether such behaviour truly involves taste-based discrimination or whether it actually is a form of statistical discrimination.

which defines our ingroup, or we do not belong to it, which constitutes our outgroup(s). This categorisation of others into ingroup and outgroup members highly affects preferences. There is vast evidence that people prefer their ingroup to their outgroups, leading to ingroup favouritism (see Balliet et al. (2014) for a meta-study). Therefore, in a provider situation, people often have preferences like the following one. Note that we denote the ingroup by $\mathcal{M}_{in}$ and the outgroup by $\mathcal{M}_{out}$, $A = \{\mathcal{M}_{in}, \mathcal{M}_{out}\}$, and assume that $\{m \in \mathcal{M}_{in} \cup \mathcal{M}_{out}\} = M$.

$$\exists x_i^{\mathcal{M}_{in}}, x_i^{\mathcal{M}_{out}} \in X : u(x_i^{\mathcal{M}_{in}}) > u(x_i^{\mathcal{M}_{out}})$$

In a receiver situation, we often have the following strong taste-based discriminatory preferences, where, for example, we can allocate money to different receivers. Note that $I = \{1, 2\}$, where 1 = “receiver gets money” and 2 = “receiver does not get money”. Additionally, although the decision-maker actually belongs to the ingroup as well, we exclude him from the ingroup and list him separately as $\mathcal{M}_{DM}$, so he becomes an individual receiver. Thus, $A = \{\mathcal{M}_{in}, \mathcal{M}_{out}, \mathcal{M}_{DM}\}$.¹⁴

$$\begin{aligned} & x_1^{\mathcal{M}_{DM}^\circ}, x_2^{\mathcal{M}_{DM}^\circ} \in \mathbb{X} : u(x_1^{\mathcal{M}_{DM}^\circ}) > u(x_2^{\mathcal{M}_{DM}^\circ}) \\ & \wedge x_1^{\mathcal{M}_{DM}^\circ}, x_1^{\mathcal{M}_{in}^\circ}, x_1^{\mathcal{M}_{out}^\circ} \in X : u_{\mathcal{M}_{in}}(x_1^{\mathcal{M}_{in}^\circ}) > u_{\mathcal{M}_{in}}(x_1^{\mathcal{M}_{DM}^\circ}) \geq u_{\mathcal{M}_{in}}(x_1^{\mathcal{M}_{out}^\circ}) \\ & \wedge x_1^{\mathcal{M}_{DM}^\circ}, x_1^{\mathcal{M}_{in}^\circ}, x_1^{\mathcal{M}_{out}^\circ} \in X : u_{\mathcal{M}_{out}}(x_1^{\mathcal{M}_{out}^\circ}) > u_{\mathcal{M}_{out}}(x_1^{\mathcal{M}_{DM}^\circ}) = u_{\mathcal{M}_{out}}(x_1^{\mathcal{M}_{in}^\circ}) \\ & \wedge x_1^{\mathcal{M}_{DM}^\circ}, x_1^{\mathcal{M}_{in}^\circ}, x_1^{\mathcal{M}_{out}^\circ} \in X : [u(x_1^{\mathcal{M}_{in}^\circ}) \geq u(x_1^{\mathcal{M}_{DM}^\circ}) > u(x_1^{\mathcal{M}_{out}^\circ})] \\ & \dot{\vee} [u(x_1^{\mathcal{M}_{in}^\circ}) > u(x_1^{\mathcal{M}_{DM}^\circ}) \geq u(x_1^{\mathcal{M}_{out}^\circ})] \dot{\vee} [u(x_1^{\mathcal{M}_{in}^\circ}) > u(x_1^{\mathcal{M}_{out}^\circ}) > u(x_1^{\mathcal{M}_{DM}^\circ})] \end{aligned}$$

We already find such preferences in case of young children. A study conducted by Fehr et al. (2008) revealed that 3–7-year-old children display more altruistic behaviour towards ingroup members than outgroup members in various economic games.¹⁵ Moreover, Jordan et al. (2014) let 6–8-year old children play a third-party punishment dictator game. This game proceeds like a normal dictator game except that the distribution is observed by a third person, who is equipped with

¹⁴ Although this is hypothetical, we suppose that the ingroup weakly prefers the alternative where the decision-maker receives characteristics 1 to that where the outgroup receives characteristics 1.

¹⁵ Appendix A which can be found in the electronic supplementary material introduces three common economic games that are employed in many of studies that we discuss in this and the following chapters; the dictator game, the ultimatum game, and the public goods game.

money as well. After the allocation, this person gets the chance to punish the dictator. Yet, punishment is costly.

Before we get to the results, we formalise the decision-situation because it differs from a situation where someone allocates money or electro shocks. Let's say that the decision-maker (*DM*) has to pay \$10 so as to take \$10 away from the dictator's (*DI*) endowment and in this way punish him. So, within a hypothetical isolated choice set $\mathbb{X}$, the decision-maker simply has two alternatives: lose \$10 ($x_1^{DM^\circ}$) or do not lose \$10 ($x_2^{DM^\circ}$). Yet, within the actual choice X , both the decision-maker and the dictator either lose or do not lose \$10, depending on the alternative. We assume that the dictator prefers not losing \$10 ($x_2^{DM^\circ, DI^\circ}$) to losing \$10 ($x_1^{DM^\circ, DI^\circ}$) and therefore $x_1^{DM^\circ, DI^\circ}$ is a punishment for him. Moreover, within an isolated choice set $\mathbb{X}$, the decision-maker also prefers $x_2^{DM^\circ}$ to $x_1^{DM^\circ}$. Yet, in a situation where receivers' outcomes are dependent, he might rather lose \$10 in order that the dictator loses \$10 too than do not lose \$10 but the dictator does also not lose \$10. Formally spoken:

$$\begin{aligned} x_1^{DM^\circ}, x_2^{DM^\circ} &\in \mathbb{X} : u(x_1^{DM^\circ}) < u(x_2^{DM^\circ}) \\ \wedge x_1^{DM^\circ, DI^\circ}, x_2^{DM^\circ, DI^\circ} &\in X : u_{DI}(x_1^{DM^\circ, DI^\circ}) < u_{DI}(x_2^{DM^\circ, DI^\circ}) \\ \wedge x_1^{DM^\circ, DI^\circ}, x_2^{DM^\circ, DI^\circ} &\in X : u(x_1^{DM^\circ, DI^\circ}) > u(x_2^{DM^\circ, DI^\circ}) \end{aligned}$$

If this is the case, the decision-maker displays antisocial behaviour because he is willing to bear costs so as to provide a disbenefit to the dictator.¹⁶ But whether the decision-maker truly behaves that way might depend on the group membership of the dictator, how fairly he behaved, and the group membership of the second player.

After this little parenthesis, let us continue with the results. Jordan et al. (2014) find that 6-year-old children punished selfishness more harshly when it negatively affected an ingroup member and when it came from an outgroup member. Meanwhile, 8-year old children did also punish egoistic outgroup dictators more harshly than egoistic ingroup dictators. But they did not differentiate between disadvantaged ingroup recipients and outgroup recipients. However, it would be wrong to

¹⁶ It could also be argued that this is actually prosocial behaviour because he punishes the unfair behaviour of someone else.

declare this change in behaviour from 6-year-old to 8-year-old children as universal. Bernhard et al. (2006) played the third-party punishment dictator game with two native groups of Papua New Guinea. They found the exact opposite of what Jordan et al. did in case of 8-year old children. On one hand, the third person punished selfishness less severely if the disadvantaged recipient was not in his group. On the other hand, punishers were indifferent to the group affiliation of the dictator. They punished dictators of each group equally harshly even though dictators expected that given the third person is in their group he punishes more leniently. So, there seems not to be a clear pattern for how people behave in third-party punishment dictator games. Nevertheless, ingroup favouritism is detectable in all three cases.

As previously mentioned, we are part of countless groups. From ethnic background to gender to profession to nationality to religion, our ingroup can be composed in various ways. In the experiment of Fehr et al. (2008) presented above, the children's ingroup was defined as being from the same playschool, kindergarten, or school. Consequently, participants that came from another playschool, kindergarten, or school formed the outgroup. Jordan et al. (2014) induced artificial groups as part of their experiment. The children were randomly assigned to either the "blue" or "yellow" team, which in turn constituted their ingroup and, in this way, also their outgroup. In the experiment of Bernhard et al. (2006), the indigenous tribes Wolimbka and Ngenika constituted the ingroup-outgroup context. Thus, all three experiments seem to have had clear group boundaries. But why did the children not perceive all participants as part of their ingroup? Why did the Wolimbka and Ngenika members form their ingroup and outgroup based on their tribes and not more generally on being from Papua New Guinea, which would have included both tribes? In other words, what does ultimately define which of our many group memberships is currently salient and thereby determines our perceived ingroup and the respective outgroups? And to put this into technical terms, what defines a decision-maker's set A ?¹⁷

The self-categorisation theory of Turner et al. (1987) provides an answer for that question. The theory says that self-categorisation can take place on different levels of abstraction, where a priori no level is more valid than another one. These levels can be narrowly defined such as me myself, a bit more general such as me a Swiss German or very broad such as me a human being. Which specific level and

¹⁷ For repetition, A is a subset of Ψ with the requirement that the elements of A are disjoint and their union leads to M . In turn, Ψ is the power set of M and M is the set of all individuals involved in a decision situation. So, A could have various manifestations as soon as M has more than one element. Yet, in a decision situation only one manifestation can be salient.

thereby group applies in a given situation depends on three components (Haslam et al., 2010).

(1) The comparative fit refers to the meta-contrast principle whose underlying assumption is as follows: Perceived stimuli are categorised in such a way that the differences between stimuli within a category are minimal whereas those between categories are maximal. The meta-contrast principle is then defined by the ratio of the averagely perceived differences between categories and the averagely perceived differences within a category.

$$\text{Meta-contrast principle:} = \frac{\varnothing \text{ perceived difference between categories}}{\varnothing \text{ perceived different within a category}}$$

The higher this ratio the more likely categorisation occurs along these categories. Moreover, if the ratio is smaller than one, there is no categorisation along these categories since there are bigger differences within than between categories. The meta-contrast principle can be illustrated through the following example: A Swiss is more likely to define himself as Swiss if he is interacting with a German than if he is interacting with another Swiss (Haslam et al., 2010).

(2) The normative fit implies that self-categorisation does not only need a meta-contrast ratio greater than one but also correspondence between the person's expectations of a category and its meta-contrast (ebd.). For example, a study conducted by Oakes et al. (1991) reveals that science students are more likely to be categorised as science students (and not simply students) if art and science students are perceived as holding different views about the value of science and these different views were compatible with stereotypic beliefs about the two groups.

(3) Ultimately, comparative fit and normative fit interact with perceiver readiness, also called accessibility. This means that a person does never execute a categorisation detached from all biographical background. He always does so in context of his beliefs, expectations, and motivations. In turn, these beliefs, expectations, and motivations are influenced by already existing salient group affiliations (Haslam et al., 2010).

We see that perceived similarity within a group and dissimilarity between groups is crucial for categorisation. These similarities and dissimilarities have to be compatible with our expectations of the categories. Consequently, it is not the objectively existing but subjectively perceived similarity between people that determines social categorisation. In turn, our subjective perception of similarity depends on prior and momentary expectations, beliefs, and motivations.

Although it is unclear whether group thinking played any role in the Elaine experiment of Batson et al. (1981) presented before, the results could at least

be explained by use of it. Subjects that were told that Elaine has similar views and interests as themselves often displayed altruistic behaviour towards her. The reason might be that in this case they perceived Elaine as “one of us”. So, Elaine benefited from ingroup directed altruism. However, when participants were told that Elaine has different views and interest she was perceived as “one of them” and as a result received help less frequently.

There are other experiments that reveal that a cue of similarity or relatedness can bolster altruism. For example, Krupp et al. (2008) let participants play a one-shot public goods game. While playing, subjects saw a photo of the face of the other players. These faces were either strangers or computer manipulated faces that resembled the participant.¹⁸ The results show that the more the faces of players in the group resembled the participant the more he contributed in the public goods game.

Pavey et al. (2011) manipulated subject’s level of relatedness, competence, or autonomy by use of different primes. (a) Participants had to solve a sentence unscrambling task, which in the relatedness condition contained words such as *community*, *together*, *connected*, or *relationship*. Additionally, they had to do a word completion task, where in the relatedness condition the words to be completed were *connect*, *relate*, and *share*. (b) Participants had to answer eight yes-or-no-questions. Given they answered with yes, they were asked to provide a short example. For instance, in the relatedness condition one of the questions was: “Have you ever felt a strong bond with someone you spend time with?” The results show that the relatedness-priming through the sentence unscrambling task and the word completion task led to higher interest in volunteering and intentions to volunteer relative to the other conditions. Moreover, relatedness manipulation participants also donated significantly more money to charity than did participants that were given a neutral task.¹⁹ Lastly, writing about relatedness experiences amplified feelings of connectedness to others, which in turn led to greater prosocial intentions. So, the authors infer that highlighting relatedness seems to increase altruistic behaviour (or at least altruistic behavioural intentions). This is all in line with self-categorisation theory and ingroup favouritism. As the similarity between us and “the others” is highlighted we rather categorise them as part of our ingroup and thereby act more prosocially towards them.

¹⁸ In the experiment, the computer manipulated faces that resemble the participant should serve as a cue for kinship. Why kinship is important for altruistic behaviour will be discussed in section 3.3.1.

¹⁹ In the experiment that led to this result, the authors implemented a relatedness-priming and a neutral task but no autonomy-priming and competence-priming.

A study by Levine et al. (2005) beautifully demonstrates how our momentarily salient ingroup can be manipulated. The authors conducted a study where subjects were self-identified supporters of the Manchester United Football Club. There were two experiments: One primed subjects to highly identify with their soccer club, the other with soccer in general. Regarding the procedure, the priming was induced at the beginning of the experiment by means of a questionnaire with open questions (e.g. “Why do you support Manchester United?” (Manchester United prime) or “When did you first become interested in soccer?” (general soccer prime)). Then, participants had to go to another room and as a consequence walk over the campus. There, a confederate ran past, fell, and held his ankle while screaming out of pain. The question of interest was whether the subject helps the runner or not. Both experiments had three conditions: (1) The jogger wore a plain shirt. (2) The jogger wore a Manchester United shirt. (3) The jogger wore a shirt of the FC Liverpool, Manchester United’s rivalry team. The results confirmed the hypotheses of the authors. One on hand, given participants were primed for Manchester United, 12 out of 13 helped the confederate in condition one but only 3 out of 10 in condition three. The latter is comparable to condition two where 4 out of 12 helped. On the other hand, if subjects were primed for soccer in general, 8 out of 10 helped the runner in condition one and 7 out of 10 in condition three. Both rates are substantially higher than in the second condition where solely 2 out of 9 helped. Consequently, something as small as a few open questions can decide whether you see the similarity between you and someone else (he is also a soccer fan) or the dissimilarity (he is a Liverpool fan). In turn, this evaluation strongly affects whether that other person receives our help or not.

So, up until now we know that people behave more altruistically towards fellow ingroup members than outgroup members and that comparative fit, normative fit, and perceiver readiness define our ingroup. Yet, why do we actually act more prosocially if it concerns someone from our ingroup compared to someone from our outgroup? The key concept to explain this question is social identity (Tajfel, 1970, 1974, 1982). Social identity is “that part of an individual’s self concept which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership” (Tajfel, 1974, p. 69). As we categorise the social world into ingroup and outgroup we automatically derive our social identity from the identified ingroup.

Social identity theorists have proposed two hypotheses for ingroup favouritism (Kite & Whitley, 2016). The first one is called the categorisation-competition hypothesis. It implies that categorisation itself leads to intergroup competition.

This is partly due to social biases.²⁰ For example, we perceive the outgroup as more homogenous, are more likely to attribute their achievements to chance and failures to their abilities, and given they are the minority overestimate their display of negative behaviour. Additionally, some cultures such as the Northern American one convey that relations between groups are naturally competitive. You should not trust the others because they try to get *our* resources (Insko & Schopler, 1987). Because of that, mere categorisation already rises feelings of competition and the desire to win. It is either us or them. Understandably, in such a situation you prefer us to them and as a result favour your own group so as to defend its (and your) interests.

The second hypothesis is called the self-esteem hypothesis. It contains the idea that we favour our ingroup because ultimately this increases our self-esteem. Social identity theory of Tajfel and Turner (1979, 1986) explains why this should be the case. Its first postulate is that people are motivated to uphold a positive self-identity. Second, our social identity is a part of our self-identity. Thus, the more positive our social identity is, the more positive our self-identity is. Third, through comparing our group status with the statuses of other groups we can evaluate how positive our social identity and thereby self-identity is. Now, if this comparison does not turn out advantageously, individuals can apply three main strategies. In case that group boundaries are permeable and/or our identification with the group is low, we escape, avoid, or deny belonging to the low-status group. This is called social mobility. Given group boundaries are not permeable and/or we identify strongly with that group, there are two different strategies, depending on whether the status hierarchy is stable or not. If it is stable, we can try to redefine the for the intergroup comparison relevant characteristics. This strategy has the name social creativity.²¹ If the status hierarchy is not stable, we can take action in order to change the standing of our group. This is called social competition and leads to ingroup favouritism because the more cohesion and cooperation a group displays the more likely it socially outcompetes others (Tajfel, 1982).²²

²⁰ We will discuss such social biases in section 4.1.2.

²¹ Let us exemplify the strategy of social creativity. A soccer team has lost a game, which, as a consequence, leads to a less positive social identity. Now, the players might say to themselves that they indeed scored only one goal whereas the opponent scored two but that their scored goal was more spectacular or that they have won more titles overall. By doing so, the relevant characteristic for intergroup comparison is no longer who has won the match but who has scored the more spectacular goal or has won more titles. In both cases the intergroup comparison turns out more advantageously.

²² In fact, social competition is not only a strategy of the low-status group to gain more status but also of the high-status group to maintain its status. Because as the low status group starts

One of the main social psychological findings that social identity theory aimed to explain was the so-called minimal group paradigm. It was inspired by a classic in social psychology. In the late 1950 s, early 1960 s, Sherif et al. (1961) conducted a number of field experiments that became to be known as the “Robbers Cave Experiment”. In a summer camp, Sherif randomly assigned 22 boys into two teams. The teams did not know about each other’s existence and were isolated for five days so as to form a group spirit. Then, the two teams had to compete in games where the winner was awarded with valued prizes. This led to massive hostility which interventions such intergroup contact (eating together) could not diminish. Not until the experimenters created scenarios with superordinate goals and thereby a positive interdependency between the groups, they started to cooperate. In the end, group boundaries almost disappeared entirely.

Now, five days of group binding activities seem to lead to strong ingroup favouritism. Tajfel (1970) wanted to know how much these group binding activities can be reduced that they still produce ingroup favouritism. In order to find that out he conducted a minimal group experiment. There are six requirements for a minimal group: (1) no face-to-face interaction; (2) complete anonymity of group membership; (3) no rational or instrumental link between the categorisation of the groups and the nature of the responses requested from the subjects; (4) all choosers should have the same choices regarding material payoffs; (5) competition between group motivation and some other motivation; and (6) the decision should be made as important as possible to the participant. For example, in Tajfel’s experiment, participants were assigned to one of two groups based on whether they preferred a painting of Kandinsky or Klee.²³ Astonishingly, even in these most minimal conditions categorisation affected individual behaviour and led to ingroup favouritism. In fact, participants did not choose the allocations that would simply maximise their ingroup outcome but the allocations that maximised the difference between groups. This phenomenon came to be known as the minimal group paradigm.

How does social identity theory explain these findings? Participants’ social identity is derived from the minimal group because the group-distributional choices make it salient. In such a situation, the Kandinsky or Klee lovers build the outgroup with which subjects compare themselves. Here, the only way to achieve

to compete with the high-status group the latter has (or wants) to defend its position, which produces ingroup favouritism.

²³ The groups were actually randomly set up in order to exclude that Kandinsky lovers and Klee lovers might have substantially different preferences and as a result the groups are not comparable.

a positive intergroup evaluation is through applying the social competition strategy. In this distributional competition, not the absolute payoff but the relative payoff is decisive, which is why subjects choose maximum group difference over maximum ingroup profit (Tajfel & Turner, 1979).²⁴

To summarise, the categorisation of the social world into ingroup and outgroup is reflected in our preferences. We are more altruistic within and concerned about our ingroup than outgroup, which is called ingroup favouritism. However, the ingroup is not at a static but both a dynamic and variable construct. According to the self-categorisation theory of Turner et al. (1987), comparative fit, normative fit, and perceiver readiness define our currently salient ingroup. These factors are situation-dependent. The salient ingroup yields our social identity. In turn, social identity is part of self-identity that we strive to perceive positively. Thus, we also strive to possess a positive social identity and have three strategies to achieve (or maintain) it: social mobility, social creativity, and social competition. The latter leads to ingroup favouritism. This human predisposition seems to be deeply rooted because it can even be observed in the most arbitrarily formed anonymous groups whose members neither had intragroup nor intergroup contact.

3.1.3 Ingroup Love or Outgroup Derogation?

The minimal group paradigm has been replicated several times in various kinds of economic games such as the prisoner dilemma (Ahmed, 2007), the dictator game (Chen & Li, 2009), or the public goods game (Kramer and Brewer, 1984; Brewer and Kramer, 1986). Moreover, at the beginning of the last chapter we discussed the experiment of Jordan et al. (2014). Here, by randomly and anonymously assigning children to either the “blue” or “yellow” team, the experimenters also set up a minimal group experiment. So, there is ample evidence for the phenomenon. However, the minimal group paradigm as described so far might lead to a wrong conclusion. Tajfel’s experiment seems to imply that people not only favour their ingroup but also disfavour their outgroup. Otherwise the participants would not have chosen the maximum group difference option but the maximum ingroup profit option. Yet, these minimal group experiments are often designed as zero-sum games, meaning the ingroup’s win is the outgroup’s loss and vice versa. So, by

²⁴ Not all social psychologists approve this explanation of the minimal group paradigm. The most prominent other explanation is given by the bounded generalised reciprocity model (Yamagishi et al., 1999; Yamagishi & Mifune, 2008). We will discuss it in section 3.2.2.

expressing ingroup favouritism you also automatically express outgroup hostility even if you are actually neutral towards the outgroup.

Why is this differentiation relevant for taste-based discrimination in the first place? It tells us how our tastes for groups actually look like. We said that strong taste-based discrimination is always constructed through a combination of agent-relativity and a certain type of social preferences. The last chapter has revealed that the ingroup and outgroup are the dominant dividing line regarding agent-relativity and thus that social identity influences taste-based discrimination. Now, in this chapter, we examine the second ingredient of taste-based discrimination, namely social preferences. In so doing, we ask whether it is primarily altruistic behaviour towards the ingroup (ingroup love), antisocial behaviour towards the outgroup (outgroup derogation), or both that give(s) rise to ingroup favouritism. We start with ingroup love.

Ingroup love involves the idea that people have a stronger desire to help ingroup members compared to the outgroup members because they care more about the well-being of ingroup than outgroup members (Everett et al., 2015). In other words, they gain more utility if they help ingroup compared to outgroup members. We can formulate this in four steps: (1) The decision-maker knows that both ingroup and outgroup members prefer characteristics 1 to characteristics 2. (2) He gains more utility if $\mathcal{M}_{in}$ receives 1 compared to if $\mathcal{M}_{in}$ receives 2. (3) He gains more or equivalent utility if $\mathcal{M}_{out}$ receives 1 compared to if $\mathcal{M}_{out}$ receives 2. (4) He gains more utility if $\mathcal{M}_{in}$ receives 1 compared to if $\mathcal{M}_{out}$ receives 1.

$$\begin{aligned}
 & x_1^{\mathcal{M}_{in}^\circ}, x_2^{\mathcal{M}_{in}^\circ} \in X : u_{\mathcal{M}_{in}}(x_1^{\mathcal{M}_{in}^\circ}) > u_{\mathcal{M}_{in}}(x_2^{\mathcal{M}_{in}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{out}^\circ}, x_2^{\mathcal{M}_{out}^\circ} \in X : u_{\mathcal{M}_{out}}(x_1^{\mathcal{M}_{out}^\circ}) > u_{\mathcal{M}_{out}}(x_2^{\mathcal{M}_{out}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{in}^\circ}, x_2^{\mathcal{M}_{in}^\circ} \in X : u(x_1^{\mathcal{M}_{in}^\circ}) > u(x_2^{\mathcal{M}_{in}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{out}^\circ}, x_2^{\mathcal{M}_{out}^\circ} \in X : u(x_1^{\mathcal{M}_{out}^\circ}) \geq u(x_2^{\mathcal{M}_{out}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{in}^\circ}, x_1^{\mathcal{M}_{out}^\circ} \in X : u(x_1^{\mathcal{M}_{in}^\circ}) > u(x_1^{\mathcal{M}_{out}^\circ})
 \end{aligned}$$

As a consequence, if the decision-maker also has altruistic preferences, he gains more utility if he acts in a way that is costly for himself but provides a benefit

to $\mathcal{M}_{in}$ compared to if he acts in a way that is costly for himself but provides a benefit to $\mathcal{M}_{out}$.

An explanation for such preferences provides a phenomenon that Brewer (1999) calls depersonalisation. It implies that through categorisation of and identification with the ingroup the individual partly loses his own identity and adopts the identity of the group.²⁵ Through that process, his interests adjust themselves to the group's interests and thereby helping himself becomes equivalent to helping the group. Kramer and Brewer (1984) describe the effects of social identification as follows: "[Actors] attach greater weight to collective outcomes than they do to individual outcomes alone. Inclusion within a common social boundary reduces social distance among group members, making it less likely that individuals will make sharp distinctions between their own and others' welfare." (p. 1045) A minimal group experiment by Simpson (2006) where participants were exposed to a prisoner's dilemma confirms this view. The results reveal that not alterations in how participants expected their fellow ingroup members to act were responsible for ingroup favouritism but how they weighted the payoffs of fellow ingroup members.

Given group identification really leads to depersonalisation which in turn leads to ingroup favouritism, the more someone identifies with his group the more he should put the group's well-being before his own.²⁶ A study conducted by de Cremer (2002) shows exactly that. In order to manipulate group identification, he let participants fill out a small personality test that categorised them as either Type O or Type P personality. The Type P personality was positively connoted and described as caring, honest, consistent, confident, and more socially skilled. In comparison, the Type O personality was less positively connoted so as to make it desirable to be a Type P personality. Half of the participants were told that their responses placed them just inside the Type P category. The other half was told that their answers were clear examples of a Type P personality. While the former should lead to low group identification the latter should induce high group identification.²⁷ Then, participants had to play a public goods game where all other players were said to be Type P personalities. Here, the high identifiers were generally more cooperative than the low identifiers. De Cremer infers that "[c]ore group members [the high identifiers] ... seem to have incorporated the

²⁵ It says partly here because we know from optimal distinctiveness theory of Brewer (2012) that people normally seek both inclusion and differentiation within the ingroup.

²⁶ More precisely, the group's well-being becomes his own well-being. Thus, the two are actually no longer separable.

²⁷ There was also a manipulation check that asked how typical of their group participants perceived themselves to be and to what extent they felt they belonged to this group.

group as an important aspect of one's self" (p. 1339). Therefore, group identification appears to have led to depersonalisation, which in turn generated ingroup directed altruism.

Van Vugt and Hart (2004) confirm this argument. They used a public goods game in order to examine cooperative behaviour. Group identification was manipulated as follows: Half of the participants were told that the study examines how well students from different universities would perform individually in the game. The other half was told that it investigates how well groups of students from different universities would perform in the game.²⁸ The authors find that the more participants identified with their public goods game group, the more altruistically they behaved in the game. Additionally, high identifiers also made less use of an attractive exit option that would have increased their personal outcome. Van Vugt and Hart conclude that high identifiers' group loyalty emerged due to an extremely positive impression of their group affiliation and thus, social identity seems to have acted as a social glue.

Let's continue with the empathy-altruism hypothesis of Batson (2015).²⁹ It says that empathy (more precisely empathic concern) leads to other-oriented motivation and thereby altruism. Thus, altruistic behaviour could be explained by empathy-based social preferences, where the awareness of another person's need arouses empathy, which in turn raises altruistic motivation (Everett et al., 2015). For example, a study conducted by Rumble et al. (2010) demonstrates that empathy is able to sustain cooperation in a public goods game. The reason for this is that empathy reduces "the detrimental effects of 'negative noise,' or unintended incidents of non-cooperation". (p. 856) Moreover, participants that were induced to feel empathy in a prisoner's dilemma behaved more cooperatively than a control group (Batson & Moran, 1999). This is even true when subjects knew that their co-player had already made a competitive choice (Batson & Ahmad, 2001). Consequently, empathy seems to be an important part of social preferences. However, regarding agent-relativity, the question of course is whether we feel the same amount of empathy for every person in a needy situation.

²⁸ The design of the study is a bit problematic because it might have led to desirability. Participants who were told that the study examines individual (group) performance might have behaved more egoistically (altruistically) to approve the authors' hypothesis that participants anticipated.

²⁹ In this dissertation, we understand empathy as "an affective reaction caused by, and congruent with, another person's inferred or forecasted emotions: that is, feeling good in response to someone experiencing a positive event (e.g., when Emile wins an award), and feeling bad in response to someone experiencing a negative event (e.g., when Rebecca's paper is rejected)" (Cikara et al., 2014, p. 111).

Apparently, the answer is no. According to Cikara et al. (2014), humans have a predisposition called the intergroup empathy bias. It implies that we tend to empathise more with ingroup than with outgroup members. Several neuroscientific studies have found that people display more neural activation in pain and empathy circuits (especially the insula) given they observe an ingroup compared to an outgroup member being in pain (Cheon et al., 2011; Chiao & Mathur, 2010; Gutsell & Inzlicht, 2010, 2012; Xu et al., 2009). Thus, these findings are compatible with the idea that through identifying with a group, other ingroup members' interests become our interests as well (at least to a certain degree). In turn, having these neural activations serves as a predictor for ingroup favouritism on a behavioural level (Mathur et al., 2010). A study by Hein et al. (2010) nicely demonstrates this. The authors took soccer fans so as to induce an ingroup and an outgroup. Subjects either witnessed a fan of their favourite team (ingroup) or their rival team (outgroup) suffering pain. Then, they could choose whether or not they wanted to relieve the person in pain through enduring physical pain themselves. Regarding the ingroup, helping behaviour was forecasted best by anterior insula activity and self-reports of empathic concern. This suggests that participants were empathising with the fellow ingroup member in need and thus helped. Contrary to that, if an outgroup member was suffering pain, non-helping behaviour was predicted best by nucleus accumbens (NAcc) activity and how negative the outgroup member was evaluated.³⁰ To conclude, "empathy-related insula activation can motivate costly helping, whereas an antagonistic signal in nucleus accumbens reduces the propensity to help." (p. 149) As we have seen, the activation of these two brain areas depends on the group membership of the person in need.

To summarise the connection between social preferences and ingroup love, group identification leads to depersonalisation, meaning that we adjust our interests to the groups' interests. Because of that our utility is (partly) derived from our fellow ingroup members' (and not outgroup members') utility which inevitably leads to ingroup favouritism. Empathy seems to be an important mediator of this whole process.

Let us continue with how outgroup derogation affects social preferences.³¹ Here, it is not the pleasure of the ingroup but the displeasure of the outgroup that provides individuals utility. At the beginning of section 3.1.2, we discussed

³⁰ In fact, brain signals predicted helping behaviour more accurately than what people said (Singer, 2015).

³¹ Sometimes, outgroup derogation is also called outgroup hate. The terms can be used interchangeably.

that, in a third-party punishment dictator game, participants punish other (especially selfish) players even if punishment is costly and has no strategic value (Bernhard et al., 2006; Jordan et al., 2014). Moreover, Anderson and Putterman (2006) reveal that the level of punishment depends on how expensive punishing is and how egoistically the person to be punished behaved. This suggest that the act of punishment and thereby retaliation gives utility to the punisher. Otherwise it is unclear why someone would pay for it.

If in certain situations the disutility of others increases our utility, an explanation for ingroup favouritism is that people gain more utility by the disutility of outgroup members than by the disutility of ingroup members. We can formulate this in four steps and exclude the possibility of ingroup love³²: (1) The decision-maker knows that both ingroup and outgroup members prefer characteristics 1 to characteristics 2. (2) He gains equivalent or less utility if $\mathcal{M}_{in}$ receives 1 compared to if $\mathcal{M}_{in}$ receives 2. (3) He gains less utility if $\mathcal{M}_{out}$ receives 1 compared to if $\mathcal{M}_{out}$ receives 2. (4) He gains less disutility if $\mathcal{M}_{in}$ receives 1 compared to if $\mathcal{M}_{out}$ receives 1.

$$\begin{aligned}
 & x_1^{\mathcal{M}_{in}^\circ}, x_2^{\mathcal{M}_{in}^\circ} \in X : u_{\mathcal{M}_{in}}(x_1^{\mathcal{M}_{in}^\circ}) > u_{\mathcal{M}_{in}}(x_2^{\mathcal{M}_{in}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{out}^\circ}, x_2^{\mathcal{M}_{out}^\circ} \in X : u_{\mathcal{M}_{out}}(x_1^{\mathcal{M}_{out}^\circ}) > u_{\mathcal{M}_{out}}(x_2^{\mathcal{M}_{out}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{in}^\circ}, x_2^{\mathcal{M}_{in}^\circ} \in X : u(x_1^{\mathcal{M}_{in}^\circ}) \leq u(x_2^{\mathcal{M}_{in}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{out}^\circ}, x_2^{\mathcal{M}_{out}^\circ} \in X : u(x_1^{\mathcal{M}_{out}^\circ}) < u(x_2^{\mathcal{M}_{out}^\circ}) \\
 & \wedge x_1^{\mathcal{M}_{in}^\circ}, x_1^{\mathcal{M}_{out}^\circ} \in X : u(x_1^{\mathcal{M}_{in}^\circ}) > u(x_1^{\mathcal{M}_{out}^\circ})
 \end{aligned}$$

As a consequence, if the decision-maker also has antisocial preferences, he gains more utility if he acts in a way that is costly for himself but provides a disbenefit to $\mathcal{M}_{out}$ compared to if he acts in a way that is costly for himself but provides a disbenefit to $\mathcal{M}_{in}$.

The reason behind this explanation can again be found in the concept of empathy. So far, we have only discussed half of the intergroup empathy bias.

³² But of course, it is also possible that a decision-maker shows both ingroup love and outgroup derogation.

We do not only exhibit more empathy for ingroup members but also counter-empathy for outgroup members. Thus, we experience *schadenfreude* because of the outgroup's adversities whereas their triumphs give us displeasure, called *glückschmerz* (Leach et al., 2003; Smith et al., 2009a; Cikara et al., 2011). This phenomenon is independent of ingroup love.³³ Cikara et al. (2014) found that the intergroup empathy bias also persisted after one's ingroup had defeated their outgroup competitors. Only by giving subjects cues that reduces group entitativity, the intergroup empathy bias could be attenuated. As a consequence, the authors infer that the intergroup empathy bias is (mainly) driven by outgroup antipathy and not extraordinary ingroup empathy.

However, there is other evidence which claims that not outgroup derogation but ingroup love is the more potent driver for ingroup favouritism. A game designed by Halevy et al. (2008) called the "intergroup prisoner's dilemma—maximizing difference" should enable to detect the motivation behind self-sacrificial behaviour in an intergroup situation. Implementing this game in a minimal group experiment, Halevy et al. (2012) concluded that it is not the aggressive drive to hurt the outgroup but the altruistic desire to help the ingroup which produces the minimal group paradigm. Moreover, Gaertner et al. (2006) show that group formation can occur without an outgroup, only by intra-aggregate factors that promote entitativity. The group affiliation that emerged from that increased cooperative behaviour in a prisoner's dilemma although there was no outgroup that would have enabled an intergroup comparison. Finally, in their meta-analytic analyses of 212 intergroup cooperation studies, Balliet et al. (2014) conclude that "intergroup discrimination in cooperation is the result of ingroup favoritism rather than outgroup derogation". (p. 1556)

In conclusion, even though outgroup derogation certainly plays a role in ingroup favouritism, it seems not to be as important as ingroup love. Or to put it differently, our preferences for positive ingroup outcomes are more pronounced than our preferences for negative outgroup outcomes. Therefore, our taste for the ingroup particularly stems from the willingness to support the ingroup and not the willingness to hurt the outgroup.

³³ So, it is not like in a zero-sum game where the expression of ingroup love cannot be distinguished from outgroup derogation.

3.1.4 Tastes Outside the Ingroup-Outgroup Context

Social identity theory is the most prominent theory so as to describe intergroup behaviour and, from this perspective, commonly applied on the topic of discrimination (Kite & Whitley, 2016). Yet, do our tastes always have to stem from an ingroup-outgroup context which is necessary for social identity theory to be applicable in the first place?

Let us look at the example of reciprocal social preferences which consider the fairness of other agents' actions (Everett et al., 2015). They imply that if someone treated you (or someone else) nicely, you treat him nicely in return. This is called positive reciprocity. For instance, Fischbacher et al. (2001) have found such preferences in a public goods game. Here, 50% were conditional cooperators, meaning that they did only cooperate if others cooperated as well. Additionally, there is also negative reciprocity which involves that if someone treated you (or someone else) badly, you treat him badly in return. Such behavioural patterns could be seen in case of the public goods game with a punishment option. Here, some players reciprocated the uncooperative behaviour of other players through punishing them (Fehr & Gächter, 2002). So, regardless of an ingroup-outgroup context, many people have a taste for those who behave fairly and distaste for those who behave unfairly.

It is important to notice that such reciprocal behaviour is not strategic. So, you do not return a favour because you expect that the beneficiary or someone else will again return your favour in the future. Or you do not punish another player in a public goods game because you expect that this punishment will pay off later. If that were the case, we would speak of weak reciprocity. Yet, reciprocal social preferences require strong reciprocity which imply that "people willingly repay gifts and punish violation of cooperation and fairness norms even in anonymous one-shot encounters with genetically unrelated strangers" (Fehr & Henrich, 2004, p. 55). So, unlike weak reciprocity, strong reciprocity excludes that behaviour is (solely) driven by strategic egoism (Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006). Finally, reciprocal social preferences are not limited on how someone actually behaves but can also take into account the intentions behind that behaviour (Falk et al., 2003). For example, Guroglu et al. (2011) let participants play an ultimatum game where some proposers were forced to make a rather unfair offer. The authors found that in such cases recipients were more likely to accept an unfair allocation compared to when proposers had deliberately chosen it.

Although reciprocal social preferences can be completely detached from social identity, there is evidence indicating that the two also interact. Boldizar and

Messick (1988) found that group membership of actors influences the fairness evaluation of their behaviour: While ratings of ingroup actors were fairer than those of outgroup actors if the performed behaviour was fair, this was precisely vice versa if the performed behaviour was unfair (which came as a surprise to the authors).³⁴ Moreover, Chen and Li (2009) implement a response game so as to examine how people reciprocate fair/unfair behaviour in a dictator game setting. First, the authors found that participants were 19% more likely to respond altruistically to a player that treated them prosocially if he was an ingroup relative to an outgroup member. Second, given that a player behaved unfairly, participants were 13% less likely to punish that player if he was part of the ingroup and not the outgroup.³⁵ Thus, it seems that after all reciprocal social preferences are still affected by ingroup-outgroup categorisation.

Let us continue with a different phenomenon that can also lead to taste-based discrimination despite the absence of an ingroup-outgroup context, namely disgust. Disgust is commonly defined as the rejection of unpleasant stimuli based on smell, sight, or even mere thought (Kiss et al., 2018). Its elicitors can stem from various sources. Kiss et al. name five disgust domains that have been identified: (1) core; (2) animal-reminder; (3) interpersonal; (4) moral; and (5) sexual.³⁶ So, while rotten food and eczemas can evoke disgust, which then would be called core disgust, this is also possible in case of violations of social and moral boundaries, which then would be called moral disgust.

We first consider a group which elicits mainly core disgust, meaning disgust that functions as a protective mechanism against potential sickness: ill people. In case of ill people, the purpose of disgust is not far-fetched. Since many pathogens are communicated via inter-personal contact it can be adaptive to avoid such people so as not to get contaminated (Schaller et al., 2003). So, disgust serves as a disease-avoidance mechanism that makes us distance ourselves from ill people (Oaten et al., 2009). In order to detect the presence of disease in others we

³⁴ The reason for this rather surprising finding might be that participants distanced themselves from the unfairly behaving ingroup member through declaring his behaviour as particularly unfair. In turn, this helped them to uphold a positive self-identity. Boldizar and Messick (1988) write: “[T]he confrontation with an ingroup member performing an unfair behaviour may have induced a feeling of dissimilarity to the ingroup. This lack of identification with the ingroup, coupled with a lack of opportunities to increase favorable outcomes for the ingroup, may have minimized the effects of an ingroup favoritism bias.” (p. 108)

³⁵ Yet, as we will see in section 3.2.2, the pattern regarding punishment of unfair ingroup and outgroup behaviour is more complex/unclear.

³⁶ Of these five domains of disgust, only interpersonal disgust is directly linked to an ingroup-outgroup context.

may rely on heuristic signals, such as coughing, behavioural tics, spasms, and skin lesions. For instance, individuals afflicted with illnesses that affect the skin, such as leprosy, were often segregated from the community (Plageron, 2005). Yet, disgust as a disease-avoidance mechanism appears to be overinclusive and can be activated even if we know that a disease is non-contagious or actually not a disease in the first place (Oaten et al., 2009). For example, disgust as a disease-avoidance mechanism has also been observed in case of cancer (Greene & Banerjee, 2006), mental illness (Stier & Hinshaw, 2007), physical disability (Park et al., 2003), or obesity (Harvey et al., 2002). Finally, disgust sensitivity also influences our attitude towards such groups, leading to distastes for them (Oaten et al., 2009; Lieberman et al., 2012).

Next, let us get to a group that can not only elicit core disgust but also other domains of disgust such as moral disgust: homosexuals and in particular gay men. Kiss et al. (2018) mention mainly two reasons why some people are morally disgusted by gay men. On one hand, gay men destabilise the idea of heteronormativity, which means that heterosexuality is not simply a sexual orientation but, rather, a socially agreed-upon and normalised set of behaviours (Jackson, 2006). In this connection, gay men are for example accused to infiltrate “heterosexual institutions” such as marriage. On the other hand, several religions forbid homosexuality and describe it as impure. “[C]oncepts such as purity and symbolic cleansing (e.g., baptism, mikven) play an important role in most popular religions (Terrizzi et al., 2012). Purity and sanctity also are crucial elements of moral disgust. Religious beliefs frequently frame gay men as abnormal and depraved and, thus, devoid of sanctimony (Devos et al., 2002; Helminiak, 2008).” (Kiss et al., 2018, p. 7)

Now, as in case of ill people, disgust also influences the attitude towards and thereby promotes a distaste for gay people. Kiss et al. (2018) conducted a meta-analytic review of 17 studies that investigated the relationship between disgust and homonegativity. There are two main results: (1) There is a moderate to large effect of disgust sensitivity on homonegativity; (2) There is a large effect of disgust induction, as for example via using a fecal odor, on homonegativity.

The distaste for homosexuals and in particular gay men brings us to another kind of social preferences that is (at least not directly) triggered by an ingroup-outgroup context, namely type-dependent preferences. Fehr and Schmidt (2006) define type-dependent preferences as follows: “According to type-based reciprocity, an individual behaves kindly towards a “good” person (i.e. a person with kind or altruistic preferences) and hostilely towards a “bad” person (i.e. a person with unkind or spiteful preferences).” Such preferences could be compatible with a distaste for homosexuals because perceived morality plays an important role

regarding whether we evaluate someone as good or bad (Everett et al., 2015). For example, Brambilla et al. (2013) found that participants reported less desire to interact with others that were said to lack moral qualities compared to those that were said to be highly moral. Importantly, this finding was independent of whether the potential counterpart was an ingroup or an outgroup member.³⁷ Therefore, in respect to some people such as religious fundamentalists, homosexuality elicits, among others, moral disgust which should lead to the evaluation that homosexuals are immoral and thus bad (Morrison et al., 2019).³⁸ In turn, due to type-based social preferences, these apparently immoral people are then treated worse than those they perceive to be moral.

However, although perceived morality can breach ingroup favouritism as Brambilla et al. (2013) have shown, often the two go together. According to Brewer (1999), groups believe in their own moral superiority. She writes: “To the extent that all groups discriminate between intragroup social behavior and intergroup behavior, it is in a sense universally true that “we” are more peaceful, trustworthy, friendly, and honest than “they”.” (p. 435) Similarly, disgust is often mentioned to be important in an ingroup-outgroup context as well. For example, Cottrell and Neuberg (2005) state that outgroups which threaten an ingroup’s values primarily evoke disgust (and to a lesser extent also fear and anger). Moreover, disgust sensitivity predicts negative outgroup evaluations and discriminatory resource allocations (Hodson & Costello, 2007; Hodson et al., 2013). Thus, while disgust (and in particular core disgust) can promote distastes for certain groups despite the absence of an ingroup-outgroup context, it also does so within an ingroup-outgroup context. Likewise, while type-based preferences do not have to be influenced by ingroup-outgroup categorisation, social identity still seems to be important within such preferences (Everett et al., 2015).

Let us finish this chapter with a taste that is independent of an ingroup-outgroup context and neither linked to fairness, nor disgust, nor morality. Imagine someone who has a cat allergy. Due to that allergy he prefers situations where he does not come in contact with cats to situations where he does come in contact with cats. In other words, we could say that the individual has a “distaste for coming in contact with cats” and thus is a non-social discriminator. Now, when invited for dinner, he always asks whether the hosts have a cat and only accepts

³⁷ Yet, the authors state that the effect of morality on behavioural intentions was mediated by perceived group image threat for ingroup counterparts and safety threat for outgroup counterparts.

³⁸ This is independent of whether they behave fairly or not (so as to delimitate reciprocal social preferences) since sexual orientation has nothing to do with fairness (cf. Brambilla & Leach, 2014).

if they do not. Therefore, the individual categorises people into cat owners and non-cat owners and by always rejecting invitations of the former seems to show a distaste for them. But is this truly a distaste for the group of cat owners? Not really, because if cat owners would invite him to a restaurant where no cats are present, he would happily accept. So, his apparent distaste for cat owners solely stems from his distaste for coming in contact with cats. And given that cat owners provide the same characteristics as non-cat owners, such as going out for dinner at a restaurant without cats, he does no longer differentiate between cat owners and non-cat owners. Likewise, in a dictator game, where there is no potential contact with cats anyway, he would also not treat cat owners and non-cat owners differently.³⁹

However, what if an individual does not want to come in contact with a group itself? For example, let's assume an individual avoids physical contact with everything that is contagious such as contagious objects, contagious animals, and also contagious people. In such a case, the individual would have a distaste for contagious people. This is because the group of contagious people is defined by their contagiousness and this is precisely what he wants to avoid. But then again, if this distaste for contagious people is restricted to avoidance of physical contact with that group, contagious and non-contagious people should be treated equally in non-contact situations. For instance, he should not *prima facie* prefer a book written by a non-contagious person to a book written by a contagious person. Similarly, he should not give non-contagious people more money in a dictator game than contagious people.⁴⁰

All in all, this chapter tried to demonstrate that not all tastes have to stem from an ingroup-outgroup context: For example, we have tastes for fair people and for good/moral people as well as distastes for people who make us feel disgusted and people who we perceive as a threat. Importantly, this list does not claim to be comprehensive and there certainly are more such sources.⁴¹ Yet, despite the fact that tastes can also stem from a non-ingroup-outgroup context, such tastes are often still intertwined with social identity (Cottrell & Neuberg, 2005; Hodson et al., 2013; Everett et al., 2015; Boldizar & Messick, 1988; Chen & Li, 2009).

³⁹ Of course, it is also possible that an individual with a cat allergy truly develops a distaste for cat owners and also treats them differently than non-cat owners even if they provide/receive the same characteristics. Yet, this does not have to be the case.

⁴⁰ Yet, as this chapter has shown, contagious people often elicit disgust which in turn promotes a general distaste for them.

⁴¹ For instance, not only disgust but also fear can promote distastes (Cortell & Neuberg, 2005). We will discuss its relevance in section 3.3.3, when we present the anxiety about the unknown.

This is why this dissertation primarily discusses taste-based discrimination from an ingroup-outgroup context.

To summarise the whole section 3.1, the categorisation in ingroup and outgroup frequently defines the dividing line between whom we treat more favourably and who we treat less favourably. Thereby, the precise manifestation of the salient ingroup is changeable. Social identity theory provides an explanation for ingroup favouritism: We partly derive our self-identity from our social identity and therefore the groups we are part of. This leads to ingroup love and outgroup derogation because it boosts a positive social identity, whereby ingroup love is more prevalent than outgroup derogation. Ultimately, tastes can also stem from a non-ingroup-outgroup context. Yet, as it seems, such apparently “non-ingroup-outgroup context-based tastes” are nevertheless often connected to social identity.

3.2 Is All Discrimination Ultimately Statistical Discrimination?

Let’s resume an example that we have already used once. It consists of two statements: (1) If a good friend asks you to assist him moving, you do so. (2) If a far relative communicates his moving date, you pretend to be out of town that day. We assumed that this is a demonstration of strong taste-based discrimination. You bear costs (e.g. in form of time) when you help someone to move and provide a benefit to the moving person. Therefore, if you help someone to move, you must have social preferences. Then, you only help your close friend but not your far relative which indicates agent-relativity. Both together lead to strong taste-based discrimination. However, what if we also had the following information: (1) Among your close friends, there is the informal rule that you help each other move. (2) Someone who offends this rule cannot expect that he receives help in case of a future move. (3) There is no such rule among far relatives. (4) You yourself plan to move soon and hope that others will help you. Considering this additional information, is your willingness to help your friend move still altruistic or simply strategic because you do not want to lose your friends’ manpower when you move at some point in the future?

We see that in such a situation, the identity of the receiver of an alternative’s characteristics can influence these characteristics. Let’s say all alternatives have the same characteristics i , which is “help receiver move”. As we have just learned, these characteristics i probably have different consequences or different probabilities on consequences if the receiver is a close friend (CF°) or a far relative

(FR°). Therefore, if a decision-maker prefers $x_i^{CF^\circ}$ to $x_i^{FR^\circ}$, this does not have to imply that he is a taste-based discriminator. He could also simply be a statistical discriminator in a situation of uncertainty and actually prefer $f_i^{CF^\circ}$ to $f_i^{FR^\circ}$.⁴² The uncertain part of the decision situation is that he does not know the (subsequent) consequences of his actions for sure.⁴³ Maybe his friends are generous and still help him when he moves at some point in the future. Maybe his far relative will be disappointed and never invites him to his new mansion, which would be quite a loss for the decision-maker. The fact is that we do not know the objective probabilities of these scenarios and thus, among others, use group (or individual) specific beliefs so as to form predictions about them.

If we develop these deliberations further, we could even form the hypothesis that all what seems to be taste-based discrimination actually is statistical discrimination. If that were true, ingroup favouritism would not be an expression of a taste for the ingroup but a strategic way to behave in for an egoistic decision-maker.⁴⁴ Regarding economic games, there is plenty of research which demonstrates that what on first sight looks like ingroup favouritism becomes strategic egoism on a second sight. Following the classification of Everett et al. (2015), we examine three areas in this chapter where ingroup favouritism can function as an expected utility maximising belief of a decision-maker with egoistic preferences: interdependence of outcomes and direct reciprocity, indirect reciprocity and reputational concerns, and cooperative norm violation.

3.2.1 Interdependence of Outcomes and Direct Reciprocity

The first ingroup favouring belief suggests that results of distributional games, which imply ingroup favouring social preferences, can be explained by perceived

⁴² For repetition, the littler star (*) next to the i indicates that the alternatives' characteristics are influenced by the receivers' identity and thus that there is statistical discrimination.

⁴³ In contrast, if you perfectly know the subsequent consequences of two alternatives (more precisely the objective probabilities of the scenarios they involve) that have the same immediate but different subsequent consequences, the two alternatives do not have equivalent characteristics in the first place.

⁴⁴ It is also possible that what seems to be ingroup favouritism is actually statistical discrimination of a decision-maker with social and agent-neutral preferences: You help your ingroup because from a statistical perspective they benefit more from your help than the outgroup. Yet, as section 3.3 will reveal, such unconditional social preferences are difficult to explain from an evolutionary biological perspective. Moreover, no paper could be found that pursues this approach to ingroup favouritism, which is why we neglect it.

outcome interdependence and expectations of reciprocity. Rabbie et al. (1989) stated an early critique on the interpretation of Tajfel and his colleagues regarding their minimal group experiments (Tajfel et al., 1971; Tajfel & Turner, 1979). They argued that instead of ingroup favouring social preferences, the allocations within these experiments were grounded on beliefs about outcome interdependence. So, participants (at least implicitly) thought that their own outcome depends on their choices. In the words of Rabbie et al. (1989): “[A]lthough subjects in the standard MGP [minimal group paradigm] cannot directly allocate money to themselves, they [think that they] can do it indirectly, on their reasonable assumption that the other ingroup members will do the same to them. By giving more to their ingroup members than to the outgroup members—in the expectation that the other ingroup member will reciprocate this implicit cooperative interaction—they will increase their chances of maximizing their own outcomes.” (p. 176)

Locksley et al. (1980) provide evidence for this hypothesis. The first two experiments of their paper showed that social categorisation via a lottery procedure produced ingroup favouring allocation. However, the second two experiments revealed that ingroup favouritism could be extinguished by means of the following condition: Subjects were told that neither their fellow ingroup members nor outgroup members depend their allocations on group membership. Given participants really had had ingroup favouring social preferences this condition should not have affected their allocation. Yet, it did. Therefore, beliefs about how other group members would behave were obviously of great importance. In the experiments of Locksley et al. (1980), subjects apparently believed that their outcome was more strongly dependent on their fellow ingroup members because ingroup members are more likely to reciprocate their behaviour. This and not ingroup favouritism is the reason why they favoured the ingroup in their allocations. And as soon as a condition eliminates this belief, it also eliminates ingroup favouritism. Rabbie et al. (1989) call this the reciprocity hypothesis.

There are two versions of this theory: the unbounded reciprocity hypothesis and bounded reciprocity hypothesis (Everett et al., 2015). The former implies that group membership per se is irrelevant for the allocation. You simply allocate more resources to those you think your outcome is dependent on, anticipating that they reciprocate this favourable treatment. Our default belief might be that ingroup members are those on which our outcome more heavily depends. However, if we learned that our outcome more heavily depends on the outgroup, we would treat the outgroup more favourably than the ingroup. So, unlike outcome interdependence, group membership only serves as a proxy and has not a moderating effect itself. This is different in case of the bounded reciprocity hypothesis. Here, our beliefs about reciprocity are not only affected by perceived outcome

interdependence but also group membership. To put it differently, social categorisation bounds our expectations of reciprocity. This might be because repeated interactions with ingroup members are more likely than with outgroup members (ebd). In turn, repeated interactions increase the chances of a beneficial reciprocal relationship. Outcome interdependency cannot (totally) overrule this effect. So, even if participants know that their outcome depends on the outgroup, they still do not treat outgroup members better than ingroup members (Gaertner & Insko, 2000).

Stroebe et al. (2005) tested whether the unbounded or bounded version of the reciprocity hypothesis applies in the minimal group experiment. As in case of Locksley et al. (1980), they found that participants gave less to ingroup members if they knew that their outcome is not dependent on them. Moreover, subjects also gave less to outgroup members if they knew that their outcome is not dependent on the outgroup. This shows that not only beliefs about the ingroup but also about the outgroup are important and thus seems to confirm the unbounded reciprocity hypothesis. However, to say that the bounded reciprocity hypothesis is therefore wrong is not correct because subjects still made more ingroup-favouring reward allocations across all conditions. So, even in the mere outgroup outcome dependent condition ingroup favouritism prevailed, suggesting that our expectations of reciprocity are at least partly bounded.

There are several other experiments which suggest that ingroup favouritism does not emerge due to ingroup favouring social preferences but expectations about reciprocity. Most famous are the studies conducted by Yamagishi and colleagues (Karp et al., 1993; Jin and Yamagishi, 1997; Yamagishi et al., 1998, 1999). For example, Karp et al. (1993) implemented the classic minimal group experiment and a modified version of it. In this modified version, players were told that in the end they would get a fixed amount of money which is independent on others' allocation decisions. While the classic minimal group experiment led to ingroup favouritism, the modified version did not. This result confirms the importance of beliefs. Gaertner and Insko (2000) also conducted a minimal group experiment but varied whether the other allocator was part of the ingroup or outgroup and whether subjects would personally get rewards or not. Again, the authors only found ingroup favouring allocations if participants' outcomes were dependent on another ingroup member.

All these findings regarding expectations of reciprocity and interdependence support "a model where individuals respond to the dependence structure and then reciprocate with favoritism towards those on whom they are dependent, with this effect considerably stronger for the ingroup" (Everett et al., 2015, p. 12). This is due to the general assumption of the ingroup as a container of generalised

reciprocity.⁴⁵ Thus, our expectations of reciprocity are (at least partly) bounded. The meta-study of Balliet et al. (2014) that we already cited in section 3.1.2 also emphasises the importance of outcome interdependence. The authors found stronger ingroup favouritism in experiments that involved interdependence of outcomes compared to those without outcome interdependence. For example, the effect size of ingroup favouritism in social dilemmas was 0.42, whereas the one in dictator games was 0.19. Yet, this also makes clear that outcome interdependence and thereby direct reciprocity cannot explain all observed ingroup favouritism, which brings us to indirect reciprocity and reputational concerns.

3.2.2 Indirect Reciprocity and Reputational Concerns

According to Everett et al. (2015), indirect reciprocity means that it is not the person that profits from your beneficial treatment who is expected to return your favour but someone else. This someone else is expected to do so because he knows that you previously treated others in a generous way. In other words, you build up a good reputation which will be beneficial for you in future interactions. In this way, seemingly altruistic behaviour that leads to no chances of direct reciprocity can in the long run still be utility maximising for someone with egoistic preferences. Yamagishi and colleagues have created a model called the bounded generalised reciprocity model that explains why indirect reciprocity provokes ingroup favouritism (Yamagishi & Kiyonari, 2000; Kiyonari & Yamagishi, 2004; Yamagishi & Mifune, 2008, 2009). To put it simple, group identification activates a default group heuristic strategy that leads to more prosocial behaviour within the ingroup. The first of the three core ideas of the bounded generalised reciprocity model tells us why this is the case: While humans have depersonalised and generalised trust in other ingroup members willingness to cooperate, this does not apply to outgroup members.⁴⁶ The other two core ideas of the model are then an ingroup specific variation of the indirect reciprocity definition given at the beginning of this paragraph: (1) Humans are motivated to build up and maintain a cooperative reputation within the ingroup because such a reputation leads to

⁴⁵ The expression “the group as a container of generalised reciprocity” stems from Yamagishi and Kiyonari (2000) and will be further discussed in the next paragraphs.

⁴⁶ According to Yamagishi and Mifune (2008), this is due to our evolutionary history. Section 3.3.1 will explore the evolution of indirect reciprocity in more detail.

strategic advantages. (2) Humans expect other ingroup members to behave prosocially towards them even though these ingroup members might not have benefited from our own cooperative/prosocial behaviour (so far).

Yamagishi and Mifune (2008) provide empirical evidence for their model. In a dictator game, participants distributed more money to fellow ingroup members compared to outgroup members. However, this was no longer true if participants were told that recipients would not know their group membership. In this condition, there was no significant difference between the giving rate regarding ingroup or outgroup recipients. These findings show the importance of reputation building in ingroup favouring behaviour. Without the ingroup recipient knowing that you are part of his group, your generosity will not lead to a positive reputation within your group. As a consequence, you behave less prosocially. Consistent with Yamagishi and Mifune (2008), Mifune et al. (2010) found that subjects only behaved in an ingroup favouring manner if there was a cue for monitoring. The authors let participants play a dictator game. While they knew whether the recipient was an ingroup or outgroup member, they were told that the recipient would never know the dictator's group membership. The experiment had two conditions: (1) The screen of the computer, on which the game had to be played, is neutral. (2) The computer screen displays a painting of eyes that critically stare at the player. The painting of the eyes should function as a cue for monitoring. In turn, monitoring implies that the way you behave in is not without consequences for your reputation. Mifune et al. found that in condition 1, dictators did not significantly differ between ingroup and outgroup recipients. However, condition 2 produced ingroup favouring allocations and thereby demonstrates the importance of reputational concerns in ingroup favouritism.

All these experiments presented regarding direct and indirect reciprocity have one substantial limitation. They only used artificial groups. Therefore, it is unclear whether these results also apply to real groups. For example, there are indications that punishment behaviour in a third-party punishment game depends on whether the experimenters examined real or artificial groups. Experiments with artificial groups tend to lead to less harsh ingroup than outgroup punishment (Jordan et al., 2014; Butler et al., 2013; Chen & Li, 2009; Goette et al., 2012) whereas experiments with real groups tend to lead to similar or even harsher ingroup punishment (Goette et al., 2006, 2012; Bernhard et al., 2006; Shinada et al., 2004; Mendoza et al., 2014).⁴⁷ For example, Goette et al. (2012) tested both randomly assigned

⁴⁷ Fehr et al., 2008 and Kubota et al. (2013) are exceptions to this rule.

real and artificial groups.⁴⁸ They found that real groups led to more ingroup favouritism. Moreover, the groups differed in their norm enforcement patterns. While in case of artificial groups punishers punished selfish ingroup vs. outgroup dictators more leniently, this was not true in case of real groups. The authors explained these results as follows: Members of real groups share a social history of social interactions and social ties, which raise empathy between group members. On one hand, this increased empathy reinforces the willingness to treat ingroup members more prosocially than outgroup members. On the other hand, it also reinforces members willingness to punish ingroup dictators who treated ingroup members badly. It is important to notice that increased empathy has nothing to do with beliefs about direct or indirect reciprocity but with ingroup love. Thus, the behaviour of real groups seems not to be solely describable by means of ingroup favouring beliefs.

Jackson (2008) provides further evidence for this argument. In his experiments, members of real groups behaved more cooperatively in simultaneous social dilemmas compared to members of artificial groups. This effect was mediated by group identification and thereby confirms previous findings of the connection between social identity and cooperative behaviour (Kramer & Brewer, 1984; de Cremer & van Vugt, 1999).⁴⁹ Nevertheless, as a study conducted by Ockenfels and Werner (2014) demonstrates, ingroup favouring beliefs are also of importance for real groups. They let participants play a dictator game in various versions, in which university affiliation always served as the line between ingroup and outgroup. In version 1, both the dictator and the recipient knew each other's group affiliation. In version 2, only the dictator knew the other's group affiliation. In version 3, the dictator could choose whether he wants to know the recipient's group affiliation. If he wanted to know it, the recipient would also be told the dictator's group affiliation. Version 4 is the same as version 3 except that here, the recipient would not be told the dictator's group affiliation if the dictator wanted to know the recipient's group affiliation. The authors attained the following results: (1) Public knowledge of group identities led to substantial ingroup favouritism. (2) There was less ingroup favouritism given the recipient was unaware (vs. aware) of the dictator's group affiliation. (3) Dictators wanted to know recipients' group affiliation less often if this created public knowledge (version 3) compared to if only they got to know the other's group affiliation (version 4). Ockenfels and Werner

⁴⁸ Platoons of the Swiss military, to which soldiers are randomly assigned to, functioned as the real groups. The artificial groups were formed via a lottery mechanism.

⁴⁹ Yet, it could be argued that group identification does not only increase the desire to positively evaluate the status of the respective group but also outcome interdependence and the possibility of indirect reciprocity.

(2014) conclude that “[t]he evidence supports the view that ingroup favoritism is partly belief-dependent” (p. 453). Therefore, both ingroup love and ingroup favouring beliefs appear to influence inter- and intragroup behaviour in real groups. Yet, further research is needed in order to assess how strongly each of the two affects ingroup favouritism.

3.2.3 Cooperative Norm Violation

The third ingroup favouring belief suggests that we behave more prosocially towards ingroup than outgroup members because we perceive social norms that recommend us to do so. There are several studies that show that group identification leads to higher adherence to group norms and that one of these norms typically is ingroup cooperation (Tajfel & Turner, 1979; Terry & Hogg, 1996; Jetten et al., 1997). Moreover, if someone strongly identifies with a group and follows its norms, he also anticipates that other ingroup members follow the group’s norms as well (Terry & Hogg, 1996; Mullin & Hogg, 1998). In turn, this reinforces ingroup cooperation. For example, Seinen and Schram (2006) found that participants acted more prosocially if they expected that other players behave prosocially as well.

Of course, the higher adherence to group norms and the consequent ingroup favouritism can be explained by ingroup love and thereby social identity. However, there is also a belief-based explanation because violating social norms can be costly (Fehr & Fischbacher, 2004). As a consequence, if an egoistic person believes that the overall utility of acting “egoistically” and thereby bearing the costs of norm violation is smaller than acting “altruistically” and thereby following the norm, he acts “altruistically”.⁵⁰ Now, given that norm violation and thus acting “egoistically” is costlier if it strikes ingroup compared to outgroup members, ingroup favouritism emerges.

This kind of reasoning is supported by Shinada et al. (2004) and Mendoza et al. (2014). The former found that noncooperative ingroup members were punished more severely than noncooperative outgroup members in a gift-giving game. Mendoza et al. (2014) implemented an ultimatum game where participants received a distribution offer and could accept or decline it. In the first study, black and white people played the game. Given the proposer had the same skin colour,

⁵⁰ Egoistically and altruistically are put into brackets because ultimately, a decision-maker with egoistic preferences always behaves egoistically. Yet, his actions might seem altruistic.

he was punished more harshly for an unfair offer than a proposer with a different skin colour. Their second study replicated this finding with college instead of racial group membership. Additionally, here, the authors discovered that the more students identified with their ingroup, the more they punished unfair ingroup members. Their third study revealed that the stricter punishment of ingroup members was mediated by fairness perception and not proposer evaluation. Unfair ingroup members violated the participants' fairness expectations and as a consequence had to be punished. Thus, both Shinada et al. (2004) and Mendoza et al. (2014) suggest that the costs of acting "egoistically" are higher if the action concerns an ingroup compared to an outgroup member, leading to ingroup favouring beliefs. However, there are also studies that found no such effect or even a contrary one (Bernhard et al., 2006; Goette et al., 2012; Kubota et al., 2013).

As a side note, such social norms which impose that you should favour the ingroup might also be relevant in a situation where an agent-neutral decision-maker is indifferent between alternatives. For example, a person can either give a certain amount of money to an ingroup member or an outgroup member and does not care about who gets it. Now, one option would be to flip a coin so as to define the final receiver. Another option would be to consider social norms so as to define the final receiver. Regarding this second option, the decision-maker would give the money to the ingroup member since social norms say that you should favour the ingroup. Now, it is important to notice that this decision would neither be based on a taste for the ingroup nor the fear of costs that might come along with norm violation. In fact, according to this dissertation's definition of discrimination, the decision-maker would not discriminate at all because he is indifferent between the two alternatives. Nevertheless, in the state of indifference, he might still always choose the alternative that favours the ingroup because he uses a respective social norm in order to reach a decision. Therefore, while the decision-maker is indifferent between the actual alternatives, he might not be indifferent to how he handles this indifference. This is why we could define such behaviour as "second-order discrimination".⁵¹ And if this "indifference-handling-rule" or more precisely its content treats people/groups differently, as it might be in case of social norms, there is second-order social discrimination. So, second-order discrimination might be of importance in certain decision. Nonetheless, the focus of

⁵¹ Let us look at an example of second-order discrimination in a non-social context. A decision-maker is indifferent to whether he wears his watch on his left or his right wrist. Now, he could flip a coin every time he puts on a watch so as to decide whether he wears it on his left or right wrist. But instead, he prefers to consider social norms so as to reach a decision. These norms involve that watches are worn on the left wrist which is why he always wears his watch on the left wrist.

this dissertation lies on possible “first-order discrimination” which involves the preference relations within a given choice set (and not on how someone handles indifference within that choice set). This is why we do not further elaborate on second-order discrimination.

To summarise, while it is often difficult to empirically separate ingroup favouring beliefs from ingroup love, it appears to be undeniable that such beliefs affect ingroup favouritism. However, only if a seemingly ingroup loving action is the sole product of ingroup favouring beliefs, it can be described as pure statistical discrimination. The experiments discussed in this chapter suggest that this seldomly is the case. Thus, the hypothesis that all discrimination is ultimately statistical discrimination is rather unlikely. It seems that we are not only statistical discriminators but also taste-based discriminators. Yet, this requires that we have ingroup favouring and/or outgroup derogating social preferences.⁵² So far, we simply assumed that they exist. In the next chapter we examine whether they truly do.

3.3 The Evolution of Agent-Relative Social Preferences

Out of an evolutionary perspective, strong taste-based discrimination poses a two-fold problem. The first one is that of social preferences and thereby altruism in general, whereby altruism implies “behaviors that are beneficial to the recipient and costly to the actor” (Silk, 2015, p. 64) for evolutionary biologists.⁵³ The evolutionary biological issue with altruism is as follows: If a group has both altruists and egoists, the latter should supersede the former sooner or later. This is because if an egoist is in need, she gets help from an altruist. In turn, if an altruist is in need, she cannot expect any help from egoists. So, while altruists for example share their food and thereby seem to decrease their fitness⁵⁴ because by doing so they have less food, egoists only profit from altruists and never sacrifice any fitness for others. As a consequence, egoists should have higher fitness than altruists. The second problem, which is of particular interest for this dissertation, is that of *agent-relative* social preferences. Why should it be adaptive to be altruistic within the ingroup but less altruistic, egoistic, or even hostile towards the outgroup?

⁵² Actually, only strong taste-based discrimination requires social preferences.

⁵³ Thus, from an evolutionary perspective, the concept of altruism is very close to that of economists.

⁵⁴ The concept of fitness will be defined a few paragraphs below under “*Kin Altruism*”.

In this chapter, we first examine the evolution of social preferences in general. Here, we present four concepts that explain why altruistic behaviour has been an evolutionary stable strategy in the course of evolution. Since these four concepts cannot satisfactorily explain all human altruism we then investigate the influence of culture on the evolution of altruistic behaviour. Finally, we discuss the conditionality of altruism and in so doing the idea of parochial altruism, which provides an ultimate explanation for agent-relative social preferences.

3.3.1 Why Altruistic Behaviour Can Be Adaptive

In order that altruism is adaptive it has to lead to higher fitness than egoism. Yet, as said above, the very concept of altruism involves that while an action benefits others, it is costly to oneself. Therefore, the only solution to this problem is that costly altruistic behaviour pays off in the long run. In this chapter, we discuss the following four evolutionary concepts where altruism ultimately leads to enhanced fitness: kin altruism, reciprocal altruism, indirect reciprocity, and costly signalling theory.

Kin Altruism

So as to understand kin altruism we first have to make an important distinction regarding the idea of fitness. On one hand, there is direct fitness which comprises the amount of my genes that spread within the direct family line (parent = >children). On the other hand, there is indirect fitness which comprises the amount of my genes that spread within the extended family via relatives. So, my fitness is not limited on how much offspring do I have but also involves how much offspring does my family excluding me has. Both together then result in inclusive fitness, which is what we refer to when we talk about fitness in this dissertation (Grafen, 2006; Scott-Phillips et al., 2011).

The concept of kin altruism precisely is based on the distinction between direct and indirect fitness. High cooperation between family members is very common in everyday life and can be explained by kin altruism (Burnstein et al., 1994).⁵⁵ Since relatives share a part of our genes it can be adaptive to help them, provided that the ratio of cost and benefit is positive. Hamilton (1964) formalised this insight which led to the Hamilton's rule: $r \times b > c$. Written out, the formula has

⁵⁵ For example, the experiment conducted by Krupp et al. (2008), which we presented in section 3.1.2, provides evidence for kin altruism. Here, participants were more cooperative in a public goods game the stronger the cue for kinship between players was.

the following implication: Altruism is adaptive if the fraction of genes the helper shares with the recipient of the help (r) multiplied by the benefit the recipient receives (b) is bigger than the costs the helper bears (c). A quote by Haldane illustrates what this means in practice: “I’d lay down my life for two brothers or eight cousins.” Brothers share half of our genes, whereas cousins share one-eighth of our genes. As a result, two brothers or eight cousins carry as many of Haldane’s genes as he does.

While kin altruism can be widely observed in human behaviour, there are animals where it is even more dominant, namely social insects such as ants and bees. Due to the haplodiploidy⁵⁶ of these insects it is adaptive for the workers to sacrifice their reproduction so as to serve their queen (Queller & Strassmann, 1998). Sherman (1977) provides another impressive example of kin altruism in wildlife. He studied the alarm calls of squirrels. The evolutionary puzzle of these alarm calls is as follows: While an alarm call might save the surrounding squirrels, it puts the squirrel that makes it at risk because it draws the raider’s attention to itself. So, squirrels that make these alarm calls are more likely to be killed and, as a consequence, such behaviour should extinct. Yet, Sherman found that in the context of kin altruism these alarm calls become an evolutionary stable strategy. To conclude, kin altruism is a ubiquitous phenomenon. Yet, it requires a non-negligible degree of kinship. We know that humans also help each other even if they are not related. Therefore, kin altruism is not sufficient to explain the whole spectrum of human altruism.

Reciprocal Altruism

The proverb “you scratch my back and I’ll scratch yours” contains the main idea of reciprocal altruism. Trivers (1971) first mentions reciprocal altruism and argues that “natural selection favours these altruistic behaviours because in the long run they benefit the organism performing them.” (p. 35) Therefore, it is an evolutionary stable strategy to cooperate with non-kin given the long-term fitness benefits of cooperation are higher than its costs. So, what seems like altruistic behaviour is actually egoism in disguise. We already discussed such behaviour in section 3.2.1 and called it direct reciprocity there. The key requirements for direct reciprocity are repeated interactions because otherwise your favour cannot be returned, which undermines reciprocal altruism. Experimental evidence confirms that. In a two-person interaction, the more probable future interactions are, the higher

⁵⁶ Haplodiploidy means that, regarding a certain species, males only have one chromosome set, whereas females have two chromosome sets. Due to that females share three quarters of their genes with their sisters, enabling stronger kin altruism.

the rate of cooperation gets (Andreoni & Miller, 1993; DalBo, 2005; Gächter & Falk, 2002). Furthermore, Trivers (1971) says that psychological adaptations such as “friendship, dislike, moralistic aggression, gratitude, sympathy, trust, suspicion, trustworthiness, aspects of guilt and some forms of dishonesty and hypocrisy” (p. 35) improve the functioning of reciprocal altruism. This is because they help us maintaining a beneficial dyadic cooperation and distinguishing between good and bad cooperators.

If reciprocal altruists cooperated with more or less every interaction partner as long as they assume that there will be future interactions, egoists would constantly exploit them. As a consequence, the ability to distinguish a like-minded reciprocal altruist from a selfish cheater would be decisive. There is evidence that humans actually have such a skill. For example, Mealey et al. (1996) found that participants recognised photos of people better when these people had been labelled as “untrustworthy” at first exposure compared to other adjectives. Additionally, we are not only able to identify cheaters but also to quickly recognise altruists (Brown & Moore, 2000). An experiment of Frank et al. (1993) confirms this insight. Before playing a one-shot prisoner’s dilemma, the authors let participants communicate face-to-face. The results reveal that “subjects who interacted for thirty minutes before playing one-shot prisoner’s dilemmas with two others were substantially more accurate than chance in predicting their partner’s decisions”. (p. 247)⁵⁷

Is reciprocal altruism an exclusively human phenomenon? Apparently not. Rutte and Taborsky (2008) found direct reciprocity among Norway rats in an adjusted version of a repeated prisoner’s dilemma. Here, rats preferentially helped cooperators instead of defectors. Dolivo and Taborsky (2015) even revealed that rats are able to differentiate between cooperators depending on the quality and the delay of their help. Moreover, other well-studied animals regarding the display of reciprocal altruism are for example bats (Carter & Wilkinson, 2013, 2015). And although Zentall (2016) argues that these behaviours are actually not the product of reciprocal altruism but laboratory induced Pavlovian conditioning, there are good arguments why this is not the case (see Dolivo et al., 2016).

So, reciprocal altruism seems to be part of (some) animals’ nature as well, which makes the phenomenon and its adaptivity even more robust. Nonetheless, the theory has two strong restrictions. First, reciprocal altruism only functions if there is a random number of repeated interactions. Second, its explanatory power

⁵⁷ However, other studies conclude that most humans are not better than chance in detecting liars and thus question the existence of such a skill (Eckman & O’Sullivan, 1991; Frank & Eckman, 1997).

is limited to few-person interactions (Fehr & Fischbacher, 2005). However, on one hand, humans often cooperate in large groups. On the other hand, people also behave altruistically in anonymous one-shot interactions where the possibility of direct reciprocity is excluded. Ultimately, altruistic punishment, as we have seen it in section 3.1.2, is not explainable by reciprocal altruism. Thus, while this concept provides an important supplement to kin altruism, it still leaves a lot of unsolved problems regarding altruism.

Indirect Reciprocity

We already discussed indirect reciprocity in section 3.2.2. As we know from that chapter, reputation is the key word in indirect reciprocity. Now, let us look at indirect reciprocity from an evolutionary perspective. The model (Alexander, 1987; Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998) states that helping non-kin results in a good reputation. In turn, having a good reputation rises the likelihood of receiving someone's help in the future even though there are no further interactions with that person. Hence, people behave altruistically in order to attain a good reputation, which is beneficiary in the long run. In previous chapters, we already presented laboratory experimental evidence for indirect reciprocity (Yamagishi & Mifune, 2008; Ockenfels & Werner, 2014). Additionally, there is also field experimental evidence for indirect reciprocity. In a large-scale field study conducted by Yoeli et al. (2013), reputational concerns tripled participation in a public-goods-game-like program of an electric utility company. Offering \$25 as an incentive to participate was four times less effective. Ultimately, studies suggest that children and even infants display indirect reciprocity (Kato-Shimizu et al., 2013; Meristo & Surian, 2013).

Indirect reciprocity solves one major problem of reciprocal altruism. There is no longer a necessity for repeated interactions because actors can build up an interaction superordinate reputation. As a result, altruism in one-shot interactions can be adaptive. Yet, notwithstanding how promising this approach is so as to explain aspects of human altruism that are inexplicable by kin altruism and reciprocal altruism, there are a few drawbacks. First, Leimar and Hammerstein (2001) found in their simulations that cooperativeness only emerges if groups are more or less isolated and there is no genetic mixing between groups. Second, it is unclear how the concept of good reputation should be modelled. Does not helping a person with a bad reputation jeopardise one's good reputation (e.g. Nowak & Sigmund, 1998) or not (e.g. Leimar & Hammerstein, 2001)? According to Fehr and Fischbacher (2005), "this question is intrinsically related to society's prevailing norms, which are themselves the product of evolutionary forces." (p. 34) As a consequence, indirect reciprocity is in need of another theory that explains which

norms prevail in a given society. Third, indeed, there are examples where indirect reciprocity led to cooperation in larger groups (Milinski et al., 2002; Panchanathan & Boyd, 2004). However, many non-cooperative equilibria are possible as well. Furthermore, hunter-gatherers had to collect and recall a lot of information in order to rightly assess the willingness for cooperation of each group member. Besides, in reality, information is often private and self-evidently, the whole process of indirect reciprocity becomes more and more complex the larger the group is (Bowles & Gintis, 2011). Finally, kin altruism, reciprocal altruism, and indirect reciprocity together can still not explain the phenomenon of strong reciprocity (Fehr & Gächter, 2000; Fehr & Fischbacher, 2005).

Costly Signalling Theory

Costly signalling theory provides a fourth explanation of altruism. The idea behind the theory is over a century old. In “The Theory of the Leisure Class”, Thorstein Veblen (1899) introduced the expression “conspicuous consumption”, which involves a hard-to-fake signal for wealth that should enhance prestige among the rich. More than 70 years later, Spence (1973) applied Veblen’s idea on the job market and argued that educational qualifications are taken as a signal for the employee’s productivity. Another two years later, signalling reached evolutionary biology. Zahavi (1975) used the approach so as to explain the helping behaviour in Arabian babblers.

The idea of signalling is as follows: Individuals give honest information about themselves by displaying behaviour that is costly. Yet, this costly behaviour benefits the individual because ultimately it increases reproduction and overall fitness (McAndrew, 2002). According to Smith and Bird (2000), a costly signal needs to fulfil four qualities. First, it has to be an honest signal of quality. Second, the costs which the signal involves must not be compensated by reciprocity. Third, others must be able to easily observe the signal. Fourth, the signal has to be beneficial, which means the signaller has to gain a net benefit. Now, behaving altruistically could be such a signal. As Gintis et al. (2001) argue: “[C]ooperation ... constitutes an honest signal of the member’s quality as a mate, coalition partner or competitor, and thus results in advantageous alliances for those signaling in this manner.” (Gintis et al., 2001, p. 103) Following this interpretation, costly signalling theory could for instance explain why societies have hunting games where they use a rather difficult instead of an efficient hunting technique or provide excessive amounts of food at feasts (Boone, 1998; Gurven et al., 2000; Smith & Bird, 2000; Sosis, 2000; Hawkes et al. 2001).⁵⁸

⁵⁸ However, one major weakness of costly signalling theory is as follows: The signalling of unobservable traits need not to manifest as altruistic acts but can also occur in other forms

Indirect reciprocity and costly signalling theory apparently have an overlap. In both models, the payback for the person's cooperative behaviour comes from third parties. Yet, Bowles and Gintis (2011) note the following difference: "[I]n the signalling model the third party responds favourably because the signal is correlated with some desirable but unobservable property of the actor; in the indirect reciprocity model the signal (cooperating with those in good standing) is the desirable property itself." (p. 71) However, as indirect reciprocity, it is not able to provide a solid explanation for all aspects of human altruism such as strong reciprocity.

The problem of strong reciprocity could be solved by group selection (Wilson, 1997; Boehm, 1999; Sober & Wilson, 1998).⁵⁹ While strong reciprocity decreases individual fitness, it raises group fitness since it sustains cooperation (Fehr & Gächter, 2000; Bowles & Gintis, 2002). Therefore, groups of strong reciprocators supersede groups of egoists. But this concept of group selection seems to be in conflict the basic idea of natural selection. Genes are the ones that are passed on to the next generation and individuals function as their vehicles in this transfer (Dawkins, 1976). Yet, if we go one level up, there are neither replicators (such as DNA information) nor vehicles (such as individuals) (Dawkins, 2012). Consequently, a trait that is exclusively beneficial to the group still has to be transmitted via genes. Due to that group selection can at best be relevant in small isolated groups since intragroup selection against strong reciprocators in combination with migration is a much stronger force than intergroup selection. According to Fehr and Fischbacher (2003): "The migration of defectors to groups with a comparatively large number of altruists plus the within-group fitness advantage of defectors quickly removes the genetic differences between groups so that group selection has little effect on the overall selection of altruistic traits (Aoki, 1982). Consistent with this argument, genetic differences between groups in populations of mobile vertebrates such as humans are roughly what one would expect if groups were randomly mixed (Long, 1986). Thus, purely genetic group selection is ... unlikely to provide a satisfactory explanation for strong reciprocity and large-scale cooperation among humans." (p. 789)

such as antisocial acts (Fehr & Fischbacher, 2003). So, given that costly signalling theory applies in case of altruistic behaviour, there needs to be another theory that explains why the prosocial-signal equilibrium prevailed (Bowles & Gintis, 2011). We will present such a theory in the next chapter, namely cultural group selection.

⁵⁹ The theory in which group selection is nowadays embedded is called multi-level-selection. Wilson and Wilson (2008) describe its basic notion as follows: "Adaption at a level X requires a corresponding process of selection at level X and tends to be undermined by selection at lower levels." (p. 381) Thus, both selection at the gene-level and the group-level are possible.

So, how can the remaining forms of human altruism be explained then? One explanatory approach is to identify them as maladaptations. Richard Dawkins (2006), who is a proponent of this explanation, writes: “Throughout most of our prehistory, humans lived under conditions that would have strongly favoured the evolution of all four [kin altruism, reciprocal altruism, indirect reciprocity, and costly signalling] ... most of your fellow band members would have been kin, more closely related to you than to other members of the band ... plenty opportunities for kin altruism to evolve. And ... you would tend to meet the same individuals again and again throughout your life—ideal conditions for the evolution of reciprocal altruism. Those were also ideal conditions for building reputations for altruism and the very same ideal conditions for advertising conspicuous generosity.” (p. 220) Therefore, strong reciprocity is a vestige of ancient times. It used to be advantageous because the environmental conditions in the late Pleistocene promoted such a trait. But these conditions changed and as a result the trait became disadvantageous. Nowadays, we neither sufficiently differentiate between one-shot and long-lasting interactions nor between strangers and intimates (Cosmides & Tooby, 1992; Price, 2008).

The maladaptation theory has some discrepancies though. First of all, group sizes of ancestral human societies seem to have been rather large and therefore suboptimal for reciprocal altruism (Gintis et al., 2008; Bowles & Gintis, 2011). Second, hunter-gatherers appear to have traded in distances over hundreds of kilometres and thereby probably had contact with various strangers (Keats, 1977; Fehr & Henrich, 2004). Thus, it should have been essential for them to distinguish between strangers and intimates as well as one-shot and long-lasting interactions (Bowles & Gintis, 2011).⁶⁰ Third, there is ample evidence that hunter-gatherer groups were neither isolated nor stable, which dampens the effect of kin altruism (Harpending & Jenkins, 1974; Lourandos, 1997; Howell, 2000; Woodburn, 1982; MacDonald & Hewlett, 1999; Fix, 1999; Moreno-Gamez et al., 2011). Due to these three problems we look for a further explanation of strong reciprocity, which brings us to culture.

3.3.2 The Role of Culture in Evolution

Richerson and Boyd (2005) define culture as follows: “Culture is information capable of affecting individuals’ behavior that they acquire from other members

⁶⁰ Besides, there are strong indications that humans can very well differentiate between one-shot and long-lasting interactions (Andreoni & Miller, 1993; DalBo, 2005)

of their species through teaching, imitation, and other forms of social transmission.” (p. 5)⁶¹ So, traits cannot only be transmitted genetically but also culturally via social learning (Creanza et al., 2017). The importance of such culturally transmitted knowledge becomes obvious if we image the situation of being lost in nature. We do not know how to make fire. We do not know which plants are poisonous. We do not know how to make arrows, nets, and shelters or how to hunt. Our ancestors once knew how to do these things, yet, today they are no longer culturally transmitted which is why modern humans have never learned them.⁶² The fact that we have to learn these abilities demonstrates that they do not have a genetic but cultural background (Chudek et al., 2015).

Yet, this shall not imply that genes and culture are exclusive concepts. The two can overlap. This is called gene-culture coevolution (Gintis, 2011; Richerson & Boyd, 2005; Henrich, 2011). It means that cultural traits that a group transmits from generation to generation can create a group structure that influences individual fitness or co-form the environment to which individuals adapt (Gintis et al., 2008; Feldman & Zhiotovskiy, 1992). In other words, a genetic change can be initiated by a former cultural change. A classic example of this process provides some humans’ ability to digest lactose after weaning. Areas where this is a common trait in the population (e.g. Northern Europe) correlate with the distribution of the earliest European cattle farms (Beja-Pereira et al., 2003). Therefore, the cultural invention of dairy farming initiated the natural selection of people with lactose tolerance since milk provided an additional nutrition form. Bersaglieri et al. (2004) found genetic evidence for the adaptation which enables the digestion milk products after weaning. It took place in the last 5’000 to 10’000 years and is said to be one of the strongest selections yet seen for any gene in the genome.

Comparable to genetic evolution, cultural traits “reproduce themselves from brain to brain and across time, mutate and are subject to selection according to their effects on the fitness of their carriers” (Gintis, 2011, p. 879). Thus, if a cultural adaptation directly leads to more individual fitness, it is little surprising that it prevails. For example, let us assume a hunter-gatherer invents a new arrow with small feathers at the end. These feathers stabilise the arrow’s trajectory and enable a harder and more precise shot. Since the new arrow makes hunting both

⁶¹ By information, Richerson and Boyd (2005) mean any kind of mental state that is acquired or modified by social learning and affects behaviour.

⁶² There is also no need to learn these abilities in a modern society because with respect to today’s environment they no longer contribute to survival.

more effective and efficient, every individual that adopts it increases her fitness.⁶³ As a consequence, the new arrow supersedes the old one and its production is from now on culturally transmitted. But can also norms emerge that (at least at the beginning) are costly for the individual but beneficial for the group? To put it differently, might strong reciprocity be a cultural adaptation?

There is ample evidence which suggests that altruistic behaviour varies with local cultural environments. Henrich et al. (2001) let 15 small-scale societies play the ultimatum game and found substantial differences between these societies. For instance, the Lamaleras, a whale hunting society, are dependent on cooperation in their daily life since you cannot catch a whale alone. After a successful hunt, they distribute the catch among all members of the group. This cooperativeness is mirrored in how they played the ultimatum game. 63% of proposers allocated half of the amount to the responder. Those who distributed differently normally gave even more, resulting in an overall average offer of 57%. In contrast, the Machiguenga, which is a Peruvian tribe, offered on average 26% of the pie and only one out of 21 responders rejected the offer. This outcome reflects the cooperativeness in their everyday life. Cooperation, sharing, or exchange beyond the family unit is uncommon. Accordingly, the Machiguenga do also not fear social sanctions or having a bad reputation. So, altruism seems to have a cultural component.

We know that the environment of our ancestors was not perfectly stable (Martrat et al., 2004). This circumstance promoted ways of fast adaptation such as cultural transmissions. Strong reciprocity could be one of these cultural inventions and enabled high cooperativeness in large groups even with migration. Still, how could this cultural norm spread out within a group even though it appears to be costly for the individual that adheres to it? According to Fehr and Fischbacher (2003), given there are enough strong reciprocators in a group, acting selfishly is no longer fitness enhancing because egoists get punished by strong reciprocators. Moreover, if even pure cooperators (individuals who cooperate but do not punish defectors) get punished for not punishing defectors, behaving like a strong reciprocator leads to highest individual fitness within a group. Besides, the more cooperators a group has, the less often strong reciprocators have to punish defectors. As a result, the intragroup disadvantage of strong reciprocators relative to pure cooperators gets smaller and might even vanish at one point. "At the limit, when everybody cooperates, punishers incur no punishment costs at all and thus have no disadvantage." (Fehr & Fischbacher, 2003, p. 790)

⁶³ We assume that producing these new arrows is only marginally costlier than the old arrows without small feathers.

This is how strong reciprocity could become dominant within a group. Here, it is important to remember that one great difference between cultural and genetic adaptations is their speed. Unlike genetic adaptations, cultural adaptations can occur within a single generation. So, a group of egoists can become a group of strong reciprocators in few decades. Due to this, the situation where an insufficient number of upcoming strong reciprocators gets superseded by egoists might get bypassed.⁶⁴ But how could strong reciprocity spread between groups? One possible answer is that groups of strong reciprocators simply had higher rates of reproduction. However, there is another concept that provides an answer to this question, namely cultural group selection (Henrich & Boyd, 2001; Boyd et al., 2003). There is ample evidence which implies that our ancestors experienced many intergroup conflicts (Jorgensen, 1980; Otterbein, 1985).⁶⁵ In such conflicts, a group of altruists that follows the cultural norm of strong reciprocity displays a high level of cooperativeness and consequently outcompetes a selfish group. Here, outcompete does not mean that the defeated group gets eliminated. It is their cultural norm of selfishness that vanishes because the losing group is forced to adapt the winner's cultural norms and institutions (Kelly, 1985; Soltis et al., 1995).

Thus, if we look at evolution from a dual inheritance perspective, which includes both genetic and cultural adaptations, we realise that the two inheritances can lead to two different selection processes. On one hand, we have gene-level selection. On the other hand, cultural group selection ultimately provokes a group (norm) selection mechanism. Moreover, in the course of evolution, some cultural adaptations might have found their way into our genes via gene-culture coevolution. Human morality and our ability to internalise norms could be the product of such a process (Gintis et al., 2008; Gintis, 2003). First, brain regions involved in moral judgements and behaviour such as the prefrontal cortex or the orbitalfrontal cortex are virtually unique to or most highly developed in humans and without doubt evolutionary adaptations (Moll et al., 2005; Schulkin, 2000). Second, the emergence of human morality is closely tied to the evolution of the human prefrontal cortex (Allman et al., 2002). Third, Gintis (2011) states that “[t]he social environment of early humans was conducive to the development of prosocial traits, such as empathy, shame, pride, embarrassment and reciprocity, without which social cooperation would be impossible.” (p. 879) Following this

⁶⁴ Here, insufficient means that there are too few strong reciprocators in order that strong reciprocity becomes the best behavioural strategy.

⁶⁵ We will further investigate the importance of war regarding the evolution of altruism in the next chapter.

line of argumentation, morality is a proximate mechanism that serves as a psychological rewarding and/or punishment system which ultimately maintains strong reciprocity. Or to put it in more drastic words, the cultural norm of strong reciprocity got directly encoded into the human brain. Here, it is important to notice that strong reciprocity as a universal structure of human morality only acquires concrete content in the context of specific cultural values regarding the legitimate rights and obligations of individuals (Gintis et al., 2008). This explains why Henrich et al. (2001) found considerable variance in how members of 15 small-scale societies behaved in the ultimatum game. In contrast, studies conducted in advanced industrial societies led to rather similar results since individuals of such societies considerably agree on the content of moral behaviour (Fong et al., 2005; Gintis et al., 2008).

Thanks to gene-culture coevolution and cultural group selection we might have found a conclusive explanation for strong reciprocity. However, while there is little doubt that elements of culture adapt over time (Bentley et al., 2004; Durham, 1991; Gabora, 1995, 2011; Mesoudi et al., 2004, 2006; Orsucci, 2008), the analogy between genetic and cultural adaptations and the consequent idea of dual inheritance is not undisputed. Most commonly, critics say that “the gene is a well-defined, discrete, independently reproducing and mutating entity, whereas the boundaries of the unit of culture are ill-defined and overlapping” (Gintis, 2011, p. 879). Yet, in the same paragraph, Gintis counters that this conception of well-defined genes is out-dated, which is a valid point, considering the epigenetics revolution (Carey, 2012). Gabora (2011) criticises that there is neither an objective benchmark for determining cultural fitness nor do cultural “mutations” occur randomly. Additionally, Tooby and Cosmides, (1992) claim that at least some behaviour, whose origin is said to be cultural, can be explained by biology alone. Nevertheless, despite these objections, it seems inappropriate to simply characterise gene-culture coevolution and cultural group selection as incompatible with natural selection and thus wrong (Fehr & Fischbacher, 2003; Gintis, 2011; Richerson et al., 2016).

3.3.3 Why Altruism Is Conditional

So far, we only discussed how social preferences could evolve. However, the title of section 3.3 is “The Evolution of *Agent-Relative* Social Preferences”. Section 3.1.2 revealed that whether or not we behave altruistically (partly) depends on the group membership of the receiver. If the receiver is a fellow

ingroup member, we treat her prosocially. If the receiver is an outgroup member, we treat her less prosocially, neutrally, or even antisocially. So, evolution has not generated universal but conditional altruism. This subchapter investigates why this conditionality might be the missing piece of the jigsaw in order to attain the ultimate explanation of human altruism.

When we discussed cultural group selection it was already mentioned that group conflicts and war were substantial parts of our ancestors' lives (Jorgensen, 1980; Otterbein, 1985). The growth rate of human population can serve as an indicator for how frequent clashes of groups must have been. From 100'000 BC until 20'000 BC, growth was close to zero, ranging from 0.002% to 0.1% (Bocquet-Appel et al., 2005; Hassan, 1980). Yet, the environmental conditions should have allowed a rate of about 2% (Hassan, 1980; Johansson & Horowitz, 1986). This gap between possible and actual growth suggests that humans themselves were their own worst enemy (Bowles & Gintis, 2011).

In the late Pleistocene, which also comprises the ending of the last glacial period, the climate was volatile and led to unpredictable natural disasters (Martrat et al., 2004). These unstable conditions laid the foundation for intergroup wars. On one hand, groups fought for resources so as to ensure immediate survival. On the other hand, they also wanted to protect themselves against future disasters and in so doing did not back away from attacking other groups that might endanger their future survival (Wendorf, 1968; Ember & Ember, 1992). Additionally, the unstable environment led to long distance migrations. Here, groups who had no established political relations frequently encountered each other, provoking conflict (Bowles & Gintis, 2011).

Archaeological findings are in line with the idea of belligerent ancestors. Bowles (2009) examined bones on marks of violent death. He infers that in the late Pleistocene and early Holocene the mortality rate which can be traced back to warfare was approximately 14%. Although this is an impressive number, there are three reasons why we have to treat it with caution. (1) It is not possible to differentiate between deaths caused by intergroup conflicts and deaths caused by intragroup conflicts. (2) Not all violent deaths leave marks in bones. (3) So far, only a tiny fraction of our ancestors' bones was found and thus could be analysed. As a result, Bowles' violent death rate of 14% is not representative. Nonetheless, the number probably points in the right direction. In the late Pleistocene, hunter-gatherers did not only behave altruistically. Intergroup conflicts seem to have been frequent and widespread.

Human's tendency for belligerence towards people from the outgroup, so-called parochialism, is puzzling out of an evolutionary perspective. This is because

such a trait should decrease the fitness of an individual. In comparison with selfish but tolerant individuals, parochialists have a higher risk of death and are less likely to benefit from intergroup relationships. Consequently, tolerance should supersede parochialism. But like in the case of egoists who should outcompete altruist, reality proves the opposite. Both altruism and parochialism are commonly observable human traits. Now, the dazzling idea of Choi and Bowles (2007) is as follows: While neither altruism nor parochialism can be an evolutionary stable strategy on its own, both together can. This intersection of the two concepts is called parochial altruism.

How do Choi and Bowles reason the notion of parochial altruism? We know that, on a group-level, altruists outcompete egoists due to the former's higher level of cooperation. Yet, we also know that group selection in and of itself is controversial. Given selection exclusively occurs on the gene-level, the advantage of altruism on the group level becomes irrelevant, unless another mechanism fosters intergroup competition and in this way a kind of group selection. Parochialism could function as such a mechanism. If intergroup hostility leads to sufficient conflicts, traits that are for the good of the group can prevail because those who have them outcompete those who do not. In the end, this provokes a sort of group selection.

This last sentence makes clear how the two contrary behaviours might complement each other. On one hand, altruism alone increases group fitness, however, there is no selection process on the group level. Thus, the individual costs of behaving altruistically are higher than its benefits. On the other hand, parochialism alone provides a mechanism for group selection. However, selfish parochialists would not voluntarily engage in intergroup conflicts because "they are not willing to risk death in order to benefit their group members." (Choi & Bowles, 2007, p. 637) Nonetheless, unlike tolerant egoists, they bear the extra cost of parochialism. As a consequence, even though parochialism leads to a group selection mechanism, the trait does not prevail because it is neither advantageous on the individual nor on the group level. So, we see that both behaviours vanish if they evolve alone. But if there is a co-evolution of altruism and parochialism, they back each other and become complementary since only parochial altruists start war and, in this war, risk their lives for the good of the group.

The decisive question for the evolution of parochial altruism is as follows: Were there sufficient group conflicts in order that intergroup selection was not (entirely) superseded by intragroup selection? At the beginning of this chapter we said that warfare was probably common in the late Pleistocene and early Holocene. With regard to ancestral hunter-gatherer societies, Bowles (2009) states that "the estimated level of mortality in inter-group conflicts would have had

substantial effects, allowing the proliferation of group beneficial behaviours that were quite costly to the individual altruist.” (p. 1293) Thus, according to Bowles, parochial altruism could have evolved. Or to put it differently, the data we have about the late Pleistocene seems not to be incompatible with such a course of evolution.

Choi and Bowles (2007) theoretically analysed the evolution of parochial altruism by means of agent-based simulations. In these simulations, there were four types: tolerant egoists, parochial egoists, tolerant altruists, and parochial altruists. The simulated environmental conditions were based on the known data of the late Pleistocene. Given at least one of two encountering groups was mainly populated by parochialists, conflict occurred. Here, the group with more parochial altruists tended to prevail. The authors let the four types interact with each other over thousands of generations and found two equilibria: “In millions of simulated evolutionary histories, the populations emerging after thousands of generations of selection tend to be either tolerant and selfish, with little warfare, or parochial and altruistic with frequent and lethal encounters with other groups.” (Bowles, 2008, p. 326) So, in their model, the emergence of parochial altruism cannot be ruled out. Other studies employing evolutionary simulations also support the prevalence of parochial altruism (García & van den Bergh, 2011; Gao et al., 2015). Nevertheless, Choi and Bowles (2007) emphasise that they merely provide a possible explanation for how humans could have become both altruistic and warlike. The paper contains no evidence of a warlike genetic predisposition and remains purely theoretical. It only states that if such a predisposition exists, it could have co-evolved in the way Choi and Bowles describe it. Finding conclusive empirical and genetic proof has to be done in other research.

So, what have other papers found? We have already discussed massive evidence for ingroup favouring and outgroup derogating behaviour in section 3.1. Such findings came from both field experiments (Voors et al., 2012; Bandería et al., 2005; Leider et al. 2009; Gneezy & Fessler, 2012) and laboratory experiments (Charness et al., 2007; Chen & Li, 2009; Leibbrandt & Sääksvuori, 2012; Abbink et al. 2010, 2012; Fowler & Kam, 2007; Ahn et al., 2011; Bernhard et al., 2006; Butler et al., 2013; Goette et al., 2006, 2012; De Dreu et al., 2015). Moreover, researchers detected a connection between altruism and parochialism in war-like situations. On one hand, Gneezy and Fessler (2012) discovered that the willingness to punish non-cooperative group members and reward cooperative ones increases during violent intergroup conflicts. On the other hand, Voors et al. (2012) found that people who are exposed to violence are more risk seeking and behave more altruistically towards their neighbours. Ultimately, such traits are

war deciding and thus, in situations where a group needs them most so as to win a conflict, we might instinctively reinforce them.

Then, section 3.2 analysed whether altruistic ingroup favouring preferences can actually be explained by selfish ingroup favouring beliefs. Here, we said that such beliefs certainly affect behaviour, yet, they are not able to explicate all altruistic behaviour. Finally, the idea of social identity theory fits that of parochial altruism well. In fact, parochial altruism could be the ultimate explanation for it. Due to social identity individuals no longer make sharp distinctions between their own and the group's welfare (cf. depersonalisation), leading to behaviour that is for the good of the group. Additionally, the desire to improve or maintain one's positive social identity by means of group comparison can give rise to social competition. In turn, this promotes a group selection mechanism. Therefore, social identity and its implications could be the proximate mechanisms of parochial altruism, or in other words, the evolution of parochial altruism provides an ultimate explanation for social identity theory.

However, there are also critics of parochial altruism. Yamagishi and Mifune (2016) tested three hypotheses of parochial altruism: (1) unconditional intragroup cooperation; (2) non-instrumental, non-retaliatory, and costly intergroup aggression; and (3) the positive relationship between intragroup cooperation and intergroup aggression. The authors conclude: "Laboratory experiments revealed no support for the unconditional nature of intra-group cooperation, mostly negative evidence for the non-instrumental, non-retaliatory, and costly nature of inter-group aggression, and mixed evidence for the positive relationship between intra-group cooperation and inter-group aggression." (p. 39)

How convincing is this critique? First of all, we have to keep in mind that Yamagishi, who is the founder of the bounded generalised reciprocity model, is an early critic of social identity theory. Thus, it is little surprising that he also criticises parochial altruism since the two concepts are connected. Second, although Yamagishi and Mifune claim that there is no unconditional intragroup cooperation, we came to a different conclusion in section 3.2. For example, the meta-analysis of Balliet et al. (2014) revealed that the effect size of ingroup favouritism is indeed higher given there is mutual interdependence. Yet, there is also ingroup favouritism in anonymous dictator games that neither enable direct nor indirect reciprocity. Third, Yamagishi and Mifune have a point when they say that ingroup favouritism is mainly the product of ingroup love and not outgroup derogation. For example, Halevy et al. (2012) demonstrated that if a game allows to express ingroup love and outgroup derogation separately, players mostly express ingroup love and not outgroup derogation. Balliet et al. (2014) or Aaldering et al. (2018) come to a similar inference. Indeed, there is also evidence for outgroup

antipathy as for example in case of *schadenfreude* (Cikara et al., 2014). Furthermore, three newer experiments further support the idea of parochialism. De Dreu et al. (2015) manipulated cognitive self-control via a Stroop Interference Task (Stroop, 1935). The authors found that compared to the easy task, the difficult one led to more parochially altruistic behaviour in an IPD-MD game (cf. Halevy et al., 2008).⁶⁶ Cacaault et al. (2015) provide evidence for unprovoked parochial altruism. In their experiment, participants tended to benefit the ingroup at the cost of the outgroup even if they could have reached the same outcome without harming the outgroup. Böhm et al. (2016) confirm these findings (Rusch et al., 2016). Yet, despite this evidence it is unclear whether human outgroup hostility was truly strong enough so as to produce sufficient outgroup derogation in order that a group selection mechanism emerged. Fourth, Yamagishi and Mifune admit that more studies are needed that examine the relationship between intragroup altruism and intergroup parochialism. So far, most evidence of how the two concepts are linked is indirect, revealing that they correlate with the same factors, as for example intergroup competition, social distance, and testosterone (De Dreu et al., 2015; Diekhof et al., 2014; Reimers & Diekhof, 2015). The only study Yamagishi and Mifune cite that examines the correlation on an individual level is one they conducted themselves (Yamagishi & Mifune, 2009). Here, they found a negative and not a positive relationship. However, for example in case of sport fans, a strong identification with one's club promotes ingroup favouritism and can also lead to outgroup hostility (Lee, 1985). So, here, we seem to find a direct positive individual correlation between the two concepts. But maybe, this is due to the zero-sum game character of sports.

Thus, while two arguments of Yamagishi and Mifune (2016) are questionable, one argument is rather strong, namely that there is little non-instrumental intergroup aggression. Still, unlike Yamagishi suggests, generalised reciprocity appears not to be able to explicate the ingroup favouring behaviour that experiments reveal. Interestingly, Böhm (2016) came up with a distinction between two manifestations of parochial altruism: a weak and a strong one. He proposes “a semantic differentiation between effects that are based on a lack of positive attitudes toward the out-group, i.e., weak parochial altruism, and effects that are due to negative attitudes toward the out-group, i.e., strong parochial altruism” (p. 2). Thus, what researchers might mainly find in their experiments is not strong but weak parochial altruism, implying ingroup favouritism and outgroup neutrality. This of course raises the following question: If today humans primarily display weak parochial altruism, was ancestral intergroup aggression strong enough to

⁶⁶ This suggests that parochial altruism might operate by means of intuitive mechanisms.

create a group selection mechanism? And consequently, provided that parochialism was not strong enough to evoke a group selection mechanism, is there another explanation for ingroup favouritism?

The evolution of weak parochial altruism could have been possible by means of cultural adaptations, cultural group selection, and gene-culture coevolution. Here is how that might have occurred. As described in the last chapter, during the late Pleistocene, the social norm of strong reciprocity emerged because it helped to maintain a high level of cooperation in a changing environment. This norm was bounded to the ingroup. So, on one hand, while hunter-gatherers cooperated with fellow ingroup members, they treated outgroup members neutrally. On the other hand, while they harshly punished selfish ingroup members, they behaved more leniently towards selfish outgroup members (cf. Shinada et al., 2004; Mendoza et al., 2014). This includes cases where an outgroup member does not treat an ingroup member or another outgroup member prosocially. In this way, a group of weak parochial altruists could protect itself against selfish outgroups that wanted to exploit them. This is due to two reasons. First, the weak parochial altruists approached their outgroups with a selfish attitude as well. Second, since weak parochial altruists treat the outgroup neutrally, they do not engage in costly punishment of selfish outgroup behaviour. Thus, the norm of strong reciprocity prevails in a limited and therefore controllable scope and thanks to the quickness of cultural adaptations it can emerge within a single generation. Next, since groups of weak parochial altruists are fitter than groups of egoists, the cultural norm of ingroup bounded strong reciprocity spreads via cultural group selection. Ultimately, the process of gene-culture coevolution engraves the norm into our hardware and thereby genes in form of social identity (cf. Ihara, 2011).

It has to be highlighted that the above paragraph is a hypothesis and needs further proof. Yet, it provides a comprehensive explanation for how altruism evolved and why it is particularly prevalent in case of the ingroup without simply describing it as a maladaptation. Additionally, unlike the evolution of strong parochial altruism, it is not dependent on substantial intergroup aggression. Admittedly, cultural group selection also requires some sort of group conflicts. Yet, such conflicts are inevitable in an unstable environment and do not need additional outgroup hostility. If a group of egoists runs out of food because the territory provides too little food so as to feed all selfish groups in it, they also start fighting against the other groups. This is because in so doing, there is at least the chance to survive. Otherwise, they are dead for sure.

After examining parochial altruism, let us consider another explanation for agent-relative social preferences. It comprises the idea that humans rather interact

with people they are familiar with than unfamiliar people. We call this phenomenon *anxiety about the unknown*. Such anxieties can be observed in an intergroup context. An interaction with an outgroup member leads to more stress and anxiety than an interaction with a fellow ingroup member (Shelton et al., 2009; Trawalter et al., 2012). Moreover, there is a positive correlation between the anxiety about the unknown and ingroup favouritism (Paolini et al., 2006). Thus, our agent-relative social preferences seem not only to derive from the groups we are part of and those we are not part of but also from groups we know and those we do not know. Of course, it seems natural that our ingroup is also the group we know and the outgroup the one we do not know. This insight leads to the following reasoning: If we simply get to know the outgroup better, our anxiety and stress produced by the outgroup decreases. In turn, this should shrink ingroup favouring preferences.

This is precisely what the contact hypothesis describes: Provided that the conditions for contact are advantageous, contact between members of two groups reduces prejudices towards the outgroup (Allport, 1979).⁶⁷ There are three psychological processes behind the contact hypothesis: decategorisation, attitude generalisation, and recategorisation. First, the outgroup member with whom you interact is no longer perceived as part of the outgroup but as an individual which reduces prejudices towards that specific outgroup member (Brewer & Miller, 1984, 1988; Miller, 2002). Second, this change of attitude towards the specific outgroup member is transferred to the outgroup as a whole (Brown & Hewstone, 2005; Wilder et al., 1996). Third, due to these changes of attitudes towards the outgroup the two groups are reappraised and might ultimately form a common ingroup identity (Gaertner et al., 2016; Kite & Whitley, 2016).

There is ample evidence for the contact hypothesis. Pettigrew and Tropp (2006) conducted a meta-analysis, which consisted of 515 studies. They found a negative effect of intergroup contact on prejudice with an effect size of 0.22. Given the four prejudice reducing conditions of contact defined by Allport (1979) were encountered, the effect size even rose to 0.29. Thus, contact seems to truly reduce prejudices.

⁶⁷ Allport (1979) defined four conditions for positive intergroup contact: (1) Groups have equal status. (2) Groups have to work together cooperatively in order to achieve a superordinate goal. (3) Groups have the possibility to get acquainted with each other (and to become friends (Pettigrew, 1998)). (4) Intergroup contact occurs in a context of supporting of authorities, law, or customs. While these conditions are promotive for positive intergroup contact, they are not necessary (Pettigrew & Tropp, 2006).

So, is ingroup favouritism simply a question of familiarity? The answer is no. First, as can be seen, even with advantageous conditions the prejudice reducing effect of contact is barely moderate. Second, even if prejudices get smaller, this does not necessarily have to affect behaviour. For example, Jackman and Crane (1986) indeed found evidence that contact has a positive impact on standard measures of racial affect. However, this impact had little effect on white people's support for political policies designed to redress racial inequalities: "In other words, intimate contact promoted emotional acceptance of Blacks, just as the contact hypothesis predicts. However, it left unaltered a resilient core of conservative attitudes that led members of a dominant group to defend their privileges and to accept the kinds of inequalities that prevent the optimal conditions for contact from being implemented."⁶⁸ (Dixon et al., 2005, p. 706) Yet, other studies come to a different conclusion. They find that if the advantaged group has contact with the disadvantaged one, the former is more likely to approve political measurements that improve the situation of the latter (Dixon et al., 2007; Cakal et al., 2011). Third, there is evidence which suggests that a common ingroup identity increases outgroup derogation towards those groups that both the former ingroup and outgroup perceive as outgroups. For instance, Kessler and Mummendey (2001) examined group identity prior and after the German reunification. Prior to the reunification, West and East Germany had viewed each other as outgroups. Then, after the reunification, there were two identity-clusters. Some Germans developed a strong common ingroup identity as simply Germans, whereas others mainly derived their ingroup identity from regional markers and thus developed a weaker common identity. The authors found that on one hand, those with the stronger common ingroup identity displayed less prejudice towards the former outgroup than those who developed a weaker common ingroup identity. Yet, on the other hand, those who strongly identified themselves as Germans after the reunification expressed more prejudice against non-Germans compared to the rather regional identifiers. Therefore, through intergroup contact, overall ingroup favouritism has not vanished. Instead, the categorisation of ingroup and outgroup has simply changed. In conclusion, anxiety about the unknown appears to play a role in ingroup favouritism. However, familiarity alone is not the reason why humans display ingroup favouritism.⁶⁹

⁶⁸ It has to be noticed that to defend one's privileges does not require ingroup favouring preferences. Someone with egoistic preferences might do the same. Yet, it excludes that white people developed altruistic preferences towards black people.

⁶⁹ The ultimate explanation for anxiety of the unknown will be given in section 4.1.1, where we introduce the belief in the superiority of familiar alternatives.

To finish this chapter, let us examine the following thought: As we have seen, culture and cultural norms seem to have been essential in the evolution of human altruism. So, would it be possible to alter culture in such a way that ingroup favouritism vanishes? The ideas and theories presented in this chapter suggest that (weak) parochial altruism, which might originally have been a cultural adaptation, is encoded into the human brain. Concurrent with that, “in all societies, individuals view themselves as part of defined social groupings (ingroups) characterized by mutual cooperation and reciprocal obligation (Levine & Campbell, 1972; Sumner, 1906)” (Brewer & Yuki, 2007, p. 307). This seems to imply that ingroup favouritism cannot be fundamentally eliminated by culture (at least not in the short run).

Yet, even though the capacity for social identity seems to be hardwired and universal, where we draw the line between ingroup and outgroup is not (Turner et al., 1987). As mentioned in section 3.1.2, perceived similarity and dissimilarity plays an important role in social categorisation. But whether we perceive someone as similar and thus part of the ingroup or dissimilar and thus part of the outgroup can be manipulated (cf. Levine et al., 2005). Therefore, a culture that emphasises similarity between all individuals could be able to diminish ingroup favouritism. In so doing, it “tricks” the apparent human nature to mainly be altruistic towards the ingroup by making us perceive more and more people as ingroup members.⁷⁰ Theoretically, it is even possible that the ingroup at one point includes all humans and as a result there is no outgroup left. However, it is unclear whether such a situation would lead to universal altruism or complete personalisation. Maybe humans always need an outgroup in order to define the ingroup towards which they behave altruistically (Hogg, 2001). In the absence of an outgroup, altruism would decay. Yet, there is also evidence indicating that a sense of “Us” is possible without “Them” which might enable universal altruism (Gaertner et al., 2006). As a consequence, while culture cannot alter our predisposition for parochial altruism and social identity in the short term, it should be able to change the scope of the ingroup towards which we behave prosocially. And given that the ingroup either includes all humans or only the individual himself, ingroup favouritism could disappear.

To quickly summarise this subchapter, there are two evolutionary concepts that could have led to agent-relative social preferences: strong parochial altruism and weak parochial altruism. Strong parochial altruism requires substantial human

⁷⁰ Similarly, a culture that involves the exact opposite of what we describe in this paragraph might also diminish ingroup favouritism. By emphasising dissimilarity, it leads to smaller ingroups which at one point might only include the individual himself. Such a situation would lead to complete personalisation.

belligerence because only if this is the case, a group selection mechanism emerges that makes both parochialism and altruism adaptive. In turn, weak parochial altruism requires cultural adaptations, cultural group selection, and gene-culture coevolution, yet, no non-instrumental intergroup aggression. These two evolutionary theories are not mutually exclusive. Moreover, they can be complemented with the idea of anxiety about the unknown. Further research is needed so as to define how important each of these three concepts were in the course of human evolution.

To conclude the whole section 3.3, social preferences have different sources. On one hand, there are the four widely accepted evolutionary theories of altruism, namely kin altruism, reciprocal altruism, indirect reciprocity, and costly signaling. On the other hand, there are the more controversial ideas of gene-culture coevolution combined with cultural group selection and parochialism that might have provoked a sort of group (norm) selection mechanisms. By means of these mechanisms we can explain why human preferences are agent-relative. But although especially gene-culture coevolution in combination with cultural group selection appear to be promising candidates so as to explicate all aspects of altruism, the existence of these mechanisms is still disputed. Given they have not existed, we have to declare agent-relative preferences as maladaptations. Yet, this hypothesis is not really convincing, which is why we do not stick to it in this dissertation. Then, the anxiety about the unknown might also have played a role in the formation of our preferences. This is because it leads to mistrust of strangers and since strangers are typically outgroup members, this mistrust might spread to the outgroup in general. Finally, the evidence presented in this chapter suggests that agent-relative social preferences truly evolved. As a result, taste-based discrimination seems to actually exist and is not simply statistical discrimination in disguise.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.