Cooperation in Social Groups:

Reactions to (moral) deviants

Dissertation zur Erlangung des akademischen Doktorgrades

Doctor philosophiae (Dr. phil.)

Vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften

der Friedrich-Schiller-Universität Jena

seit 1548

Von Dipl.-Psych. Stefanie Hechler

Geboren am 13.05.1985 in Nürnberg

Gutachter:

1. Prof. Thomas Kessler
2. Prof. Franz J. Neyer
3. PD Dr. Raoul Bell

Tag der mündlichen Prüfung:

Table of Content

Research Line I

Research line II

Research Line III

Appendix

List of Tables

List of Figures

Summary

The present dissertation examines the influence of self-involvement with perpetrators and victims on third-party reactions to deviants. Dealing with others' social behaviors regulates social life and successful cooperation between interaction partners. Third-party reactions to deviants are sensitive to group context, and thereby more likely to protect ingroup interests. Such biased reactions raise the question of how much they are triggered by involvement (i.e., shared group membership, empathy) with perpetrators or victims of deviance. The three reported lines of research extend the current knowledge on cognitive (memory), emotional (anger), and behavioral (punishment) reactions to deviance within and between social groups.

Research Line I examined whether accurate memory for persons' social behavior is group-specific. Two studies investigated memory for uncooperative individuals (i.e., unfair or cheating) in minimal group contexts. Uncooperative ingroup members were remembered better than other ingroup and outgroup members. In contrast, guessing behavior indicated that participants assumed more cooperative ingroup member than outgroup. In a third study, a salient ingroup enhanced memory for ingroup deviants (i.e., cheating and trustworthy) compared to outgroup deviants in a natural group context.

Research Line II investigated whether involvement with victims is crucial for anger about deviance. Therefore, emotional reactions to the wrongfulness (i.e., perpetrator's intentions) and the harmfulness (i.e., consequences for a cared-for-other) of deviance were examined. Across three studies, the perpetrator's intentions to cause harm elicited more anger than the harmful consequences. A clear separation of the two features demonstrated that anger reacts to intentions, whereas empathy is sensitive to harm. The results indicate that moral outrage can emerge irrespectively of empathic anger.

Research Line III examined how involvement with perpetrators or victims of deviance influences anger and punishment. Five studies modulated unfairness, perpetrator and victim group membership orthogonally. Anger and (altruistic) punishment emerged consistently as responses to unfairness, even in outgroup interactions. Negative reactions to unfair distributions towards ingroup victims increased with higher identification with a minimal ingroup. Ingroup perpetrators elicited more anger and were punished harsher in a natural conflictive than in a cooperative (inter)group context.

Taken together, memory, anger, and punishment are sensitive to perpetrators' and victims' group memberships, and also emerge irrespective of self-involvement. The discussion addresses how such reactions facilitate social life and cooperation in groups.

Zusammenfassung

Die vorliegende Dissertation untersucht inwiefern ein persönlicher Bezug zu Tätern und Opfern die Reaktionen Dritter gegenüber Devianten beeinflusst. Reaktionen auf das soziale Verhalten anderer unterstützen ein friedliches Zusammenleben und Kooperation zwischen Interaktionspartnern. Die Reaktionen Dritter gegenüber Devianten sind gruppenspezifisch, und schützen häufig die Interessen der eigenen Gruppe. Inwiefern sind diese Reaktionen abhängig von der Beziehung, die zwischen Beobachter, Tätern und Opfern von Devianz besteht? Die folgenden drei Forschungslinien erweitern den Stand der Forschung zu kognitiven (Gedächtnis), emotionalen (Ärger) und Verhaltensreaktionen (Bestrafung) gegenüber Devianten innerhalb und zwischen Gruppen.

In *Forschungslinie I* wurde untersucht, ob Erinnerung an soziales Verhalten anderer ein Gruppenphänomen ist. Zwei Studien zeigten den Einfluss von Kategorisierung in minimale Gruppen auf das Gedächtnis für unkooperative (d.h. unfaire oder betrügerische) Personen. Unkooperative Eigengruppenmitglieder wurden besser erinnert als andere Gruppenmitglieder oder Fremdgruppenmitglieder. Im Gegensatz dazu wurden unbekannte Eigengruppenmitglieder eher als kooperativ eingeschätzt als Fremdgruppenmitglieder. Eine dritte Studie zeigte, dass ein natürlicher und salienter Gruppenkontext zu einem besseren Erinnerungsvermögen an Devianten (d.h. Betrüger und besonders Vertrauenswürdige) aus der Eigengruppe als an Devianten aus der Fremdgruppe führt.

In *Forschungslinie II* wurde erforscht, ob Ärger über Devianz maßgeblich von der Anteilnahme am Leid der Opfer abhängt. Deshalb wurden emotionale Reaktionen auf die moralische Verwerflichkeit (d.h. Intention des Täters) und die Konsequenzen (d.h. Schaden an den Opfern) von Devianz untersucht. Die Intention des Täters erregte in allen drei Studien mehr Ärger als der Schaden an den Opfern. Mitgefühl mit den Opfern und Ärger über den Täter unterschieden sich voneinander, wenn die Konsequenz der Tat deutlich von der Intention des Täters abgegrenzt wurde. Die Ergebnisse zeigen, dass moralischer Ärger auch unabhängig von empathischem Ärger auftritt.

In *Forschungslinie III* wurden die Auswirkungen von gemeinsamer Gruppenzugehörigkeit mit Tätern oder Opfern von Devianz untersucht. In fünf Studien wurden unfaires Verhalten, Gruppenzugehörigkeit der Opfer und der Täter orthogonal manipuliert. Unfaires Verhalten löste in allen Studien Ärger und (altruistische) Bestrafung aus, auch wenn die Interaktionen ausschließlich zwischen Fremdgruppenmitgliedern stattfanden. In minimalen Gruppen zeigten hoch identifizierte Gruppenmitglieder härtere Reaktionen gegenüber Tätern, die Eigengruppenopfer unfair behandelten. Täter aus der eigenen Gruppe riefen in einem natürlichen konfliktbeladenen (Inter-) Gruppenkontext mehr Ärger und Bestrafung hervor, als in einem kooperativen.

Zusammenfassend bleibt festzuhalten, dass Gedächtnis, Ärger, und Bestrafung von Devianz durch eine gemeinsame Gruppenzugehörigkeit mit Tätern und Opfern verstärkt werden. Allerdings rufen moralische Vergehen auch unabhängig von persönlichem Bezug zu den Beteiligten emotionale Reaktionen hervor. Die Ergebnisse werden hinsichtlich ihrer Funktion für das soziale Zusammenleben diskutiert.

1. Introduction and Overview

1.1 General Introduction

Moral deviance elicits very strong reactions in observers. People don't want to live close to convicts, feel morally outraged, call repeatedly for justice, and approve of harsh treatment of perpetrators. Dealing with deviants is important because shared morality (i.e., prescriptive norms) specify the dos and don'ts in a society. Hence, morality prevents conflict and facilitates trust and collaboration. For example, one (most often) should not cheat, murder, or apply torture, but one should pay taxes. Shared morality often shapes and supplements legal provisions. Moral concerns manifest in people's intuitions, perceptions, and emotions with regard to their own and other's behavior. Moral deviants are harshly condemned, while deviants from descriptive norms, such as barefoot people in the winter, are less harshly treated. But when and why do we react to deviance? Do our psychological mechanisms effectively deal with such threats to group life? The current work focuses cognitive, emotional, and behavioral responses to observed moral violations. It clarifies the influence of self-involvement (i.e., shared group membership with perpetrator or victim, and empathy with victims) on reactions to deviants.

To illustrate the current approach, imagine two girls, Susan and Anna. Susan and Anna meet at the school yard. All of a sudden, Susan pushes Anna so hard that she falls and hurts her ankle. Our psychological mechanisms would work as follows: We are angry at Susan, and think that she should be punished for pushing Anna. Moreover, we will remember Susan. Maybe we will tell others about her bad behavior and avoid her next time we see her. Like this example, most research on cognitive, emotional, and behavioral reactions to moral violations concentrates on interpersonal encounters (e.g., Bell & Buchner, 2012; Darley & Pittman, 2003). Morality, however, largely operates within groups (Haidt, Rosenberg, & Hom, 2003; Leach, Bilali, & Pagliaro, 2015). If Susan and Anna are our class mates, we will be upset about the pushing. If Susan and Anna belong to different classes than us, or even different schools, we might be less concerned. How does this fit with the notion that morality provides a code of conduct that regulates social life? Under which circumstances is it important that others stick to our principles? Three lines of research try to answer these questions and extend our knowledge of psychological underpinnings of morality.

Research Line I investigates whether memory for cheaters is group-bound and thus, facilitates within-group cooperation. Considering the example, Susan might be remembered by others because her pushing was wrong. Or, Susan's bad behavior might be remembered, because she was our classmate, and therefore her pushing threatens the harmony in our class.

Research Line II disentangles moral outrage from empathic anger. It differentiates anger from other emotional reactions to moral violations. Does Susan elicit outrage, and only outrage, because she intentionally pushed Anna? Alternatively, Susan might elicit outrage because we care for Anna's wellbeing.

Research Line III examines the role of a shared group membership with perpetrator and victim for moral outrage and punishment. We might be angry and punitive towards Susan, because she is in our class, and we do not tolerate pushers in our class. Or we might not tolerate harm inflicted on our classmate Anna. If we are bothered by the wrongfulness of pushing, we might be angry and punitive when Susan and Anna both belong to a different school.

1.2 Conceptual and Theoretical Framework

1.2.1 Morality and cooperation

The social function of morality

Morality is a code of conduct that defines "good" and "bad" behavior (Beauchamp, 2001; Hauser, 2006; Leach et al., 2015). Typical examples of moral violations are cheating, disloyalty, stealing, and murder. These examples illustrate acknowledged domains of morality. Morality determines that people should not harm others and treat others fairly in social exchanges. Thus, it restricts selfish behavior for the sake of others' wellbeing, and prevents conflict and facilitates social interactions.[1] A preference for care and fairness has

[1] Moral agendas sometimes extend to other domains or contents, such as divine purity (e.g., homosexuality) or authority obedience (e.g., desertion); in other cases, killing is perceived as moral (e.g., honour killings, human sacrifice). These moral agendas are applied by certain groups and are often religiously motivated (Haidt, 2007). Other authors suggest that moral rules always aim at preventing harmful acts in the widest sense (Gray, Young, & Waytz, 2012). In the course of the present work, I will concentrate on moral violations in the care and fairness domain.

been suggested to derive from basic human impulses (Lerner, 1980; Tyler & Blader, 2000).

Morality guides the formation of attitudes, values, and behavior (Beauchamp, 2001); and is thus a short-cut for decision-making. Behaving morally contributes to a positive self-concept and a positive ingroup concept (Cialdini, Reno, & Kallgren, 1990; Leach, Ellemers, & Barreto, 2007; Monin & Jordan, 2009). Additionally, people evaluate and judge others' behaviors, attitudes, and values according to how moral they are, which includes their trustworthiness, honesty, and sincerity (Brambilla & Leach, 2014; Pagliaro, Ellemers, & Barreto, 2011). Morality regulates social interactions in interpersonal encounters and groups because it sets a standard for the evaluation of social behavior (Baumard, André, & Sperber, 2013; Ellemers, Pagliaro, & Barreto, 2013; Haidt, 2008; Rai & Fiske, 2011). The belief that (most) people behave morally enables one to trust others in social interactions. In sum, morality facilitates stable and mutually beneficial relations, harmonious social life, and the maintenance of relationships.

In contrast to social conventions, people perceive morality as unalterable, universal, and obligatory (cf. Darley & Shultz, 1990; Haidt, 2001; Mikhail, 2007; Piaget, 1932; Shweder, Mahapatra, & Miller, 1987; Skitka, 2010; Turiel, 1983; Turiel, Killen, & Helwig, 1987). Developmental psychologists suggested that from adolescence, universal ethical principles guide one's own behavior and the evaluation of others' behavior (e.g., Kohlberg, 1976; Piaget, 1932; Turiel, 1983). They found that adolescents apply such moral guidelines independently of authorities, conventions, or concern for the self. People experience their moral convictions to be independently from external sources, such as social contracts or authorities; and they perceive their moral convictions as universally valid and self-evident (Skitka, 2010). Even though the content of morality varies cross-culturally, people are intolerant of moral diversity (Haidt et al., 2003). Thus, morality is psychologically different from other norms, such as descriptive norms and attitudes. For example, in our society murder is immoral, independent of where and whom you murder. In contrast, we can readily accept that in some places only driving on the right side of the road is acceptable. This suggests that any norm may be seen as moral if it is highly valued and rewarded, or/ and, whose violations cause strong reactions (Harms & Skyrms, 2008; Kessler & Cohrs, 2008).

Evolutionary approaches to morality

Since Darwin (1871/1901), evolutionary approaches suppose that organisms are mainly concerned with the management of their own fitness. This main assumption provides a framework for explaining why organisms develop some features over others. Features that enhance fitness are adaptive and outlast those that are not. According to evolutionary psychology (e.g., Buss, 2015; Caporael, 2001), what people think, feel and do, is a product of natural selection. Morality has long been a challenge to this view, because it restricts immediate self-benefit for the benefit of others. The current mainstream in moral psychology considers morality as a set of (psychological) skills that evolved to foster cooperation (Greene, 2013; Haidt & Kesebir, 2010; Tomasello & Vaish, 2013). Cooperation facilitates the provision of certain resources, such as hunting big animals, farming, building houses, and fighting enemies. This ultimately provides mutual benefits for interaction partners. Competitive individual strategies (e.g., increasing personal benefit at the expense of others') must be restricted for successful cooperation. To promote the evolution of skills of cooperation different strategies have been suggested (e.g., Nowak, 2006).

First, *reciprocal altruism* complies with the rule: "I scratch your back, you scratch mine". Thus, the support of others (even with personal costs) is equally supported by the interaction partner (Trivers, 1971). In a two person interaction, the *tit-for-tat* strategy can produce stable cooperation (Axelrod & Hamilton, 1981). Second, social interactions are not necessarily repeated interactions, and sometimes a different person may repay the efforts. "I scratch your back, you scratch Peter's back and Peter scratches my back" is termed *indirect reciprocity* (Nowak & Sigmund, 2005). Tracking accounts of moral behavior or cooperation of all available interaction partners is important to uphold indirect reciprocity. This becomes increasingly challenging with the increasing amount of available interaction partners. Third, *reputation-based cooperation* also works in groups where repeated face-to-face interaction is less common: "I am known to scratch others' backs, therefore I get my back scratched" (Nowak & Sigmund, 1998). Fourth, especially in large populations, the likelihood to interact with some people (such as co-workers or neighbors) is higher than with others (*network cooperation*; Ohtsuki, Hauert, Lieberman, & Nowak, 2006). Therefore, maintaining cooperative relationships with some people is more important than with others. The organization of network clusters for cooperation and mutual help increase potential benefits, as long as the benefit-to-cost ratio exceeds the number of individuals in the network. Last, *group selection* (or *multilevel selection*) might foster cooperation within one's group, even when cooperative behavior decreases the immediate benefit of the cooperator. Cooperative groups might outcompete less cooperative groups in intergroup conflicts. Moreover, regulated management of shared resources fosters group survival (e.g., Brewer & Caporael, 2006; Wilson & Wilson, 2010).

Psychological devices should foster one or several of the proposed strategies that grant benefits from cooperation. As mentioned, explicit contracts, such as the law, as well as implicit codes of conduct, such as morality, foster cooperation. Indeed, people most often intuitively act cooperatively across a variety of social situations. For example, stable cooperation emerges in repeated interactions between various players in economic games (Fehr & Fischbacher, 2004a), and among populations of indirect reciprocators (Nowak & Sigmund, 2005). In sum, it is assumed that psychological features of morality evolved, because social interactions and group life ultimately provides benefits of cooperation. However, peoples' cooperative tendencies differ individually, and cooperative behavior bears the risk of exploitation.

1.2.2 Dealing with cheaters

The cheater's benefit

According to Cosmides and Tooby (1992), who shaped the concept of cheating, "a cheater is an individual who illicitly benefits himself or herself by taking a benefit without having satisfied the requirement …" (p. 180). Unconditional cooperators can be easily exploited by cheaters. Unilateral defection (cheating or non-cooperation) pays, because the cheater keeps their contribution and profits from other's cooperative behavior. The co-operator loses both their own and the other's potential contribution. Despite the risk of exploitation, people withhold high levels of cooperation (Boyd & Richerson, 2009). In a population of cooperators only a certain amount of cheaters can be tolerated. The benefits of cooperation, and thus cooperative tendencies within a population, decline when cheating is frequent (Brewer & Caporael, 2006; Fehr & Gächter, 2002; Kerr et al., 2009; Yamagishi, 1986).

As cheating is tempting but threatens cooperative tendencies of others, cheaters have to be dealt with. Thus, to maintain cooperation, efficient selection or control of interaction partners is required (Noë & Hammerstein, 1994; Peck, 1993; Rockenbach & Milinski, 2006; Trivers, 1971). When cheaters are reliably detected and avoided or punished in future interactions, they will not be successful (Baumard et al., 2013; Cosmides & Tooby, 1992; Fehr & Fischbacher, 2004a). Moreover, any behavior that is constantly and predictably punished by social exclusion or other costs will be less often performed (Boyd & Richerson, 1992; Fehr & Fischbacher, 2004b). As morality specifies requirements of social life beyond (explicit) exchange rules, it is suggested that moral violations require appropriate treatment (for a review, see Haidt & Kesebir, 2010).[2]

Efficient selection of interaction partners

Cooperators benefit when they approach the cooperative and avoid the uncooperative interaction partners. Efficient partner choice excludes cheaters from beneficial social interactions (Baumard et al., 2013) or from the moral community (moral exclusion; Staub, 1990); therefore, it is important to effectively detect cheaters (Cosmides & Tooby, 1992, 2008). People predict other's uncooperative tendencies in an interaction better than chance (Frank, Gilovich, & Regan, 1993; Little, Jones, DeBruine, & Dunbar, 2013; Verplaetse, Vanneste, & Braeckman, 2007). The selection of business partners, friends, or sexual partners is based on cues of fairness and trustworthiness (Swann, 1987). People also readily gossip about others' social behavior. This enables reputation-building that is independent from personal face-to-face interaction (Dunbar, 2004; Dunbar, Marriott, & Duncan, 1997).

In addition to appearance-based cues, behavior could be used as indicator for person's cooperative tendencies. It has been suggested that violations of social contracts are detected more efficiently than violations of other conditional rules (Cosmides, 1989; Fiddick & Erlich, 2010).[3] People quickly and intuitively judge others' behaviors to be moral or immoral, independently of their personal involvement. Moral violations cause automatic evaluations and negative affect in observers (Haidt, 2001). Such intuitive reactions emerge primary to deliberate reasoning about moral judgments (for reviews, see Greene & Haidt, 2002; Haidt, 2007). The social-functionalist approach further suggests that people seek to detect and judge other's wrongdoing like "intuitive prosecutors" (Tetlock, 2002); they overestimate features of perpetrators that indicate the likelihood of future wrongdoings, such as accountability

[2] From the theoretical point of view of this work social norms, morality, and social exchange rules are hard to distinguish. I will therefore adhere to the following definitions in the course of this thesis: *deviants* violate any social norm (e.g., driving on the wrong side of the road), *perpetrators* are deviants who commit a (intentional) moral violation (e.g., murder), and *cheaters* or *non-cooperators* are perpetrators who violate (explicit or implicit) rules of cooperation and fairness (e.g., shop-lifting).

[3] The appropriateness of the Wason Selection Task for cheater detection is critized. See, for example: The solution of the Wason selection task is fostered by perspective taking (Gigerenzer & Hug, 1992), relevance-guided comprehension (Sperber & Girotto, 2002), and correct text processing (Almor & Sloman, 2000).

(Tetlock et al., 2007) or intentions (Caruso, Waytz, & Epley, 2010; Falk, Fehr, & Fischbacher, 2008; Falk & Fischbacher, 2006; Waytz, Gray, Epley, & Wegner, 2010). Moreover, blameworthy action is often perceived as intentional, even when it is not (Knobe, 2003; Rosset, 2008).

Once detected, remembering cheaters and their behavior facilitates the selection of interaction partners. Indeed, people must remember what a particular person has done previously in order to avoid exploitation in future interactions. There are mixed findings on whether cheaters' faces are better recognized than other faces (Barclay, 2008; Barclay & Lalumière, 2006; Chiappe, Brown, Dow, Koontz, Rodriguez, & McCulloch, 2004; Mealey, Daood, & Krage, 1996; Mehl & Buchner, 2008; Oda, 1997; Rule, Slepian, & Ambady, 2012; Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003). However, it was found repeatedly that cheaters (the persons and their behavior) are remembered better than irrelevant or trustworthy persons (Bell & Buchner, 2010b, 2012; Buchner, Bell, Mehl, & Musch, 2009). The memory for non-cooperators leads to distrust in repeated encounters (Oda & Nakajima, 2010; Wilkowski & Chai, 2012). The enhanced memory for cheaters (or non-cooperators) is driven by general memory mechanisms (e.g., Bell & Buchner, 2012). First, memory is better for relevant information, for example through self-involvement in cheating or justice sensitivity (Bell & Buchner, 2010a; Bell, Giang, & Buchner, 2012). Second, information that violates expectations is remembered especially well, including persons associated with cheating, disgusting, or rare behavior (Bell & Buchner, 2010b, 2012; Bell, Buchner, & Musch, 2010; Volstorf, Rieskamp, & Stevens, 2011). Nevertheless, people have a heightened memory for those behaved immoral than those who disadvantaged them personally (Bell, Schain, & Echterhoff, 2014). This indicates that threats of exploitation are not enough to explain the better memory for cheaters, but moral violations to trigger memory.

Norm enforcement through punishment

In large-scale cooperation the management of reputations is difficult. Nevertheless, people frequently cooperate with strangers in non-repeated encounters (Boyd & Richerson, 2009; Tomasello & Vaish, 2013). Therefore, punishment has been suggested to play a crucial role in the maintenance of cooperation (e.g., Boyd & Richerson, 1992). Punishment, in contrast to merely hard treatment, implies the disapproval of certain behavior (J. Feinberg, 1965). The option to punish, or merely provide feedback on other's behavior, in common good games increases cooperative behavior of interaction partners (Dawes, McTavish, &

Shaklee, 1977; Fehr & Fischbacher, 2004a). Further, people even prefer groups that offer the option to apply costly punishment over groups in which no punishment option is available (Gürerk, Irlenbusch, & Rockenbach, 2006). Individual gain from stable cooperation in large groups outweighs the costs of punishment when all potential interaction partners equally apply punishment (Boyd, Gintis, & Bowles, 2010). This indicates that punishment fosters the perpetuation of cooperation, or even any social norm that is worth upholding (Boyd & Richerson, 1992; Chudek & Henrich, 2011); thus punishment may re-establish moral standards after violations (Davis, 1949; Tyler & Boeckmann, 1997; Vidmar, 2001).[4]

In line with this argument, players frequently punish cheaters in cooperative games (for an overview, see Guala, 2012). They even invest their own resources to punish without self-serving benefits (referred to as *altruistic punishment*). Altruistic punishment has been observed in one-shot games (Fehr & Gächter, 2002), and on behalf of others (Fehr & Fischbacher, 2004b; Gintis, Bowles, Boyd, & Fehr, 2003; Henrich et al., 2006). Fehr and Gächter (2002) suggested that emotional commitment to cooperation triggers punishment in spite of self-interest. Indeed, anger has been found to mediate the relation between unfairness and third-party punishment (Fehr & Gächter, 2002; Nelissen & Zeelenberg, 2009; Seip, Dijk, & Rotteveel, 2014). Moreover, people who report more anger about unfair offers in ultimatum games are more likely to reject the offer. Consequently, neither of the players receives an incentive (Yamagishi, Horita, Takagishi, Shinada, Tanida, & Cook, 2009).

Next to cheating, other moral violations have been suggested to elicit anger and motivate punishment (Darley & Pittman, 2003; Haidt, 2003; Montada & Schneider, 1989). The intuitive reaction to a moral violation (i.e., "this is just wrong") is accompanied by negative affect (Haidt, 2001).[5] Unfairness and harm provoke anger (Gutierrez & Giner-Sorolla, 2007; Haidt, 2003; Hutcherson & Gross, 2011; Rozin, Haidt, McCauley, Dunlop, & Ashmore, 1999), which motivates people to take action, such as punish the perpetrator (e.g., Frank, 1988; Frijda, 1988). The bystanders' anger also predicts their tendency to communicate their disapproval to the perpetrator (i.e., social control; Chaurand & Brauer, 2008). "Righteous" punishment even reduces anger and increases observer's satisfaction

[4] Rewards are as effectively as punishment in fostering cooperation (Balliet, Mulder, & Van Lange, 2011). Thus, instead of the infliction of negativity on perpetrators any signal for acceptable behavior might foster cooperation.

[5] Haidt and Kesebir (2010, p. 802) define the relation as follows: "…moral intuitions … are a subclass of automatic processes that always involve at least a trace of 'evaluative feeling'. Moral intuitions are about good and bad. Sometimes these affective reactions are so strong and differentiated that they can be called moral emotions, such as disgust or gratitude, but usually they are more like the subtle flashes of affect that drive evaluative priming effects."

(Goldberg, Lerner, & Tetlock, 1999). People prefer to punish perpetrators in a morally proportional way (*just desert*) over incapacitation or deterrence (Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002; Darley, Carlsmith, & Robinson, 2000). They seem to aim at restoring a "moral balance" rather than preventing future offenses. Thus, the psychological machinery might operate detached from distal utilitarian goals, such as the enforcement of cooperation.

1.2.3 Group-bound cooperation

Interdependence within groups

The notion that psychological devices regulate social interactions does not mean that we expect everybody to behave similarly. People live in social groups, which largely define our interaction partners (Brewer, 2004, 2007; Tooby & Cosmides, 2010). Group members engage in joint action in order to accomplish common goals (stag hunt, care of the elderly, street pavement, etc.), or share scarce resources (clean water, sources of electrical energy, etc.; cf. Hardin, 1968). In contrast to social categories, which are defined only by a shared feature, social groups act as entities (Campbell, 1965; Rabbie & Horwitz, 1988). They share group-specific norms, values, and traditions that facilitate coordination and cooperation (Sherif, 1936; Terry & Hogg, 1996; Turner, 1985). Such commonalities account for ingroups, but often differ for outgroups. Likewise, morality is shared within groups, but differs across cultures (Graham & Haidt, 2010; Haidt et al., 2003; Leach et al., 2015).

Realistic Group Conflict Theory (Campbell, 1965; Sherif & Sherif, 1953; Sumner, 1906) suggests that group dynamics evolve because groups are in competition for resources and have incompatible goals. This causes ingroup members to dislike outgroup members and accentuate intergroup differences, and simultaneously behave favorably towards ingroup members (also referred to as ethnocentrism or ingroup bias). Favorable ingroup behavior increase group efficiency and/or strength, and guarantee mutual benefits of cooperation within the group. In this sense, a group is characterized by the (perceived) positive interdependence of its members. Therefore the Theory of Realistic Group Conflict is an extension of Interdependence Theory for group processes (cf. Kelley & Thibaut, 1978; Rabbie & Horwitz, 1988). Perceived threat, such as intergroup competition, even increases ingroup support, cooperation and efficacy (Benard, 2012; Bornstein, Gneezy, & Nagel, 2002; Fritsche, Jonas, & Kessler, 2011). Intergroup conflict even increases cohesion in large social groups (Brewer

& Campbell, 1976; LeVine & Campbell, 1972; Sumner, 1906). In times of war, ingroup solidarity and violence against ingroup treachery increases (Gould, 1999). Still today, we experience that citizens emphasize their national identity more eagerly during international conflicts or sports competitions. Dissidents are strongly rejected, punished, and sometimes even banned or executed. This strong commitment facilitates collective action that protects the ingroup.

Ingroup cooperation

As suggested by the Interdependence Theories in group psychology, various studies illustrate that a shared group membership facilitates cooperation, whereas members of different groups rarely collaborate. In cooperative encounters group members coordinate personal investments to acquire mutual benefits. People adapt ingroup norms, even when they oppose their personal view (Asch, 1956; for a review, see Cialdini & Goldstein, 2004). The knowledge of interacting with a fellow group member increases effort, expenditure and the success of coordination (Mehta, Starmer, & Sugden, 1994). Conversely, coordinated synchrony between persons increases liking and subsequent cooperation (Wiltermuth & Heath, 2009).

Group members evaluate and treat fellow group members more positively than outgroup members, even in large-scale groups (for a review, see Hewstone, Rubin, & Willis, 2002). Individuals reliably cooperate with ingroup members, and less with outgroup members (e.g., Kramer & Brewer, 1984). Indeed, high levels of cooperation are upheld in many different cultures (Henrich, 2004; Hill, 2002). In social interactions people expect mutual trust and support within the group (Brewer, 2007; Yamagishi, Jin, & Kiyonari, 1999). They preferably select interaction partners from their own group than an outgroup (Foddy, Platow, & Yamagishi, 2009; Sober & Wilson, 1998). A meta-analysis of 212 studies shows that people generally treat ingroup members favorably and discriminate against outgroup members in experimental games (Balliet, Wu, & De Dreu, 2014). The authors further found that this tendency increases with the number of repeated interactions between the partners and the common knowledge of group membership. Most natural groups have a social markers, preferably one that is hard to fake to illustrate shared group membership (Cohen, 2012). In sum, a shared group membership seems a reliable predictor for successful cooperation.

Dealing with ingroup deviants

Ingroup deviants threaten efficient and beneficial group behavior, and thus must be dealt with to maintain cooperation. This shows, for example, by punishment of ingroup cheaters being more frequent in high-trust than in low-trust societies (Balliet & Van Lange, 2013). In public good games, group members apply more third-party punishment to ingroup non-cooperators than to outgroup non-cooperators (Shinada, Yamagishi, & Ohmura, 2004). This ingroup punishment, but not outgroup punishment, correlated positively with anger at non-cooperators. Likewise, an unfair offer from within the group is more likely to be rejected than an unfair offer between groups (Mendoza, Lane, & Amodio, 2014). Moral violations are also punished more harshly when committed by ingroup members as they pose a moral threat to the group: Ingroup perpetrators elicit the strongest reactions when they have full knowledge about the ingroup norms, can be fully blamed and assigned guilt, and/or commit the offenses repeatedly (Gollwitzer & Keller, 2010; Kerr, Hymes, Anderson, & Weathers, 1995; Pinto, Marques, Levine, & Abrams, 2010; Taylor & Hosch, 2004; van Prooijen, 2006).

As most interaction takes place within groups, an ingroup perpetrator is likely to affect an ingroup victim. This indicates that punishment of ingroup perpetrators might emerge on behalf of fellow group members. In line with this argument, people are more willing to invest in punishment of a perpetrator who harmed family members or schoolmates in comparison to strangers (Lieberman & Linke, 2007, Study 3). Punishment of ingroup perpetrators might aim at protecting fellow group members, instead of protecting ingroup norms. Bernard, Fehr and Fischbacher (2006) investigated reactions of tribal members (Papua New Guinea) to ingroup and outgroup perpetrators who behaved unfairly towards ingroup and outgroup victims. They found that altruistic punishment was applied on behalf of ingroup victims, regardless of the perpetrator’s group membership. The same study was conducted with groups of Swiss soldiers in third-party-punishment prisoner’s dilemmas. Again, altruistic punishment was increased for offenses against ingroup members (Goette, Huffman, & Meier, 2006).

1.2.4 Psychological meaning of group membership

The selfless care for fellow group members illustrates that group-favoring behavior does not always pay immediately. Moreover, in anonymous or large-scale groups the risk of exploitation is extremely high; however, people still engage in group favoring behavior even with personal costs. This raises the question of how such ingroup biases emerge on an individual level when self-interest is not present or possible. The Social Identity Approach (e.g., Tajfel & Turner, 1979) offers a psychological explanation for group dynamics. It captures the subjective experience of being a group member, its antecedents, and its consequences.

Ingroup identification

Studies on the Minimal Group Paradigm (Tajfel, Billig, Bundy, & Flament, 1971) demonstrated that mere knowledge about one's group membership elicits ingroup favoring and outgroup discriminating behavior. Group members automatically evaluate ingroup members more favorably than outgroup members, even in the absence of intergroup competition (e.g., Brewer, 1979; Mullen, Brown, & Smith, 1992; Otten & Wentura, 1999; Perdue, Dovidio, Gurtman, & Tyler, 1990). These findings indicate that belonging to one group and not another suffices for ingroup favoritism to emerge. Through self-categorization group members "perceive themselves more as the interchangeable exemplars of a social category than as unique personalities …" (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987, p. 50). A salient ingroup leads to the group becoming part of the self (E. R. Smith & Henry, 1996), and as part of the self, an ingroup is perceived as generally positive (Brewer, 1999; Gramzow & Gaertner, 2005; Perdue et al., 1990). Identification with a group adds personal meaning and emotional involvement to group membership (Tajfel & Turner, 1979; Turner et al., 1987). In other words, a group member is psychologically attached to their ingroup.

With ingroup identification ingroup norms gain importance (Terry & Hogg, 1996). Group-specific attitudes and behavior (i.e., group norms), constitute the group prototype. The prototype is the mental representation of how group members behave and interact (or ought to interact) with each other. What is perceived as the ingroup prototype derives from a meta-contrast between the ingroup and the outgroup (Turner et al., 1987). The meta-contrast increases similarity within the group and differences between the groups. Within the ingroup, prototypicality provides personal guidelines and structure that fosters intuitively-correct behavior. The perceived consensus validates the personal interpretation of situations and the ingroup norms (*normative fit*). Normative fit reduces group-based or personal uncertainty that emerges, for example in threatening situations (J. R. Smith, Hogg, Martin, & Terry, 2007). The meta-contrast also fosters development and maintenance of a positive ingroup image (*comparative fit*). A positive ingroup image is related to increased self-esteem. As the outgroup serves as a reference group, evaluations often produce ingroup biases (Hogg & Abrams, 1988; Houston & Andreopoulou, 2003). Prototypicality is also important for the evaluation of fellow ingroup members (Brewer, 1999; Tajfel & Turner, 1979). A normative group member is granted full membership, represents an attractive interaction partner, possesses social influence, and is endorsed as a group leader (Hogg, 2001; Steffens, Haslam, Ryan, & Kessler, 2013). Moral behavior is often a crucial dimension of intergroup comparison, because morality contributes to maintaining a positive group image (Ellemers et al., 2013).[6] Moral behavior within the ingroup elicits more positive evaluations of the group than other dimensions of normative behavior, such as competence or sociability (Leach et al., 2007). Adherence to shared morality is also important for gaining ingroup respect (Pagliaro et al., 2011).

Group members are often internally motivated to adhere to group norms, even when they conflict with self-interest (see Section 1.2.3). This personal commitment derives from group salience and ingroup identification (Brewer & Kramer, 1986; Kerr, Garst, Lewandowski, & Harris, 1997; Kerr & Kaufman-Gilliland, 1994; Wenzel, 2004; Wit & Kerr, 2002). The psychological attachment to the ingroup binds ingroup members to mutual cooperation (Brewer, 1999, 2007; Haidt & Kesebir, 2010; Seewald, Hechler, & Kessler, 2016). Highly identified group members contribute more to the common good and restrain their consumption of shared resources more than low identified group members (e.g., Brewer & Kramer, 1986; De Cremer & van Vugt, 1998, 1999; Kramer & Goldman, 1995). Highly identified group members also behave more pro-socially towards fellow group members in interpersonal encounters, and thereby discriminate more between groups (for an overview, see Spears, Jetten, Scheepers, & Cihangir, 2009; Tajfel et al., 1971). Thus, identification with a group easily establishes and maintains cooperation because it fosters the adherence to ingroup norms and the favorable treatment of fellow ingroup members.

Social identity and deviance

On a psychological level, a positive social identity provides positive feedback to the self, such as “I fit with my positive group”. Whereas Interdependence Theories (Rabbie & Horwitz, 1988)suggest that ingroup cheaters threaten personal benefits of cooperation, any deviant (i.e., non-prototypical group member) threatens one’s positive social identity. Ingroup deviants cast doubt on the validity of ingroup norms and/or the positive distinctiveness of the

[6] As mentioned previously and according to this argumentation, any social norm could become “moral” (and general) with increasing importance and its attached emotional. However, central group norms are not always moral, and ingroup identification fosters the importance of salient differentiation norms over general norms (Jetten, Spears, & Manstead, 1997)

ingroup. Therefore, people derogate ingroup deviants stronger than outgroup deviants, whereas they regard normative ingroup members as more positive than normative outgroup members (Marques & Paez, 1994). This so-called Black Sheep Effect has been mainly observed for negative deviants in salient intergroup contexts (for an exception, see Abrams, Marques, Bown, & Henson, 2000, Study 1; e.g., Marques, Abrams, & Serôdio, 2001; Marques, Yzerbyt, & Leyens, 1988). The clear differentiation between "bad" and "good" group members has been suggested to reassume the group consensus and foster the positive ingroup bias (Marques et al., 2001). Ingroup norms and intergroup differentiation are more important to highly identified group members; thus, the relative derogation of ingroup deviants increases with ingroup identification (e.g., Abrams et al., 2000; Marques et al., 1988). The Black Sheep Effect might even bolster social identity, as it strengthens ingroup identification (Marques, Abrams, Paez, & Martinez-Taboada, 1998). When group members are not given the chance to devaluate, they disidentify from their ingroup (Eidelman & Biernat, 2003). Moreover, ingroup deviants from generic norms are more harshly derogated and punished than outgroup deviants, because they diffuse the group's moral standing (e.g., Abrams et al., 2000; Hutchison, Abrams, Gutierrez, & Viki, 2008; Marques et al., 2001). The negative judgment of ingroup deviants comes is accompanied by harsh punishment (Abrams et al., 2000). This has been illustrated for exertion of social control towards deviants who disobey public rules (e.g., littering; Nugier, Chekroun, Pierre, & Niedenthal, 2009); and the high tendency to exclude deviants from the ingroup when the chance is given (Eidelman, Silvia, & Biernat, 2006).

Care for fellow ingroup members

The Social Identity Approach (Tajfel & Turner, 1979) suggests a second reason for negative reactions to deviants. As mentioned, it considers the group as part of the self. The social self includes fellow ingroup members, and therefore is concerned with their wellbeing (Yzerbyt, Dumont, Gordijn, & Wigboldus, 2002). In fact, group members experience group-based emotions on behalf of the group or its members (Mackie & Smith, 2002; E. R. Smith, 1993). Group-based emotions are stronger with increasing ingroup identification, but also ingroup identification enhances with the experience of group-based emotions (Mackie & Smith, 2002; Mummendey, Kessler, Klink, & Mielke, 1999; Walker & Pettigrew, 1984). As emotions motivate action, group-based emotions can elicit behavior on behalf of the group. Most famously, collective deprivation elicits collective action in order to change unfavorable ingroup conditions (Mackie, Devos, & Smith, 2000; Mummendey et al., 1999; Runciman, 1966; van Zomeren, Spears, Fischer, & Leach, 2004).

Group members also react strongly to offenses against fellow group members in which they are not personally involved. A shared group membership with the victim triggers outrage at moral violations more than an outgroup victim (Batson, Chao, & Givens, 2009; Gordijn, Wigboldus, & Yzerbyt, 2001; Yzerbyt, Dumont, Wigboldus, & Gordijn, 2003). Contextual category salience and ingroup identification increase group-based anger and subsequent action tendencies (Gordijn, Yzerbyt, Wigboldus, & Dumont, 2006; Yzerbyt et al., 2003). People are motivated to punish perpetrators who affect fellow ingroup members (Batson et al., 2009; Gordijn et al., 2006). In sum, anger about offenses and punishment of perpetrators protects the interest of ingroup victims over outgroup victims.

1.3 Overview of the Present Research

The current dissertation project aims to investigate the influence of perpetrator and victim group membership and (im-)moral intentions on third-party reactions to deviants. People show strong reactions to deviance, often entailing the exclusion or harsh treatment of perpetrators. Three research questions specify what triggers such intense cognitive, emotional and behavioral responses: I) Does memory for deviants differ in group contexts?, II) Is moral outrage elicited by the wrongfulness of an action or the suffering of victims?, and III) What role does self-involvement with the victim or perpetrator play in the anger at and punishment of deviants?

The hypotheses are based on three theoretical assumptions. First, morality fosters mutual benefits, such as those obtained by cooperation. The detection, memory, and (third-party) punishment of deviants enforces desired behavior (cf. Section 1.2.1 and 1.2.2; Cosmides & Tooby, 1992; Fehr & Fischbacher, 2004b). Second, people mostly interact within groups. They expect favorable treatment within groups, but not between groups (cf. Section 1.2.3; Balliet et al., 2014; Kramer & Brewer, 1984). Third, ingroup identification is a psychological device for positive ingroup interactions. It increases the importance of normative behavior within groups and the care for fellow ingroup members (cf. Section 1.2.4; Brewer, 2007; Gordijn et al., 2001; Tajfel & Turner, 1979). Three lines of research were developed and tested.

Research Line I transferred memory for uncooperative individuals into an ingroup/outgroup context. It was expected that memory is enhanced for uncooperative ingroup members, but not for uncooperative outgroup members. Memory for a person's uncooperative, cooperative, or neutral behavior (also referred to as source or reputational memory) is better for information that violates expectations about the targets (e.g., Bell & Buchner, 2009; Buchner et al., 2009). Ingroup contexts so far have not been addressed in such memory paradigms. Ingroup contexts improve memory for person information (e.g., Brewer, Weber, & Carini, 1995; Howard & Rothbart, 1980; Schaller & Maass, 1989). The present studies advance existing research on memory for ingroup and outgroup members with methodological rigor by applying multinomial source models (Bayen, Murnane, & Erdfelder, 1996). This procedure allows determining memory for individual's attributes independently from other processes (i.e., old-new discrimination; guessing biases). At first glance, memorizing more uncooperative ingroup members than cooperative ones may seem inconsistent with the frequently observed phenomenon of positive ingroup bias. However, group members strive to maintain a positive ingroup image (Tajfel & Turner, 1979). Such positive ingroup assumptions might appear as *guessing bias* (e.g., in contrast to "them", "we" are likely to be cooperative), even though individual uncooperative group members become infamous.

Two studies examine the effect of minimal group categorization (Tajfel et al., 1971) on memory for uncooperative individuals. Participants evaluated several uncooperative and cooperative (and neutral) ingroup and outgroup members. Uncooperative behavior was manipulated in terms of unfair decisions in a dictator game (Study 1), and behavioral descriptions (Study 2). In a surprise memory test, participants were required to recognize target faces and recall their behavior. Study 3 tested the hypothesis that a meaningful natural group context (ingroup identification) and concern for deviants (right-wing authoritarianism) interact with memory for ingroup and outgroup members. University affiliation differentiated targets' group membership and descriptions of students' norm-violations indicated uncooperativeness.

Research Line II tested the assumption that intentionally committed harmful acts elicit moral outrage independent of their consequences. Moral violations were suggested to elicit moral outrage regardless of self-involvement (cf. Section 1.2.1; Haidt, 2003; Montada & Schneider, 1989). Batson (2011) argued that this anger emerges because the consequences affect the observer, either personally or via empathy with suffering victims. Previous studies testing this assumption (e.g., Batson et al., 2009; Batson et al., 2007), however, did not distinguish clearly between a moral violation and its consequences. In this line of research a new perspective was taken to disentangle moral outrage and empathic anger. Three studies orthogonally crossed the wrongfulness (i.e., perpetrator's intentions) and the harmfulness (i.e., consequences) that are implied in moral violations. This design enabled a comparison of anger at the perpetrator and empathy with victims when the occurrence of bad intentions and damage diverged (failed attempt, accidental damage). Study 1 examined anger and punishment towards team sport members who violated fair play rules. It was conducted with sports team members in their natural environment. Study 2 and 3 focused on strong moral violations to elicit intuitive moral affect, i.e. murdering innocents. Whereas Study 2 described the perpetrator's intentions before the consequences, Study 3 described consequences before intentions. Both studies were conducted with student samples from different universities.

In *Research Line III*, it was proposed and tested whether moral violations elicit anger and punishment irrespectively of self-involvement with the perpetrator or the victim through shared group membership. Punishment of outgroup deviants who affect outgroup members cannot be explained through self-involvement. Moreover, the role of perpetrator and victim group membership so far has not been systematically studied. Most studies either focus on the group membership of perpetrators (e.g., Abrams et al., 2000; Shinada et al., 2004), or victims (e.g., Batson et al., 2009; Gordijn et al., 2001) of moral violations. As interaction mainly takes place within groups, ingroup perpetrators are likely to affect ingroup victims. A full design is needed to differentiate reactions to ingroup perpetrators and ingroup victims. Following Bernard, Fehr and Fischbacher (2006), the present studies orthogonally cross the occurrence of deviance (fair/unfair), perpetrator group membership, and victim group membership (ingroup/outgroup). This design was applied to different group contexts and experimental approaches. The studies address the role of ingroup identification, as it increases derogation of ingroup deviants (Hutchison & Abrams, 2003; Marques et al., 1988), triggers emotions on behalf of fellow group members (Yzerbyt et al., 2003), and modifies the meaning of intergroup relations (Messick & Mackie, 1989). Study 1a, 1b and 2 investigate the role of minimal group categorization to reactions to fair and unfair dictator decisions. Study 1a and 1b focused on altruistic punishment and Study 2 on anger. Two subsequent studies investigated the design in scenarios with natural large-scale groups. Study 3 presented a cooperative (Germany/France), and Study 4 a conflictive intergroup context ("Islamic State"/Western societies). Study 3 presented police officers that treated tourists fairly or unfairly. Study 4 described Secret Services applying torture or offering a fair trial to a prisoner. The diverse approaches allow moral outrage to be disentangled from group-based concerns across a variety of group situations.

Together, the three lines of research investigate which situations trigger memory, anger, and punishment of perpetrators of moral violations. They illustrate via different approaches how the group-bases of morality and norms might shape reactions to deviants. On the up-side, morality facilitates social coordination and cooperation. On the down-side, deviants are frequently targets of hard treatment, exclusion, and other forms of punishment. Therefore, it is important to investigate the circumstances under which deviants receive special attention, and trigger negative reactions.

The three studies in *Research Line I* were conceptualized and conducted as part of the Research Project “Cooperation in Social Groups”. The research project was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Research Unit Person Perception (KE 792/4-1). Principal investigators of the project were Prof. Dr. Thomas Kessler and Prof. Dr. Franz Neyer. They were involved in the conceptualization and interpretation of the present studies. The author was responsible for extended developments in theorizing and data interpretation, programming, as well as data collection, and analysis. Section 2 constitutes two research manuscripts, which are co-authored by Franz J. Neyer and Thomas Kessler. The co-authors contributed to the framing of theory and discussion, and made stylistic improvements to the manuscript. One manuscript has been accepted for publication in a peer-reviewed journal, the second is currently under revision. *Research Line II* and *III* were conceptualized in collaboration with Thomas Kessler.

Research Line I

2. The (In) Famous Among Us: Memory for deviant group members

2.1 Introduction

A common group membership facilitates successful coordination because it raises expectations of fellow ingroup members' behaviors (Mehta et al., 1994). It elicits mutual trust between interaction partners and facilitates cooperation (e.g., Balliet et al., 2014; Brewer & Caporael, 2006; Turner, 1982). Uncooperative group members exploit cooperative tendencies within groups. However, ingroup cooperation may be maintained, as long as uncooperative individuals have an infamous reputation in their group. Moreover, overly trustworthy or cooperative group members are important to remember for efficient partner selection. In the present studies, we examine whether group membership (i.e., ingroup, outgroup) modulates reputational memory (memory for the target and their behavior) for targets that deviate from such ingroup expectations. Study 1 and 2 examine how reputational memory for cheaters and group biases are modulated by minimal group membership of the uncooperative (i.e. ingroup, outgroup). Study 3 extends this research by examining reputational memory for group members' behavior in natural group context.

2.1.1 Memory for uncooperative targets

People remember uncooperative targets better than cooperative or neutral ones (e.g., Bell, Buchner, Erdfelder, Giang, Schain, & Riether, 2012; Buchner et al., 2009), and distrust them in future interactions (Oda & Nakajima, 2010; Wilkowski & Chai, 2012). Bell, Buchner, and colleagues uncovered general memory processes that account for this effect (e.g., Bell & Buchner, 2012): first, people remember socially relevant information about a person better than socially irrelevant information (Bell, Giang, et al., 2012). Moreover, there is a memory advantage for positive and negative person information compared to neutral information (Bell & Buchner, 2010b, 2011; Bell, Buchner, Erdfelder, et al., 2012). These effects can be attributed to general effects of (self-) relevance and emotional information on memory, especially if the information is threatening (e.g., Kensinger, 2007; Kensinger & Corkin, 2003; Li, Li, & Guo, 2009).

Second, behavior that violates expectations enhances reputational memory. Uncooperative behavior is remembered better than cooperative behavior if it occurs infrequently (Barclay, 2008; Bell et al., 2010; Volstorf et al., 2011). Similarly, people remember the uncooperative behavior of trustworthy-looking targets better than the uncooperative behavior of untrustworthy-looking targets (Bell, Buchner, Kroneisen, & Giang, 2012; Suzuki & Suga, 2010), because reputational memory is generally enhanced for schema-incongruent information. A schema is knowledge about a target that leads to expectations regarding the target's attributes, such as its behavior. Schematic knowledge influences reputational memory (knowing the target's attributes) and guessing (assuming the target's attributes) differently. People more accurately remember target attributes that violate the target's schema (e.g., Bell, Mieth, & Buchner, 2015; Hastie & Kumar, 1979; Hicks & Cockman, 2003; Küppers & Bayen, 2014). Guessing represents either schema-driven biases (i.e., guessing biases; Bayen, Nakamura, Dupuis, & Yang, 2000; Küppers & Bayen, 2014) or, if available, the perceived contingency between targets and attributes (Bayen & Kuhlmann, 2011; Klauer & Meiser, 2000).

In group contexts, people recognize and recall stereotype-inconsistent information more accurately than stereotype-consistent or irrelevant information, after taking guessing into account (Stangor & McMillan, 1992). Recent research on memory in group contexts has taken a closer look at individuals' behaviors whilst controlling for item recognition and guessing biases. For example, it has been observed that strong stereotypes of a target elicit enhanced memory for any exhibited traits that are stereotype-inconsistent (Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003). Stereotypical portrait pictures (e.g., of skinheads) improve memory for unexpected target behavior (Ehrenberg & Klauer, 2005). Similar results have been found in the context of gender categorization: participants remember women's behavior better when they violate stereotypes of female cooperativeness or neatness (Kroneisen & Bell, 2012). In sum, people remember schema-incongruent person behavior better than schema-congruent person behavior.

2.1.2 Memory for ingroup and outgroup information

Intergroup contexts (i.e., ingroup, outgroup) also modulate person memory. Shared categories provide the basis for differentiating between ingroup and outgroup members. Self-categorization indicates that the self belongs to one category, but not to the other (Tajfel & Turner, 1979). This enhances the relevance of fellow group members and elicits group-based expectations (Foddy et al., 2009; Gordijn et al., 2001; Terry & Hogg, 1996).

First, the greater relevance of the ingroup versus the outgroup is reflected in differential group perception and memory. The ingroup is perceived as heterogeneous, whereas outgroups are perceived as homogeneous (e.g., Boldry, Gaertner, & Quinn, 2007; S. A. Haslam, Oakes, Turner, & McGarty, 1995). Accordingly, ingroup faces are recognized better than outgroup faces (Bernstein, Young, & Hugenberg, 2007; Hugenberg, Young, Bernstein, & Sacco, 2010). In recall tasks (e.g., the "who-said-what"-paradigm), people make fewer within-group errors (assigning behavior to the wrong member within one group) than between-group errors (assigning behavior to a person of the wrong group) when group categorization is salient. In other words, people demonstrate an individualized person memory for ingroup members, while demonstrating a stronger category-based memory for outgroup members (Brewer et al., 1995; Ostrom, Carpenter, Sedikides, & Li, 1993; Ostrom & Sedikides, 1992).

Second, ingroup indicators (e.g., "we" or "us") have a positive valence (Perdue et al., 1990). Positive perceptions of the ingroup bolster the positive self-images of group members (Tajfel & Turner, 1979). Memory biases foster this ingroup favoritism in impression formation. For example, group members use abstract knowledge to make positive ingroup judgments, whereas negative group judgments are based on the retrieval of specific (negative) ingroup behaviors (Sherman, Klein, Laskey, & Wyer, 1998). In their classic studies, Howard and Rothbart (1980) showed that the members of minimal groups tend to assign correct negative information to the outgroup more frequently than to the ingroup. This is in line with ingroup favoritism. However, the authors did not differentiate between guessing and actual memory, and their findings could be attributed to guessing biases in favor of the ingroup.

Other studies have shown enhanced memory performance for violations of ingroup positivity. For example, Schaller and Maass (1989) and Gramzow et al. (2001) found that recall and group assignment of negative and self-discrepant information was more accurate for novel ingroups than novel outgroups. In sum, an ingroup context enhances person memory, because it increases a person's relevance and creates expectations of them. In contrast to prior studies, we draw on the idea that reputational memory is not simply related to recognizing a behavioral description and assigning it to the correct group. Instead, reputational memory implies that people recognize an ingroup (or outgroup) member and remember how that particular person behaved in the past. An enhanced reputational memory for schema-inconsistent behavior (i.e., uncooperative ingroup members) is not necessarily inconsistent with a positive view of the ingroup as a whole. In other words, reputational memory is based on individual observations, while guessing is based on expectations about groups.

2.1.3 Individual Differences in concerns for the ingroup

Ingroup identification. People tend to cooperate more within social groups (cf. Section 1.2.3). Hence, it is important to decide whom to approach and whom to avoid within an ingroup. Avoidance or even punishment of uncooperative group members could enhance norm adherence within the group (Baumard et al., 2013; Fehr & Fischbacher, 2004a). In group contexts, ingroup identification is an essential factor for group behavior (cf. Section 1.2.4). With stronger ingroup identification, an ingroup becomes increasingly meaningful for their members. Highly identified group members also perceive the group norms as more important (Brewer, 1979; Livingstone, Haslam, Postmes, & Jetten, 2011; Tajfel & Turner, 1979; Turner et al., 1987). Thus, highly identified are especially sensitive to norm deviation. Norms provide behavioral guidelines through ingroup consensus and validate individual behavior (Cialdini & Trost, 1998). Ingroup deviants reduce the normative fit within the group, and the ingroup's positive distinctiveness. Therefore, they threaten the validity of the group norm and group cohesion (Oakes, Turner, & Haslam, 1991; Turner et al., 1987). In salient ingroup contexts negative and positive ingroup deviants receive particular attention, are derogated compared to normative group members or even expelled from the group (Abrams et al., 2000, Study 1; Parks & Stone, 2010).

An ingroup focus has been shown to influence differential processing of ingroup and outgroup members. A salient ingroup context leads to individualized memory for ingroup information (Brewer & Harasty, 1996; Brewer et al., 1995, Study 1; Ostrom et al., 1993; Park & Rothbart, 1982). In line, highly identified group members recognize ingroup faces better than outgroup faces (Van Bavel & Cunningham, 2012). Memory advantage for stereotype-inconsistent information about the group is moderated by ingroup identification (Doosje, Spears, de Redelijkheid, & van Onna, 2007). Likewise, individual differences have been shown to modulate reputational memory for uncooperative targets, because they express differential concern (Bell & Buchner, 2010a).

Right-wing authoritarianism (RWA). In addition to variations in ingroup and norm relevance, people vary in their motivation to foster norm adherence. Right-wing authoritarianism (RWA) is usually associated with prejudice but also with punishment of threatening deviants and cheaters (e.g., Bray & Noble, 1978; McCann, 2008). RWA consists of the three facets of conventionalism, submission to authority, and aggression against minorities and deviants (Altemeyer, 1981). These three components describe a concern for conformity towards norms and a clear antipathy for deviation. Moreover, authoritarians perceive the social world as threatening and long for structure and safety (Altemeyer, 1996). They value norm compliance, legitimization of leadership decisions, and ingroup protection as coping strategies (Duckitt, 1989; Kessler & Cohrs, 2008; Van de Wetering, 1996). Most importantly, authoritarians punish deviants quite harshly. This tendency increases when their group is threatened (Feldman, 2003; Feldman & Stenner, 1997). Therefore, RWA expresses an extraordinary concern for threatening anti-normative behavior within groups. The particular focus on norm deviation and threat of authoritarians leads to the assumption that RWA may modulate memory for threatening ingroup deviants.

2.1.4 Hypotheses

The present research investigates reputational memory for uncooperative individuals in intergroup contexts, which has (to the best of our knowledge) not been previously examined. It goes beyond recognition and group assignment by differentiating between target recognition (old-new discrimination of faces), memory for a target's behavior (reputational memory), and reputation guessing in an ingroup-outgroup context. We assume that group-based processes modulate memory for uncooperative ingroup versus outgroup targets. The *main hypothesis* states that reputational memory is better for uncooperative ingroup members, but not for uncooperative outgroup members. Furthermore, reputational memory for uncooperative ingroup members may be stronger than reputational memory for all other ingroup and outgroup members. The *second hypothesis* assumes that a positive view of the ingroup manifests in guessing biases: participants may guess that ingroup members are cooperative more often than outgroup members. The *third hypothesis* supposes that differential concerns for ingroup and outgroup behavior motivate differential retrieval of particular group members (i.e., trustworthy, cheating; Bell & Buchner, 2010a). We expect that meaningful categorization modulates reputational memory for ingroup relative to outgroup members. Specifically, higher identified, but not lower identified participants, remember relevant ingroup members better than outgroup members. As authoritarians are concerned with threatening norm deviance, we expect that people high in RWA remember uncooperative ingroup members better than uncooperative outgroup members in contrast to people low in RWA.

In the main experiments, we presented a variety of facial photographs of ingroup and outgroup members, combined with behavioral descriptions. Both the photographs and the descriptions had been pre-tested with independent samples. After encoding, a surprise memory test took place. Participants indicated whether they recognized target faces and remembered associated target behavior.

Two studies tested these hypotheses using a minimal group paradigm. This paradigm rules out the influences of previous group interactions (e.g., previous losses or victories), the influence of group-specific stereotypes (as participants have no prior knowledge about experimentally created groups), and the long-standing attachments that people develop to natural groups (see Brewer, 1979; Kessler & Mummendey, 2002). Hence, any observed effect leads back to the simple fact that people belong to one group (ingroup), but not to the other (outgroup). Study 1 tested reputational memory for within-group fairness, where uncooperative behavior was indicated by unequally sharing resources with ingroup members. Half of the targets in each group were uncooperative. Study 2 replicated and extended Study 1. Targets were described as engaging in uncooperative, cooperative or neutral behavior in short sentences. Study 2 additionally examined participants' guessing of whether ingroup or outgroup targets were associated with uncooperativeness. In Study 3, university affiliation (own vs. different) indicated targets' group membership. Cheating, neutral, and trustworthy descriptions of student social behavior indicated the targets' behavior. One third of the group members in each group were uncooperative (or cheating) in Study 2 and 3.

2.2 Study 1: Memory for unfair ingroup dictators

In Study 1, unequal sharing of resources with ingroup members in a dictator game indicated uncooperativeness. All interactions were within groups, and interaction partners knew about their group membership. Prior research has shown that group members prefer to interact with ingroup members over outgroup members, even if explicit and positive outgroup stereotypes are given (Foddy et al., 2009). Fairness restricts striving for individual benefits, which makes it important for maintaining cooperation. Sharing unequally with an ingroup member thus represents a violation of pro-social ingroup expectations (Mendoza et al., 2014; Yamagishi et al., 1999). Target pictures and (un)equal sharing were presented to participants during the encoding phase. In the subsequent test phase, old targets were presented again, in addition to new distractor targets. Participants had to indicate whether they recognized the target picture, and recall whether the target's behavior was uncooperative.

2.2.1 Method

Participants. The sample consisted of 130 participants (79 female; 51 male). Most of them were students at the University of Jena, Germany. The mean age of the sample was 24.00 years (*SD*= 5.46). Recruitment took place through university mailing lists, advertisements on Facebook, and at the university campus. Two participants were excluded from the analysis; the first due to experiencing technical problems during the experiment, and the second for having already participated in a similar study. Participants were payed 5 Euros for their participation.

Materials. Photographs of faces were employed to grant each target a distinct identity. The images were derived from available databases (Ebner, Riediger, & Lindenberger, 2010; Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & van Knippenberg, 2010; Lundqvist, Flykt, & Öhman, 1998; Minear & Park, 2004; PICS, 2008). Each picture consisted of a frontal facial shot with a neutral expression, presented in color with dimensions of 300x400 pixels. To prevent own-age or race-related biases from influencing target recognition (Hugenberg et al., 2010; Wiese, Komes, & Schweinberger, 2013), all of the photographs were of young Caucasian faces. Similarly, we attempted to prevent potential own-sex recognition biases (Herlitz & Lovén, 2013), by designing the experimental procedure so that female participants saw only female targets, and males only male targets.

In order to rule out potential memory effects due to the targets' facial appearance, we conducted a pre-test. 49 participants (29 female) who did not participate in the main study evaluated 295 photographs for likability, trustworthiness and distinctiveness on a six-point scale (1= *not at all*; 6= *very much*). On the basis of these ratings, we selected 80 photographs of each sex to be utilized in the main study. Selections were made based on how close the ratings were to the scale midpoints on all three dimensions. Overall, the mean likability of target faces was 3.22 (*SD*= .55), the mean trustworthiness was 3.35 (*SD*= .49), and the mean distinctiveness was 3.28 (*SD*= .30).

A second pre-test with 100 student (64 female; 36 male) participants provided the basis for the behavior manipulation. Participants rated the 50-50 distribution of monetary units as being fair (1= *unfair*; 7= *fair*; *M*= 6.50, *SD*= 1.11), whereas taking 90 out of 100 units was classified as unfair (*M*= 1.3, *SD*= .75). The fairness ratings of the distributions clearly differed from each other, *t*(99)= 34.30; *p*< .01; *d*= 3.48. Hence, distributions between 45 to 55 out of 100 units presented the manipulation of "cooperative behavior", and 5 to 15 units the manipulation of "uncooperative behavior" in the main study.

Procedure. The study began with participants providing informed consent for their participation. The experimenter asked if she could take a photograph of them with a neutral facial expression, allegedly for use in upcoming studies. However, the real purpose of this request was to make the target photographs employed in the study appear more realistic. After taking the photo, the experimenter accompanied participants to one of ten cubicles, for individual computer-based testing. Participants were asked to follow the instructions presented to them on the computer's screen, commencing the first proper phase of the experiment. Participants read that scientific studies had found a relationship between personality and social behavior, and that the goal of the current experiment was to investigate this topic. They were then asked to complete a perceptual task, and learned that they were either 'figure' or 'ground' perceivers (e.g., Otten & Wentura, 1999). The perceptual types (i.e., minimal group memberships) were randomly assigned, together with the colors yellow and blue to represent ingroup and outgroup. The salience of group membership was enhanced using an open question about the new group, and t-shirts in group colors that were given to the participants. A short addendum stated that there was a close connection between perceptual type and social behavior, suggesting that there were fundamental psychological differences between the categories (e.g., Forgas & Fiedler, 1996). We did not specify the sizes of either group.

Moving to the second stage of the experiment, participants read the rules of the dictator game they would soon be asked to play. One of the two players involved (i.e., the target) would decide how much of 100 monetary units to keep, and how much to transfer to an ingroup member. The instructions lead the participants to believe that they were viewing decisions made by other participants, and that some of them were currently working in the other cubicles. We also claimed that participants would execute a team task with other ingroup members after the experiment. In order to preserve their anonymity, every player would be assigned a random photograph as a personal avatar. In four preliminary rounds, participants decided how much of the 100 monetary units they would donate to another ingroup member. In the subsequent encoding phase, participants sequentially observed 40 target decisions (i.e., resource distributions in the dictator game). Photographs of faces (with a short introduction; e.g., "This is T.") and indications of group membership (background color) were presented for 2 seconds before the target's distributional decision appeared underneath. After another 4.5 seconds, the photograph and the decision disappeared, and participants were asked to rate the target's fairness (e.g., "How fair do you think T. is?") and state whether the target was associated with the ingroup or the outgroup. The next trial started after the questions were answered, continuing until all 40 decisions were observed. The order of faces was randomized between subjects, as was their behavior and group membership. Of the 40 targets viewed by each participant, 20 targets were ingroup members, of whom half donated 45 to 55 units (cooperation) and the other half 5 to 15 units (uncooperativeness). The other 20 targets were outgroup members, of whom half distributed fairly and half unfairly.

Once the encoding phase was completed, participants were allowed a short break before the final, testing phase began. In a surprise memory test, participants were shown the 40 faces they had previously seen, plus 40 new faces. Of the new faces, half were ingroup members and half outgroup members. Participants were asked to state whether they had seen each target before (old-new discrimination). If they indicated that they recognized a target, participants were asked to recall whether the target's decision during the dictator game was fair or unfair (reputational memory).

Finally, participants answered items assessing ingroup identification and group impression and completed a manipulation check. Ingroup identification was accessed through four basic items on a seven-point scale (1= *not at all*; 7= *very much*): "I feel like a figure/ground perceiver", "I am a figure/ground perceiver", "I see myself as a figure/ground perceiver", "I identify with figure/ground perceivers". Participants were thanked, debriefed and given their incentives.[7]

Design. The study consisted of a 2 (*target*: ingroup vs outgroup) x 2 (*behavior*: cooperative vs uncooperative) design with two within-subject factors. More precisely, each

[7] Additionally, a set of personality tests (right wing authoritarianism, victim justice sensitivity, social value orientation) was completed by all participants in order to control for individual differences. These variables did not influence the results, and are subsequently not included in this dissertation.

participant was confronted with the four target types during the dictator game. Four dependent variables were analyzed via ANOVA: the perception of fairness of the targets' behavior during the encoding phase (fairness), correct discrimination of previously seen 'old' faces as old in the test phase (hits), false discrimination of new faces as old (false alarms), correct discrimination of old faces as old including false alarms (recognition sensitivity; P_r), and correct classification of target behavior as either cooperative or uncooperative (correct behavior classification). We applied Greenhouse-Geisser corrections in case of violations of sphericity, which could lead to fractional degrees of freedom. Furthermore, multinomial modeling of source monitoring provided estimates of reputational memory (elsewhere source memory, see Bayen et al., 1996; Buchner et al., 2009). The sample size was estimated from prior studies on reputational memory for uncooperative individuals (e.g., Bell, Buchner, Kroneisen, et al., 2012; Bell et al., 2010). The sample of 128 participants, conducting 80 trials each (N = 10,400), provided the possibility of detecting small differences between two parameters (df = 1; $\omega \approx 0.04$) with α= .05 and 1-β= .95 (calculated in G*Power; Faul, Erdfelder, Lang, & Buchner, 2007). We used *Multitree* for multinomial tree modeling (Moshagen, 2010).

2.2.2 Results

Manipulation Checks. On average, participants kept 54.93 (*SD*= 13.21) out of 100 monetary units (MU) to themselves when sharing with an ingroup member. Thus, participants tended to distribute resources fairly. Only 4% of participants distributed less than 15 MU, thus being uncooperative. All of the participants correctly remembered their ingroup color and their perceptual type at the end of the experiment.

Identification and Ingroup Bias. Participants indicated that they identified with their ingroup (α= .85, *M*= 4.26, *SD*= 1.48). They thus perceived themselves as group members. Participants' overall impression of the ingroup (1= *negative*; 100= *positive; M*= 54.45, *SD*= 14.79) was significantly more positive than their impression of the outgroup (*M*= 49.34, *SD*=12.44), *t*(127)= 2.79, *p*< .01, *d*=.25. Participants also indicated the proportion of uncooperative targets they saw in the ingroup (1= *20% uncooperative*; 7= *80% uncooperative*; *M*= 50.9%, *SD*= 16.3), and in the outgroup (*M*= 52.4%, *SD*= 14.0). The answers resembled the actual frequency of uncooperative ingroup members during encoding (50% of the group members in each group). There were no significant differences between the perceived numbers of uncooperative targets in the ingroup and outgroup, $t(127)= 1.08, p=.14, d=.09$.

Fairness. Targets' group memberships did not influence fairness ratings, $F(1,127)= .15$, $p= .69$, $\eta^2 < .01$, though target behavior showed a main effect, $F(1,127)= 1016.47$, $p< .01$, $\eta^2= .81$. There was no interaction effect on fairness ratings, $F(1,127)= .45$, $p= .50$, $\eta^2< .01$. Distributions of 5 to 15 MU (uncooperative behavior) with an ingroup member were evaluated as less fair than distributions of 45 to 55 MU (cooperative behavior), independent of group membership (see Appendix A for descriptive statistics).

Old-New Discrimination. 1 illustrates that the mean number of hits per target type did not differ in terms of group membership, $F(1,127)< .01$, nor target behavior, $F(1,127)= .08, p= .78, \eta^2< .01$. There was no interaction regarding the number of hits, $F(1,127)= .13, p= .71, \eta^2< .01$. The mean number of false alarms did not differ between ingroup faces (M=3.23; SD= 3.14) and outgroup faces (M= 3.23; SD= 2.88), $t(127)\leq .01$. We determined how well participants discriminated between old and new target faces in terms of the P_r. The P_r is the sensitivity measure of the two-high threshold (2HT) model. It subtracts false alarm rates from hit rates. The multinomial model of source monitoring that is used in the present paper is based on the 2HT model (see Bayen et al., 1996). Therefore, the P_r is a good approximation of the recognition parameter D (Bell & Buchner, 2009; Snodgrass & Corwin, 1988). Means and standard deviations of the P_r are provided in Table 1. As expected from prior research, old-new discrimination did not differ as a function of target type (e.g., Barclay & Lalumière, 2006; Buchner et al., 2009). The sensitivity for discriminating old and new target faces P_r did not differ significantly in terms of group membership, $F(1,127) < .01, p= 1.00, \eta^2< .01$, nor target behavior, $F(1,127)= .08, p= .78, \eta^2< .01$. There was no significant interaction effect between the two factors on the P_r, $F(1,127)= .13, p= .72, \eta^2< .01$.

Table 1. Mean number of hits, recognition sensitivity indicated by the P_r, and correct behavior classifications per target type in the test phase of Study 1 (*N*= 128), Section 2

	Hits		P_r		Correct behavior classifications	
	Ingroup	Outgroup	Ingroup	Outgroup	Ingroup	Outgroup
	M (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*
Uncooperative	5.55 (1.98)	5.51 (2.10)	.39 (.21)	.39 (.19)	3.10 (1.85)	2.82 (1.87)
Cooperative	5.55 (2.12)	5.59 (2.13)	.39 (.20)	.40 (.22)	2.78 (1.70)	2.74 (1.72)

Note: M=*mean,* SD= *standard deviation, hits = number of hits in recognition of old faces,* $\mathrm{P_r}$ *= hit rate – false alarm rate, correct behavior classification = number of correct classifications of recognized faces associated with uncooperative or cooperative behavior*

Correct Behavior Classifications. Our main hypothesis states that reputational memory (i.e., memory for a particular person's behavior) is enhanced for uncooperative ingroup members. We thus first sought to examine the number of correct classifications of target behavior. Participants showed a tendency to correctly classify uncooperative ingroup members as uncooperative more often than other targets (see Table 1). However, there was no main effect of group membership, $F(1,127)= 1.90$, $p= .17$, $\eta^2< .01$, no main effect of target behavior, $F(1,127)= .99$, $p= .32$, $\eta^2< .01$, and no interaction effect on the number of correct behavior classifications, $F(1,127)= 1.88$, $p= .32$, $\eta^2< .01$.

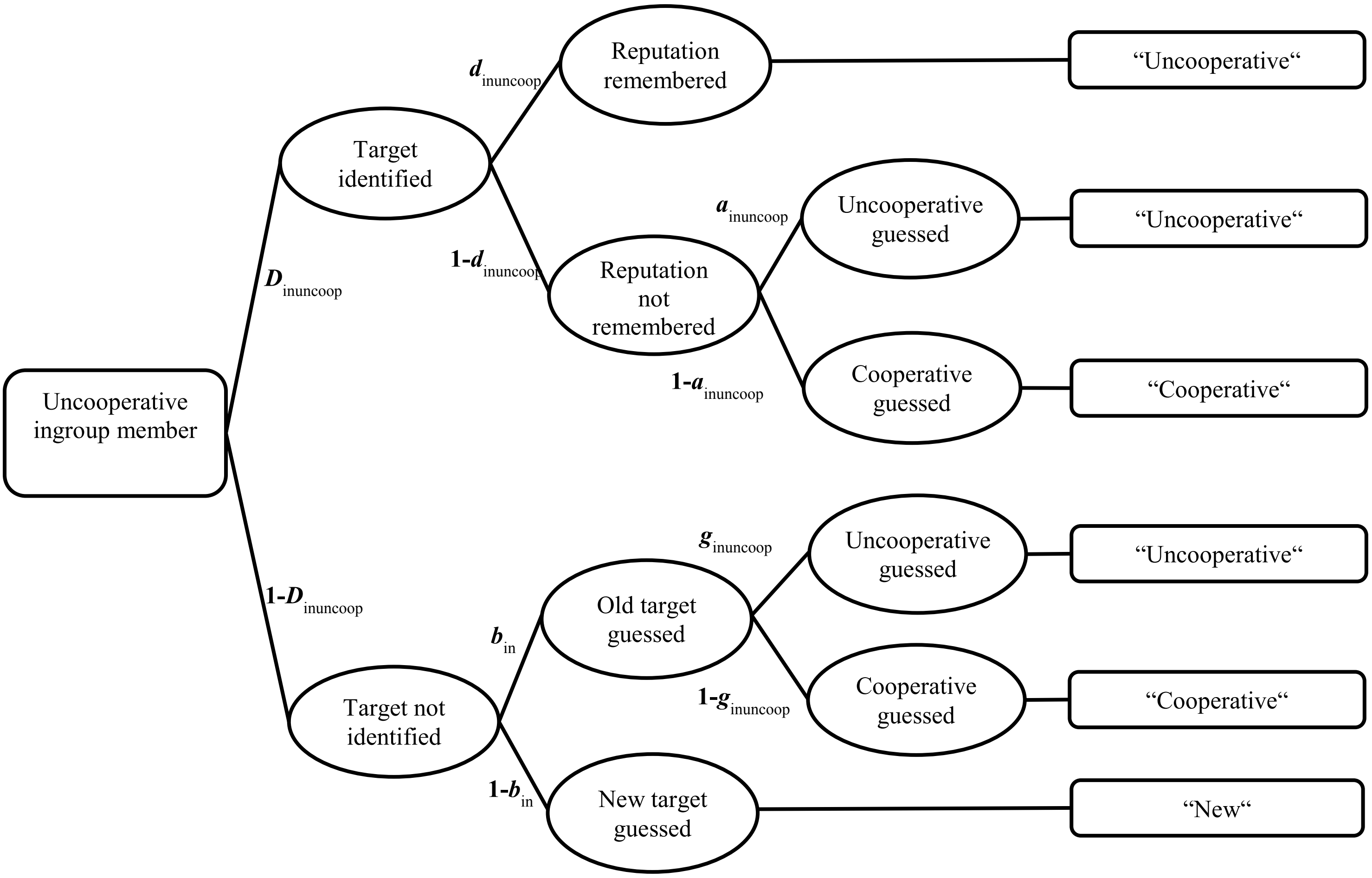
Multinomial Modeling of Source Monitoring. True reputational memory differs from mere correct behavior classification, as correct behavior classification may emerge from pure guesswork. Moreover, increasing recognition of a target type enhances the likelihood of guessing the target's reputation, as behavior was only asked for targets categorized as "old". We applied multinomial models of source monitoring to account for different cognitive processes that contribute to the participants' responses in the present memory task (e.g., Batchelder & Riefer, 1990; 1999; Bröder & Meiser, 2007). The 2HT multinomial model of source monitoring (Bayen et al., 1996) provides an appropriate approach for differentiating actual memory from the mere guessing of person information (Buchner et al., 2009). The model determines the conditional probabilities of old-new discrimination (face recognition), reputational memory (memory for the person's behavior), and guessing (old-new guessing; reputation guessing). The estimated parameters represent the probabilities for each of these processes ranging from 0 (process did not occur) to 1 (process always occurred). Parameter estimates are based on observed data frequency and obtained through an expectancy maximization algorithm (for further reading, see Batchelder & Riefer, 1999; Moshagen, 2010).

The present model consists of six trees that specify parameters for each target type: uncooperative, cooperative, and new ingroup members; and uncooperative, cooperative and new outgroup members. Figure 1 illustrates a tree that disentangles the responses towards "uncooperative ingroup members". The target type is specified by the left rectangles. The branches lead to one of three possible responses (right rectangles): new, old and associated with uncooperative behavior, or old and associated with cooperative behavior. The indices of the parameters specify the processes for different target types.

This section illustrates the model parameters, following the processes in Figure 1 from left to right. *D* is the probability that "old" faces were correctly identified as "old" and "new" faces were correctly identified as "new" (old-new discrimination). 1-*D* is the probability that participants did not detect whether the face was "old" or "new". If participants recognized an old target face, they remembered the target's previous behavior with the probability *d* (reputational memory). If they did not remember the target behavior (1-*d*), they guessed that the target was associated with uncooperative (and not cooperative) behavior (*a* = reputation guessing of identified faces). Estimates of *a* higher than .5 indicate a guessing bias in favor of uncooperativeness, whereas estimates of *a* lower than .5 indicate a guessing bias in favor of cooperation. An undetected target face (1-*D*) can still be classified as old through guessing (*b*). Half of the new target faces in the test phase were ingroup members, and half outgroup members. Therefore, the model distinguishes between guessing that an ingroup face was old and guessing that an outgroup face was old. If participants guessed that an undetected face was "old", they guessed with the probability *g* (reputation guessing for unidentified faces) that it was associated with uncooperative behavior (estimates higher than .5), and not cooperative behavior (estimates lower than .5). The probability of each response is determined by adding the probabilities from the branches that lead to the according response. For example, the following equation determines the probability that an uncooperative ingroup member is correctly classified as uncooperative: P("inuncoop"| inuncoop)= $D_{inuncoop}*d_{inuncoop} + D_{inuncoop}*1\text{-}\ d_{inuncoop}*a_{inuncoop} + (1\text{-}D_{inuncoop})*b_{in}*g_{inuncoop}$. Whereas the following equation determines the probability that a new ingroup member is classified as uncooperative: P("innew" |inuncoop)= $(1\text{-}D_{innew})*b_{in}*g_{inuncoop}$.

Multinomial modeling provides a goodness-of-fit statistic (G^2) that is approximately χ^2- distributed. The Akaike's information criteria (*AIC*) gives further information regarding the model fit. A negative *ΔAIC* indicates a good model fit (Wagenmakers & Farrell, 2004). After imposing additional restrictions, changes in the model fit (*ΔG*^{2}) indicate whether the restrictions are appropriate or not. *ωAIC* < .5 indicates that additional restrictions do not fit the model.

Figure 1. Exemplary illustration of multinomial tree model structure for "uncooperative ingroup members"

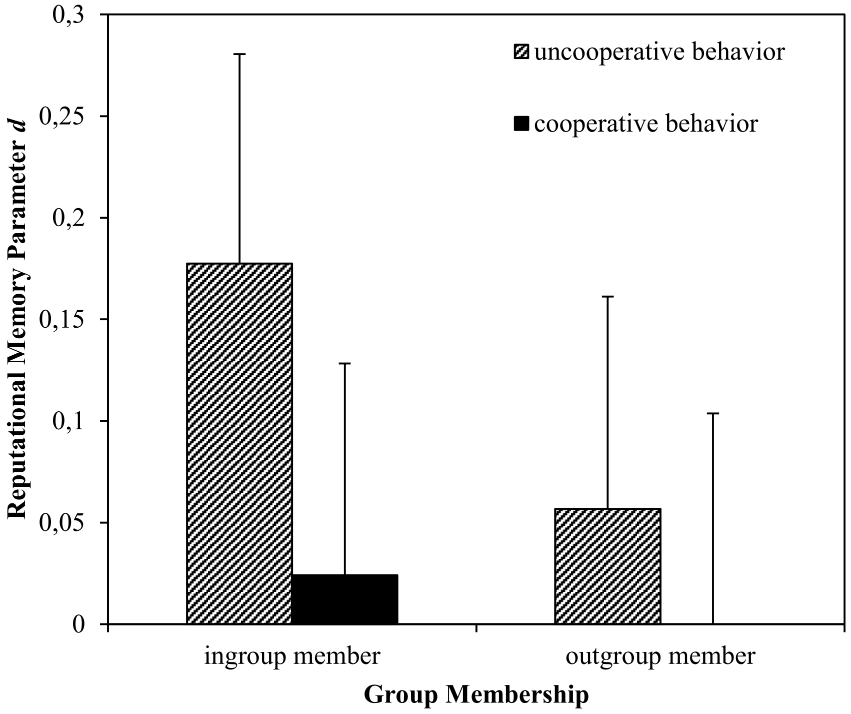


Note: The originally presented target is displayed on the left, and the behavior assigned by participants during the test phase on the right. The index "inuncoop" signals the target type "uncooperative ingroup member". The estimated parameters represent the following processes, involved in the memory task: D= *old-new discrimination of target face,* d= *reputational memory,* a= *guessing that a correctly recognized target was associated with uncooperative behavior and not with cooperative behavior (higher than .5≙ uncooperative, lower than .5 ≙ cooperative behavior),* b= *guessing that an unidentified target face has been presented before,* g= *guessing that an un identified target was associated with uncooperative behavior (higher than .5 ≙ uncooperative, lower than .5 ≙ cooperative behavior).*

Base Model Restrictions. Parameter restrictions have to be defined in order to obtain an identifiable base model (see Bayen et al., 1996). We derived base model assumptions from theoretical considerations and the results reported above. First, we restricted all *D*-parameters to being equal. The identification of an old item as old is as likely as identifying a new item as new (Glanzer & Adams, 1985, 1990; Kellen, Klauer, & Bröder, 2013). Previous research has shown that old-new discrimination is independent of both target valence and incongruity between the target schema and behavior (e.g., Bell & Buchner, 2010b; Buchner et al., 2009). Accordingly, we found that the P_r did not differ among target types. Second, we restricted both *b*-parameters to be equal ($b_{in} = b_{out}$), since target faces were randomly assigned to the ingroup and the outgroup. Third, guessing parameters on how a target behaved previously (uncooperatively, and thus not cooperatively) had to be constrained to obtain an identifiable base model and test hypotheses on reputational memory parameters. We restricted all reputation guessing parameters to a constant of .5 ($a_{inuncoop}$= $a_{outuncoop}$= $g_{inuncoop}$= $g_{outuncoop}$= .5), because participants reported that uncooperative behavior was equally distributed across ingroup and outgroup members. Guessing has been found to be in line with perceived contingency (Bayen & Kuhlmann, 2011; Klauer & Meiser, 2000). The model restrictions fit the data well, $G^2(6)$= 1.98, p= .92, ΔAIC= -10.02. Thus, the model was accepted as a base model. In line with the values of the P_r (see Table 1), the probability of correctly identifying a face as old or new was D= .39 (95% CI= [.37, .41]).

Reputational Memory. Our main hypothesis is that reputational memory is enhanced for uncooperative ingroup members, but not for uncooperative outgroup members. The parameter *d* represents reputational memory, that is, people recognize the target and remember their prior behavior. The restriction that the reputations of ingroup members were equally well remembered ($d_{iunncoop}$= d_{incoop}) led to a significant decrease in model fit, $\Delta G^2(1)$= 4.20, p=.04, ωAIC= .25. Figure 2 shows that participants remembered uncooperative ingroup members better than cooperative ingroup members. In contrast, there was no difference between the reputational memory for uncooperative and cooperative outgroup members, $\Delta G^2(1)$= .75, p=.39, ωAIC= .65. Reputational memory for uncooperative ingroup members was only numerically better than reputational memory for uncooperative outgroup members (see Figure 2), $\Delta G^2(1)$= 2.60, p=.11, ωAIC= .43.

Figure 2. Reputational memory parameters *d* for all target types, as estimated by the base model in Study 1, Section 2



Note: error bars represent the 95% confidence intervals

A nested analysis of hierarchical models tested whether reputational memory would be enhanced for uncooperative ingroup members compared to all other targets. The cooperative behaviors of ingroup and outgroup members were equally poorly remembered, $\Delta G^2(1)$= .17, *p*= .68, *ωAIC*= .71. A new base model was constructed that contained the restriction $d_{incoop=}$ $d_{outcoop}$. It fit the data well, $G^2(7)$= 2.15, *p*= 0.95, *ΔAIC*= -11.85. Reputational memory for uncooperative outgroup members did not differ from that for cooperative ingroup or outgroup members, $\Delta G^2(1)$= .58, *p*= .45, *ωAIC*= .67. A new base model was generated that included the restriction $d_{outuncoop=}$ $d_{incoop=}$ $d_{outcoop}$, which provided a good model fit, $G^2(8)$= 2.74, *p*= 0.95, *ΔAI* = -13.26. Finally, restricting reputational memory for uncooperative ingroup members to being equal to reputational memory for all other targets significantly decreased the model fit, $\Delta G^2(1)$= 6.37, *p*= .01, *ωAIC*= .10.

Confirming our main hypothesis, participants remembered uncooperative ingroup members better than cooperative ingroup members, and they remembered all outgroup members (be they uncooperative or cooperative) equally poorly. Behavior classifications tended to be correct more often for uncooperative ingroup targets than for all other targets.

The application of multinomial models revealed a significant advantage in reputational memory for uncooperative ingroup members. Participants remembered the uncooperative behavior of ingroup members better than the behavior of any other target type. The perceived fairness of targets did not differ between ingroup and outgroup. There were no differences in face recognition between target types considering the number of hits, P_rs and *D*-parameters. As predicted, the ingroup was perceived more positively than the outgroup. The perceived frequency of uncooperative behavior was equal for the ingroup and the outgroup. Nonetheless, participants remembered uncooperative ingroup members better than other targets.

Although these findings support our main hypothesis, it must be noted that Study 1 has its limitations. Though the hypothesized enhanced reputational memory for uncooperative ingroup members emerged, the memory parameters were generally quite low. This indicates that the task may have been rather difficult. Furthermore, uncooperative and cooperative behaviors were differentiated only by different patterns of resource distribution in the dictator game. This does not leave much room for discerning different targets. Moreover, previous studies have found that reputational memory for uncooperative individuals is enhanced when such behavior is uncommon (e.g., Buchner et al., 2009). A full half of the targets in Study 1 displayed uncooperativeness. We thus sought to remedy these limitations in Study 2.

2.3 Study 2: Memory for ingroup cheaters

In Study 2, we aimed to replicate the findings of Study 1 while addressing its limitations. In order to do this, we made four changes. First, we added a manipulation of neutral behavior to those for uncooperative and cooperative behavior. Adding neutrally behaving targets in equal proportion to uncooperative and cooperative targets meant that participants would see each behavior type one third of the time, reducing the frequency of uncooperative behavior. Second, the inclusion of neutral behavior permitted the estimation of potential guessing biases based on ingroup or outgroup membership. Third, we altered the descriptions of the targets we provided, adopting descriptions from prior studies (Bell, Buchner, Erdfelder, et al., 2012) where the uncooperative behavior was cheating. Cheating violates social contract rules, and thus disrupts cooperation (Cosmides & Tooby, 1992). These descriptions may also enhance reputational memory, as the behaviors are more naturalistic.

Finally, though we employed the same minimal group paradigm as Study 1, we eliminated any mention of a future group task. This left only group categorization as a modulating factor for reputational memory and guessing.

2.3.1 Method

Participants. The study sample consisted of 132 participants (63 female; 69 male). 96.21% were university students of various disciplines. Participants' ages ranged from 18 to 55 years (*M*= 23.08, *SD*= 4.27). The recruitment process and incentives provided were identical to Study 1. Seven participants were excluded from the final analysis. Four claimed to be familiar with the stimuli we used for target presentation, one had participated in a similar study before, and two told the experimenters that they did not believe our group manipulation.[8]

Material. The study's materials consisted of 72 grey-scaled facial photographs (256 bit, 116x164 pixels) and 36 behavioral descriptions, each containing about seven words. The photographs were of Caucasian males, ranging from young to middle-aged. All participants saw the same set of faces. The behavioral descriptions contained 12 uncooperative, 12 neutral, and 12 cooperative behaviors.9 Consequently, the frequency of uncooperative behavior was one third, and it did not constitute any group's norm. Material ratings provided by an independent sample are described in Bell, Buchner, Erdfelder, and colleagues (2012, p.460). The authors found that uncooperative descriptions were clearly rated more negatively (and the cooperative descriptions more positively) than the neutral behavioral descriptions. For instance, a description of an uncooperative target would read: "K.P. is a gas station attendant. He fleeces inattentive drivers of their change", whereas a description of a cooperative target would be: "M.D. is a carpenter. He reads books to lonely elderly people". A neutral description would read: "P. W. is a courier driver. He uses a big backpack for transportations“.

Procedure. The procedure was similar to Study 1, with some minor changes. First, we eliminated the request that participants have their picture taken. The experimenters individually seated participants in 1 of 10 cubicles. Participants proceeded to complete the 'perceptual types' task in order to induce a minimal group membership. Participants were

[8] We re-analyzed the data to include the whole sample, though there were no differences in the results. These analyses can be provided as a supplement if requested.

[9] We kindly thank Raoul Bell for providing the material for this study.

handed a colored scarf as a reminder of their ingroup membership after the group assignment. During the encoding phase, participants sequentially saw 36 target faces (including ingroup and outgroup members), each one appearing for two seconds before behavioral descriptions were displayed underneath. As in Study 1, the background color of the target faces indicated their group membership. After participants rated the likability of the target, the next one was presented. The test phase presented participants with the 36 old faces, plus 36 new ones, all of which were presented with their background color indicating group membership. Participants first rated the likability of targets. As in Study 1, participants stated whether a target was old or new, and classified them as having been associated with one of three categories: uncooperative, cooperative, or neutral behavior. At the end of the experiment, participants completed a manipulation check and answered items assessing group impressions, ingroup identification, and the RWA-scale.10 All participants correctly remembered their perceptual style and the color representing their ingroup. Subsequently participants were thanked, debriefed, and incentivized.

Design. The study design consisted of a 2 (target: ingroup vs outgroup) x 3 (behavior: uncooperative vs neutral vs cooperative) design with two within-subject factors. As in Study 1, we applied t-tests, ANOVAs and multinomial models of source monitoring. 125 participants, conducting 72 trials each (N = 9,000), provides the possibility of detecting small effects (df = 1; $\omega \approx 0.04$) with α= .05 and 1-β= .95.

2.3.2 Results

Identification and Ingroup Bias. An index of ingroup identification indicated that participants identified with their ingroup (*α*= .75, *M*= 4.33, *SD*= 1.20). Their overall impressions of the ingroup (0= *negative*; 100= *positive*; *M*= 52.85, *SD*= 15.07) and the outgroup (*M*= 52.10, *SD*= 14.46) were equally positive, *t*(105)= .29, *p*= .39, *d*= .03. Participants perceived significantly fewer uncooperative ingroup members (*M*=39.4%, *SD*= 13.0) than uncooperative outgroup members (1= *20% uncooperative*; 7= *80% uncooperative; M*= 44.0 %, *SD*= 14.7), *t*(105)= 2.81, *p*< .01, *d*= .28. Although ingroup bias did not emerge in the overall impression of the groups, participants perceived less uncooperativeness in the ingroup than the outgroup.

[10] Due to technical issues, we only measured the ingroup biases of 112 participants. Identification was assessed in the entire sample. The RWA-scale will not be presented here, but results could be featured in the article's supplementary material for interested researchers.

Encoding and Test Phase Likability. The kind of behavior a target performed – but not their group membership – contributed significantly to their likability. The analysis of encoding phase likability ratings showed a significant main effect of target behavior, $F(1.36,168.89)= 407.93$, $p< .001$, $\eta^2= .68$ (sphericity violations required Greenhouse-Geisser corrections). There was no main effect of group membership on encoding phase likability, $F(1,124)= .766$, $p= .38$, $\eta^2< .01$, nor any interaction effect, $F(2,248)= .46$, $p= .63$, $\eta^2< .01$. The ratings during the test phase revealed that uncooperative targets were liked significantly less than cooperative or neutral targets, $F(2,248)= 19.341$, $p< .001$, $\eta^2= .05$. There was no main effect of group membership on test phase likability, $F(1,124)= 1.03$, $p= .31$, $\eta^2< .01$, and no interaction effect, $F(2,248)= 1.05$, $p= .35$, $\eta^2< .01$. Descriptive statistics are displayed in Appendix A.

Old-New Discrimination. Participants correctly reported that a neutral target face was old less often than other target faces (see Table 2). There was a small main effect of the number of hits on target behavior, $F(2,248)= 4.46$, $p= .01$, $\eta^2= .01$. There was no main effect of group membership on the number of hits, $F(1,124)= 1.63$, $p= .20$, $\eta^2< .01$, and no interaction effect, $F(2,248)= 1.06$, $p= .35$, $\eta^2< .01$. The mean number of false alarms did not significantly differ between ingroup faces (M=1.58; SD= 2.07) and outgroup faces (M= 1.77; SD= 2.07), $t(124)= 1.19$, $p= .24$, $d= .09$. We again determined sensitivity for old-new discrimination of the 2HT model (P_r). Target behavior had a main effect on the P_r, $F(2,248)= 4.47$, $p= .01$, $\eta^2= .01$. Participants were less sensitive to recognizing neutral targets than other targets (see Table 2). There was no main effect of group membership on the P_r, $F(1,124)= .87$, $p= .35$, $\eta^2< .01$, and no interaction effect, $F(2,248)= .78$, $p= .45$, $\eta^2< .01$.

Table 2. Mean number of hits, recognition sensitivity indicated by the P_r, and correct behavior classifications per target type in the test phase of Study 2 (N = 125), Section 2

	Hits		P_r		Correct Behavior Classification	
	Ingroup	Outgroup	Ingroup	Outgroup	Ingroup	Outgroup
	M (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD*)	*M* (*SD)*
Uncooperative	4.35 (1.45)	4.20 (1.39)	.64 (.25)	.60 (.25)	2.18 (1.30)	1.93 (1.19)
Neutral	3.97 (1.54)	4.02 (1.51)	.57 (.67)	.57 (.26)	1.37 (1.25)	1.34 (1.11)
Cooperative	4.17 (1.58)	4.21 (1.30)	.61 (.27)	.60 (.24)	1.98 (1.19)	1.98 (1.22)

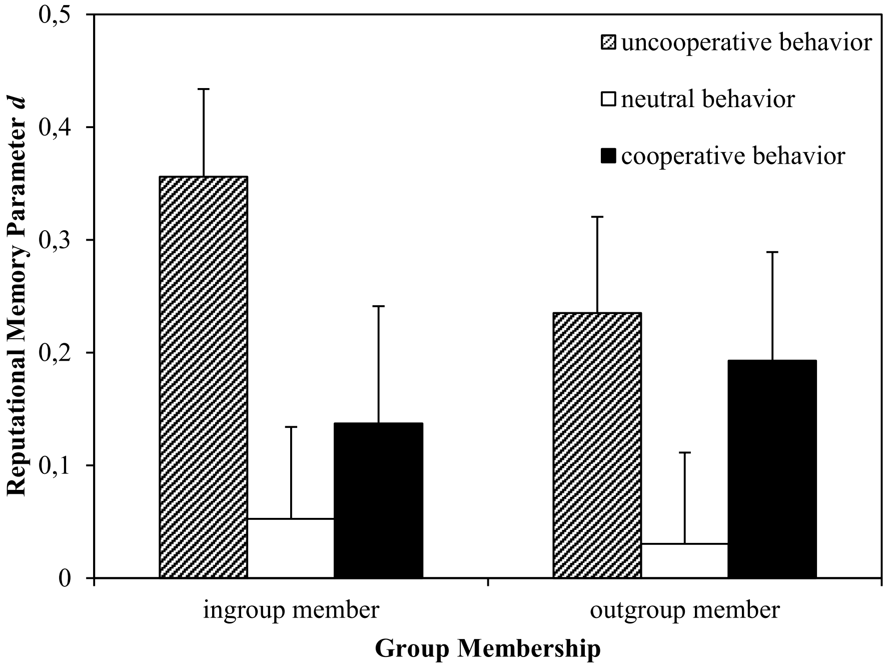
Note: *M*=mean, *SD*= standard deviation, hits = number of hits in recognition of old faces, P_r = hit rate – false alarm rate, correct behavior classification = number of correct classifications of recognized faces associated with uncooperative or cooperative behavior.

Correct Behavior Classification. Overall, participants classified fewer neutral targets correctly (as neutral) than they classified uncooperative targets as uncooperative, or cooperative targets as cooperative (see Table 2). This is demonstrated by a main effect of target behavior on correct behavior classification, $F(2,248)= 23.14$, $p< .01$, $\eta^2= .08$. There was no main effect of group membership on correct behavior classification, $F(1,124)= 1.62$, $p= .20$, $\eta^2< .01$, and no interaction effect, $F(2,248)= 1.06$, $p= .35$, $\eta^2< .01$. Numerically, uncooperative ingroup members were classified correctly more often than other targets.

Multinomial Tree Modeling of Source Monitoring: Base Model Restrictions. We applied multinomial modeling to unravel the probabilities of the underlying processes of participants' responses. The model contained two sets (ingroup and outgroup members) of four trees (old targets with an uncooperative, neutral or cooperative description, and new targets). Base model assumptions were similar to those of Study 1, except for two differences. First, all *D*s were restricted to being equal, but independent of old-new discrimination of neutral targets. We restricted $D_{inneutral}$ to be equal to $D_{outneutral}$. This was indicated by the P_r, which showed that recognition of neutral faces was poorer than for other faces (see Table 2). Second, Study 2 contained neutral ingroup and outgroup targets. The additional parameters $g_{neutral}$ and $a_{neutral}$ represent the probability that participants guessed that targets behaved neutrally compared to relevantly (i.e., cooperatively or uncooperatively; see Buchner et al., 2009). We restricted guessing parameter $a_{inneutral}$ to being equal to $a_{outneutral}$, and $g_{inneutral}$ to being equal to $g_{outneutral}$, because there was no reason to believe they would differ for ingroup and outgroup members. If participants guessed that a face was associated with relevant behavior (1- $a_{neutral}$ / 1- $g_{neutral}$), they subsequently guessed that the target was uncooperative with the probability a_{uncoop} or g_{uncoop}. As in Study 1, we restricted the probability of guessing that an unidentified target was old to being equal for ingroup and outgroup members (b_{in} = b_{out}). Other guessing parameters remained unrestricted for probability estimations. The goodness of fit showed that the base model fit the observed data, $G^2(9)$= 9.39, *p*= .40, *ΔAIC*= -8.61. As expected, participants recognized the faces of relevant targets (*D*= .61; 95 % CI = [.59, .63]) better than the faces of neutral targets ($D_{neutral}$ = .56; 95% CI= [.53, .59]), $\Delta G^2(1)$= 7.35, *p*< .01, *ωAIC*= .06. The parameter estimates again resemble the P_r values.

Reputational Memory. We tested our main hypothesis that reputational memory is enhanced for uncooperative ingroup members. As indicated by the correct behavior classification, *d*-parameters show that participants remembered uncooperative ingroup members better than neutral ingroup members, $\Delta G^2(1)$= 25.05, *p*< .01, *ωAIC*< .01. They also remembered uncooperative outgroup members better than neutral outgroup members, $\Delta G^2(1)$= 10.38, *p*< .01, *ωAIC*= .01. The restriction that reputational memory for uncooperative ingroup members and reputational memory for cooperative ingroup members are equal led to a significant change in model fit, $\Delta G^2(1)$= 9.38, *p*< .01, *ωAIC*= .02. Reputational memory for uncooperative outgroup members did not differ significantly from reputational memory for cooperative outgroup members, $\Delta G^2(1)$= .34, *p*=.56, *ωAIC*= .70. Moreover, participants remembered uncooperative ingroup members better than uncooperative outgroup members, $\Delta G^2(1)$= 4.29, *p*= .04, *ωAIC*= .24. Parameter estimates in Figure 3 show reputational memory as estimated by multinomial models of source monitoring is enhanced for uncooperative ingroup members. Correct behavior classification (not controlled for guessing) showed only a tendency to be higher for uncooperative ingroup members.

Figure 3. Reputational memory parameters *d* for all target types as estimated by the base model in Study 2, Section 2



Note: error bars represent the 95% confidence intervals

Hierarchically-nested models compared all of the reputational memory parameters associated with cooperative and uncooperative behavior. The procedure disregarded reputational memory for neutral targets, as it was generally low. The restriction d_{incoop}= $d_{outuncoop}$= $d_{outcoop}$ was compatible with the base model, *ΔG*^{2}(2)= 2.21, *p*= .33, *ωAIC*= .71. Hence, the new base model contained the two additional constraints and fit the data, *G*^{2}(11)= 11.60, *p*= .39, *ΔAIC*= -10.40. In a subsequent step, we compared $d_{inuncoop}$ with [d_{incoop}= $d_{outuncoop}$= $d_{outcoop}$] in the new base model. The assumption was clearly not compatible with the data, *ΔG*^{2}(1)= 8.77, *p*< .01, *ωAIC*= .03. We conclude that memory for uncooperative ingroup members was better than memory for all other targets.

Reputation Guessing. In the next step, we were interested in possible differences in participants' guessing behavior for ingroup and outgroup members. As reported, participants reported seeing fewer uncooperative ingroup members than uncooperative outgroup members. Our findings show that participants guessed that ingroup members were associated with uncooperative behavior less often ($g_{inuncoop}$= .35, 95% CI= [.27, .44]) than outgroup members ($g_{outuncoop}$= .54, 95% CI= [.45, .63]) when guessing the reputation of unidentified faces, *ΔG*^{2}(1)= 8.70, *p*< .01, *ωAIC*= .03. This was not the case when guessing the reputation of previously identified faces of ingroup members ($a_{inuncoop}$= .42, 95% CI= [.37, .48]) and outgroup members ($a_{outuncoop}$= .46, 95% CI= [.40, .51]), $\Delta G^2(1)$= .79, *p*= .37, *ωAIC*= .65. We presented an equal number of cooperative and uncooperative members in each group. Nonetheless, participants guessed that ingroup members were cooperative rather than uncooperative. This ingroup-favoring guessing emerged for unidentified ingroup members, $\Delta G^2(1)$= 10.45, *p*< .01, *ωAIC*= .01 ($g_{inuncoop}$= .5), and identified ingroup members, $\Delta G^2(1)$= 7.48, *p*< .01, *ωAIC*= .06 ($a_{inuncoop}$= .5). There were no favorable guessing biases for outgroup members, $G^2(1)$= .78, *p*= .38, *ωAIC*= .65 ($g_{outuncoop}$=.5); $G^2(1)$= 2.34, *p*= .13, *ωAIC*= .46 ($a_{outuncoop}$= .5). As predicted, participants guessed that ingroup members would be cooperative more often than outgroup members, and thus displayed a favorable ingroup bias.

2.4 Study 3: Memory for deviant ingroup members

Study 1 and 2 show that group categorization modulates reputational memory for cheaters. However, it is not completely clear how much group-related motivational processes or higher relevance of negative ingroup information account for the effects. Group identification may increase relevance of ingroup members and motivate in depths processing of ingroup members (e.g., Brewer et al., 1995). Study 3 focuses on the potential moderating role of higher versus lower relevance of group membership. Natural ingroups for which ingroup identification often varies substantially may reveal some of these processes. Moreover, in a context with existing groups, the groups' overall impression is less dependent on the presented targets. An ingroup focus might not lead to an enhanced reputational memory for cheaters, but an enhanced reputational memory for (relevant) ingroup members.

We suggest that high ingroup identification and high authoritarianism enhance memory retrieval for particular ingroup targets compared to outgroup targets. Study 3 measured ingroup identification and authoritarianism to determine individual differences. The procedure enables the comparison of reputational memory for ingroup and outgroup targets and moderating individual differences. Cheating, neutral, and trustworthy descriptions of student social behavior indicated the targets' behavior. University affiliation (own vs. different) specified targets' group membership. As in Study 1 and 2, we presented numerous ingroup and outgroup targets combined with behavioral descriptions (1/3rd cheater; 1/3rd trustworthy). The same amount of cheating, irrelevant and trustworthy targets was presented in either group. Multinomial models estimated true memory accuracy (old-new discrimination of target faces and reputational memory for target behavior) and guessing biases separately.

2.4.1 Method

Participants. 121 students of the University of Jena took part in the experimental sessions. We excluded seven participants from analysis, two because they accidentally received wrong instructions, one was familiar with the stimuli, and one failed to remember the ingroup's assigned color. The three other excluded participants reported not recognizing any of the faces presented in the test phase. Hence, a sample of 114 participants (63 female; M_{age}= 23.28, *SD*= 2.76) was retained for the analysis. Lab schedules were advertised on campus. Participants signed an informed consent and received five Euros as an incentive after completing the experiment.

Material and Pre-tests. Facial photographs with neutral expression depicted the targets. We presented only same-sex targets in the pre-test and the main study to avoid confounds with gender groups and own-gender biases in person memory (Herlitz & Lovén, 2013). All pictures selected for the study were rated in a pre-test by an independent student sample (*N*= 49, 29 female). The selected faces in the stimulus set appeared medium trustworthy, likable, and distinct (see method Study 1, p.27). Target behavior was described as either cheating (uncooperative), neutral, or trustworthy (cooperative) in a student context. Sentences appeared underneath the pictures. For instance, a cheater would be described as "M. plagiarizes thesis papers from articles, which are accessible by the public via the internet, if he/she doesn't feel like putting any effort into studying," whereas a trustworthy description would read: "R. completes a costly team task on his/her own, because a fellow student became ill at short notice and nevertheless hands it in as common work." A student sample evaluated the descriptions beforehand (*N*= 76, 50 female). We chose 36 sentences that were clearly tagged as cheating, neutral or trustworthy, and which appeared most realistic to participants. Participants perceived cheating and trustworthy descriptions as less realistic than descriptions of neutral behavior, *F*(2, 150)= 21.04, *p*< 001, η^2= .22. Further, the behavioral descriptions showed a main effect in perceived valence, *F*(2, 115.32)= 1955.76[11], *p*< 001, η^2= .96. It indicated that cheating was rated more negatively and trustworthy behavior was rated more positively than neutral behavior (see Appendix B for descriptive statistics).

[11] We corrected for violations of sphericity through Greenhouse-Geisser procedure

Procedure. The procedure was similar to the procedure in Study 1 and Study 2. After participants entered one of ten cubicles, they on a shirt in their "university" color. The university affiliations were represented by randomly assigned colors (yellow or blue) throughout the whole session. All instructions appeared on the computer screen. The experiment was structured into five parts. First, participants described the personal meaning of being a student at their university. Then they indicated the overall impression of in group and outgroup on a positive-negative slider-task. Second, during the encoding phase participants sequentially observed 36 individually presented target faces. The background color indicated the targets' group membership (own university vs. other university). The Universities selected (University of Jena, University of Erfurt) are located in nearby towns, but not competing with each other. After 2.5 seconds the behavioral descriptions were presented underneath the face for 4 seconds. All pictures, group memberships, and descriptions were assigned randomly, as was the order of target presentation. Participants saw six faces in each condition (cheating, neutral, and trustworthy ingroup and outgroup members). When participants had rated the likability of the target, the pictures disappeared. The third assignment was a surprise memory test that started directly after the encoding phase. In 72 trials we again presented the 36 old target faces, plus 36 new faces of ingroup and outgroup members. Participants indicated how much they liked the target, and specified whether or not each target had been presented during the encoding phase. If yes, they classified which behavior the face had been associated with (cheating, neutral, or trustworthy). Fourth, we assessed ingroup identification, overall impression of the groups, and variables to measure group perceptions after the reputational memory test. Finally, participants completed measures of RWA (Funke, 2005) and other scales for individual differences.[12] The last slide debriefed participants and thanked them for participation.

Design. The study consisted of a 2 (group membership: ingroup, outgroup) x 3 (behavioral descriptions: cheating, neutral, trustworthy) within-participants design. As in Study 1 and Study 2, we applied *t*-Tests, ANOVAs and multinomial modeling. Multinomial reputational memory models do not feature testing interactions with continuous moderators. Instead, subgroup analyses are a commonly accepted procedure for revealing moderating variables (e.g., Bell & Buchner, 2010a; Bell et al., 2014). We tested differences in target memory of high and low identified participants and of participants high and low in RWA.

[12] These were group authoritarianism (Stellmacher & Petzel, 2005), the Interpersonal Reaction Index (Paulus, 2009), and Regulatory Focus (Keller, 2008). Those scales served as controls and were not included into the further analyses.

Since we expected continuous effects, we compared extreme groups to clearly illustrate the interaction patterns.

A priori sample size computations with G*Power indicated that 113 participants were required to detect such small effects (*w*≈ 0.04) with 1-ß= .95 and α= .05. Our complete sample consisted of 114 participants (N_{trial}= 8,208).

2.4.2 Results

Intergroup Perceptions. We assessed evaluative intergroup biases twice by using a slider measurement on the general group impression (1= *negative*, 100 = *positive*). Group impressions before the encoding phase and after the test phase correlated highly ($r_{ingroup}$= .81; $r_{outgroup}$= .87), and were therefore collapsed. A one-sided *t*-test demonstrated that participants' impression of their own group was significantly more positive (*M*= 75.33; *SD*= 17.21) than their view of the outgroup (*M*= 63.14; *SD*= 18.69), *t*(1,113)= 6.83, *p*< .001, *d*= .64.[13]

Encoding and Test Phase Likability. Participants evaluated targets differentially based on their behavioral description, *F*(1.47,165.93)= 745.17, *p*< .001, η^2= .83. Despite of the positive ingroup bias in overall group impression, there was no significant main effect of group membership or interaction effect on likability ratings, other *Fs*< 2. During the test phase only the group membership of the target was presented to participants. Participants had to retrieve behavior from memory. Nevertheless, likability ratings for the faces showed the same main effect for behavioral description of the target, *F*(1.88,212.72)= 19.43, *p*< .001, η^2= .07, but no other effects, *Fs*< 2. In both phases participants rated cheaters more unfavorably than they rated neutral persons, and trustworthy targets were liked best. Group membership did not affect the likability of targets (see Appendix A).

Old-New Discrimination. Comparing the hits in each stimulus category showed a significant main effect for target behavior, *F*(2,226)= 6.42, *p*< .01, η^2=.05. Table 3 demonstrates that participants correctly categorized cheaters as old less often than other targets. There was no main effect of target group membership and no interaction effect, *Fs*< 1. Participants did not indicate more often to identify a new ingroup face falsely as old (*M*= 2.75, *SD*= 2.66) than an outgroup face (*M*= 2.66, *SD*= 2.46), *t*(113)= .49, *p*= .63, *d*= .04. Target behavior had a main effect on the P_r, *F*(2,226)= 6.42, *p*< .01, η^2=.02. Simple contrasts revealed that participants recognized faces of ingroup and outgroup cheaters less often than

[13] Further variables, such as estimated frequency of cheaters, perceived similarity of group members, perceived intergroup competition were measured at the end of the experiment. They are not further reported here.

faces of neutral targets, $F(2,113)= 11.80$, $p= .001$, $\eta^2=.10$, and less often than faces of trustworthy targets, $F(1,113)= 4.01$, $p= .048$, $\eta^2=.03$. Participants did not show any differences in sensitivity to old-new recognition (P_r) due to target group membership, *F*s< 1.

Table 3. Mean number of hits and recognition sensitivity indicated by the P_r per target type in the test phase of Study 3 (*N*= 114), Section 2

	P_r		Hits	
	Ingroup	Outgroup	Ingroup	Outgroup
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Cheater	.51 (.26)	.50 (.25)	3.96 (1.40)	3.88 (1.40)
Irrelevant	.55 (.25)	.57 (.23)	4.23 (1.25)	4.30 (1.27)
Trustworthy	.54 (.23)	.53 (.26)	4.16 (1.29)	4.05 (1.41)

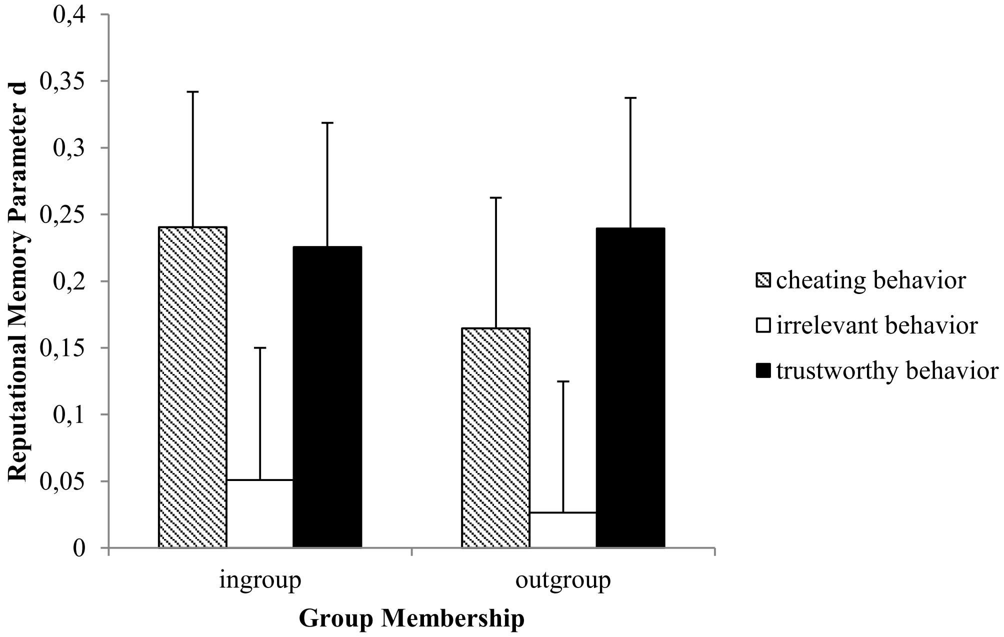
Note: M=*mean,* SD= *standard deviation, number of hits in recognition of old faces,* P_r = *sensitivity of old-new-discrimination (correct hit rate - false alarm rate)*

Multinomial Tree Modelling. We used these indications to obtain an identifiable base model in the multinomial analysis. Identifying a new face as new has been shown to be as probable as identifying an old face as old (Glanzer & Adams, 1990). Conversely, correct hits showed that participants were worse at correctly detecting cheaters as old in contrast to other target faces. Thus, we restricted all old-new-discrimination parameters *D* to being equal, except recognition of (ingroup and outgroup) cheaters. Twelve free parameters remained after applying restrictions to the multinomial tree models. The model assumptions fit the observed data well, $G^2(12)= 18.01$, $p= .12$, $\Delta AIC= -5.99$. We accepted the model as base model. Recognition of ingroup and outgroup cheaters was lower than correct identification of other targets as old or new ($D_{incheat} = D_{outcheat} = .48$, 95%CI = [.44- .52]; $D_{other} = .55$, 95%CI = [.53- .57]). The additional assumption that old-new discrimination does not differ between cheater and other faces led to a significant decrease in model fit, $\Delta G^2(1)= 8.31$, $p= .01$, $\omega AIC= .04$. This finding is in line with the literature. Two recent studies with high statistical power found that faces presented with negative descriptions were less well recognized than faces presented with other descriptions (Bell et al., 2015; Mieth, Bell, & Buchner, 2016).

Reputational Memory. To test differences in reputational memory, we applied comparisons of the reputational memory parameters *d* (see Figure 4 for parameter estimates with confidence intervals). Reputational memory for cheating targets differed from reputational memory for neutral targets, in both the ingroup and outgroup, $\Delta G^2_{in}(1)= 6.43$, $p=$

.01, *ωAIC*= .10; $\Delta G^2_{out}(1)$= 3.64, *p*= .05, *ωAIC*= .31. Reputational memory did not differ for ingroup cheaters and ingroup trustworthy targets, $\Delta G^2(1)$= .04, *p*= .84, *ωAIC*= .73. The same applied to reputational memory for cheating and trustworthy outgroup members, $\Delta G^2(1)$= .99, *p*= .32, *ωAIC*= .62. We created a new base model that contained only one parameter for reputational memory for trustworthy and cheating behavior (i.e. socially relevant behavior) in each group. The new base model fit the data well, $G^2(14)$= 19.03, *p*= .16 , *ΔAIC*= -8.97. Equality restrictions revealed that participants remembered relevant targets with the same accuracy, $\Delta G^2(1)$= .48, *p*= .49, *ωAIC*= .68. Hence, reputational memory was equally strong for trustworthy and cheating targets, but decreased for neutral targets. There were no differences in reputational memory based on group membership.

Figure 4. Reputational memory parameters *d* for all target types as estimated by base model 1 in Study 3, Section 2



Note: N=*114,* N_{trial} *= 8,208; error bars show upper 95% confidence intervals*

Guessing Bias. In case participants did not remember target reputations, they could still guess target behavior, e.g., that the target was a cheater instead of a trustworthy person (*g*-parameters). Participants guessed that an ingroup target was a cheater (g_{in}= .50, 95% CI= [.46, .54]) equally often than an outgroup target was a cheater (g_{out}= .46, 95% CI= [.42, .50]), $\Delta G^2(1)$= 1.74, *p*= .19, *ωAIC*= .53. Hence, participants guessing behavior did not indicate a group bias, despite of the more positive ingroup than outgroup impression.

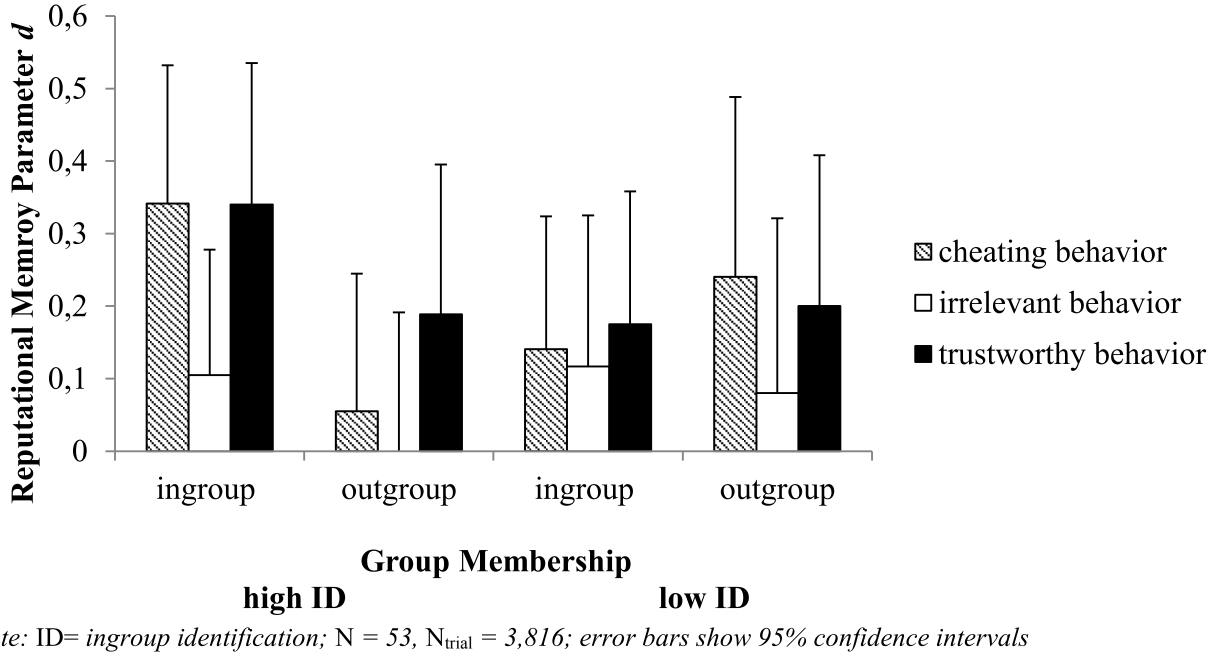
High and Low Ingroup Identification. Our results showed no differences in memory for ingroup and outgroup targets, whereas other studies found an influence of group membership on processing of person information (e.g., Brewer et al., 1995). Hence, it may be that the importance of group membership modulates reputational memory. Overall, participants identified well with their University affiliation (*α*= .89, *M*= 4.56, *SD*= .90). We defined two subsamples of low identifiers (25th percentile, *N*= 27) and high identifiers (75th percentile, *N*= 26). Mean identification in the low identified subsample was 3.31 (*SD*= .43). Mean identification of the highly identified subsample was 5.70 (*SD*= .35). Before encoding, the sample of high identifiers indicated more ingroup bias ([positivity of ingroup – pos. of outgroup]; *M*= 12.85, *SD*= 19.00) than low identifiers (*M*= 8.57, *SD*= 13.19). The difference was not significant, *t*(1,51)= .95, *p*= .35, *d*= -.26. The factor "subgroup identification" did not modulate the likability-ratings for ingroup and outgroup targets during the encoding phase, *F* < 1. As expected, the salience of the group context was higher for the highly identified subsample (Turner at al., 1987). They reported higher within-group similarity than between-group similarity in the ingroup (*M*= 1.13, *SD*= .47) and in the outgroup (*M*= 1.02, *SD*= .46). In contrast, the lowly identified subsample reported lower similarity within than between groups in the ingroup (*M*= .80, *SD*= .39) and in the outgroup (*M*= .92, *SD*= .37). This showed in a trend in high/low identification, *F*(1,51)= 9.50, *p*= .06, η^2= .07, and an interaction effect between group similarity and high/low identification, *F*(1,51)= 9.50, *p*< .01, η^2= .16.

We doubled the number of model trees in the original model and entered data from the two subsamples to test reputational memory. Restrictions were separately applied to each sample in order to define an identifiable base model. The base model was compatible with the data, $G^2(24, N_{trial} = 3{,}816)$= 13.39, *p*= .96, *ΔAIC*= -34.61. Reputational memory parameter estimates of the base model are displayed in Figure 5. First, we analysed the subsample of high identifiers. Results show that highly identified participants remembered cheating and trustworthy ingroup targets equally well, $\Delta G^2(1)$= .06, *p*= .80, *ωAIC*= .72. The same pattern emerged for cheating and trustworthy outgroup targets, $\Delta G^2(1)$= .60, *p*= .43, *ωAIC*= .67. These assumptions were integrated into a new base model, $G^2(26)$= 12.92, *p*= .94, *ΔAIC*= -31.18. Consequently, one parameter represented reputational memory of relevant ingroup behavior (i.e. cooperative, cheating) and another parameter reputational memory of relevant outgroup behavior. Equality restrictions showed that highly identified participants remembered relevant ingroup members better than corresponding outgroup members, $\Delta G^2(1)= 4.42, p= .04, \omega AIC= .25$.

We then applied the same analyses to the subsample of low identifiers. Low identifiers also remembered trustworthy and cheating ingroup members equally well, $\Delta G^2(1)= .03$, $p= .86$, $\omega AIC= .73$. The same applied for trustworthy and cheating outgroup members, $\Delta G^2(1)= .13$, $p= .72$, $\omega AIC= .72$. We defined a new base model in the low identified subsample. Then we restricted reputational memory for relevant (cheating and trustworthy) ingroup targets and simultaneously reputational memory for relevant outgroup targets to being equal. The new base model fit the data well, $G^2(26)= 13.18$, $p= .98$, $\Delta AIC= -38.82$. However, low identifiers' reputational memory for relevant targets was equally good for ingroup and outgroup members, $\Delta G^2(1)= .36$, $p= .55$, $\omega AIC= .69$.

A direct comparison of memory for ingroup and outgroup cheaters shows a marginal effect for highly identified. They remembered ingroup cheaters better than outgroup cheaters, $\Delta G^2(1)= 2.74$, $p= .098$, $\omega AIC= .41$, whereas low identifiers did not, $\Delta G^2(1)= .60$, $p= .44$, $\omega AIC= .67$. Guessing biases for ingroup and outgroup targets did not differ between the subsample of high and low identified.

Figure 5. Reputational memory parameters d for all target types as estimated in one common model containing data of subsamples of high and low identified participants, Study 3, Section 2



Note: ID= *ingroup identification;* N *= 53,* N_{trial} *= 3,816; error bars show 95% confidence intervals*

To sum up, only high identifiers remembered ingroup members' cheating and trustworthiness significantly better than corresponding outgroup members' behavior; low identifiers did not show this focused recollection.[14]

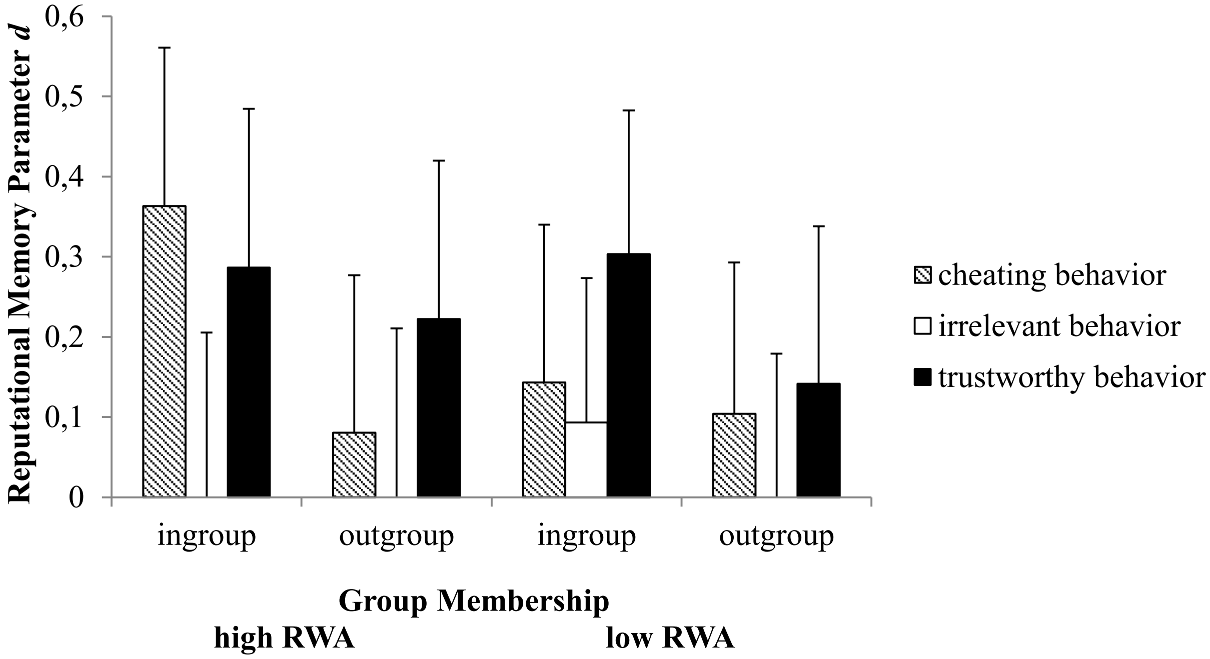
High and Low RWA. We measured RWA to see how the individual differences in concern about ingroup cheaters influence memory for person behavior. All six-point scale items of RWA were combined to one scale (α= .82, *M*= 2.59, *SD*= .75). RWA and ingroup identification are conceptually related (Duckitt, 1989; Kessler & Cohrs, 2008). Accordingly, they correlated positively, *r*= .32, *p*≤ .01. However, authoritarians display particularly negative attitudes (and behavior) towards threatening deviants (Altemeyer, 1996). Therefore, we expected people high in RWA to remember ingroup cheaters better than outgroup cheaters. The data of participants high in RWA (*N*= 31) and low in RWA (*N*= 30) was entered into one multinomial model with eight trees for each subsample. The base model restrictions fit the data, $G^2(24, N_{trial} = 4{,}392)= 24.70$, *p*= .42, ΔAIC= -23.30. Figure 6 shows reputational memory parameters for the both subsamples.

Participants high in RWA remembered ingroup cheaters significantly better than outgroup cheaters, $\Delta G^2(1)$= 4.17, *p*= .04, ωAIC= .25. The high RWA subsample remembered trustworthy targets and cheating targets in ingroup and outgroup equally well, $\Delta G^2(1)< 1$. In the low RWA subsample, participants' reputational memory was equally good for cheating ingroup and outgroup members, $\Delta G^2(1)$= .08, *p*=.77, ωAIC= .72. They remembered trustworthy and cheating targets equally well in the ingroup and in the outgroup, $\Delta G^2(1)< 1.3$. Thus, high RWA led to a better reputational memory for ingroup cheaters compared to outgroup cheaters, whereas this result did not obtain for low-RWA participants.[15]

[14] As predicted, the median split showed the same pattern of higher reputational memory of deviant ingroup members compared to deviant outgroup members only in the high identified subsample. However, the effect was weaker, $\Delta G^2(1)$= 3.13, *p*= .08, ωAIC= .36.

[15] A median split showed a tendency of higher reputational memory for ingroup cheaters compared to outgroup cheaters for people high in RWA, $\Delta G^2(1)$= 2.12, *p*= .14, ωAIC= .48.

Figure 6. Reputational memory parameters d for all target types as estimated in one common model containing data of subsamples of high and low in RWA, Study 3, Section 2



Note: RWA = *right-wing authoritarianism;* N = *61,* N_{trial} = *4,392; error bars show 95% confidence intervals*

2.4.3 Discussion

The present study examined whether a meaningful group context modulates reputational memory for ingroup and outgroup targets. As hypothesized, the results showed that ingroup identification and RWA lead to differences in reputational memory for cheating, trustworthy and neutral ingroup and outgroup members. The reported results showed that reputational memory was generally better for trustworthy and cheating targets compared to neutral targets. But memory did not differ in terms of targets' group categorization. However, we hypothesized that the importance of the ingroup would enhance reputational memory for socially relevant group members rather than social categorization. A social category is not necessarily important for the self. However, the relevance of the ingroup is expressed by ingroup identification (Brewer, 1979). Accordingly, participants who identified highly with their University ingroup remembered fellow ingroup students associated with relevant behavior (trustworthy, cheating) better than they remembered comparable outgroup students. The reputational memory effect was not related to general group impressions of high and low identified. This indicates that motivational factors of identification, i.e normative fit, are related to reputational memory advantages regarding ingroup members. This hypothesis was also supported by the influence of RWA on reputational memory. Authoritarians are especially concerned with anti-normative ingroup deviants (cheaters), which pose a threat to ingroup cooperation. In line with our assumptions, participants high in RWA remembered ingroup cheaters better than outgroup cheaters, whereas those low in RWA did not.

Ingroup Identification and Enhanced Memory for Ingroup Targets. The group context in the presented study became salient and meaningful through ingroup identification. For highly identified group members the differentiation in ingroup and outgroup is meaningful, whereas low identified group members may not care about group differences. Reputational memory differences for ingroup and outgroup members were elicited by ingroup identification with the natural ingroup. Since encoding likability ratings of the subgroups did not differ as a function of identification, this effect was not valence-driven.

Our results showed that group salience moderation, as suggested by Turner and colleagues (1987), differed in terms of identification. Highly identified participants perceived more within-group than between-group similarities, whereas this was inversed for lowly identified. Hence, ingroup identification indeed altered salience. However, participants did not remember all ingroup members better than outgroup members, but only the relevant. Ingroup identification elicits the expectation that ingroup members behave in line with norms. Own and other group members' behavior is evaluated in terms of normativity. Relevant behavior of ingroup members threatens the perceived reliability of ingroup norms (Abrams et al., 2000; Oakes et al., 1991). In accordance, with the notion that unusual behavior enhances reputational memory (e.g., Bell & Buchner, 2012), relevant ingroup behavior enhanced reputational memory compared to neutral ingroup or outgroup behavior for highly identified participants.

Despite the slightly increased ingroup bias for highly identified, reputational memory did not show inconsistency-effects with group impression in the subgroups. High identifiers have been shown to recall stereotype-inconsistent information about their ingroup better than low identifiers (Doosje et al., 2007). In the present study, participants retrieved ingroup cheaters (as opposed to a positive ingroup image) only slightly better than outgroup cheaters. Reputational memory for ingroup members was not influenced by the positive ingroup impression. Our results are in line with the notion that high identifiers are motivated to individualize fellow ingroup members, but not outgroup members. Research has shown that identified group members recognize faces of ingroup members better than faces of outgroup members (Van Bavel & Cunningham, 2012). Our findings go beyond face recognition and demonstrate that ingroup identification increases reputational memory for ingroup members.

Possibly, category-based information-processing of outgroup members accounts for the low memory accuracy of individual outgroup behavior (e.g., Brewer et al., 1995).

RWA and Enhanced Memory for Ingroup Cheaters. We additionally found that RWA enhances memory for ingroup cheaters. This promotes the idea that the perceived relevance of targets influences reputational memory. Additionally, RWA is a personality factor that describes the tendency to adhere to norms, follow leaders and punish deviants. Individuals evincing RWA highlight the importance of deviance within their society (Altemeyer, 1996). Authoritarianism has also been observed among other groups than national or societal level (Stellmacher & Petzel, 2005). Ingroup identification and RWA have been suggested to be related phenomena (Duckitt, 1989; Kessler & Cohrs, 2008). In our study cheating descriptions captured general misbehavior within the superordinate group of students. Instead of a general concern for cheaters, high authoritarians were particularly sensitive to ingroup cheaters. Whereas ingroup cheaters actively threaten cooperative tendencies among ingroup members, outgroup cheaters usually pose no threat to the ingroup. Hence, the high sensitivity to negative deviance within ingroups is in accordance with the suggestion that authoritarians foster ingroup cooperation (Kessler & Cohrs, 2008; Van de Wetering, 1996).

2.5 General Discussion

The presented studies demonstrated that reputational memory for social behavior is modulated by group membership of the targets (i.e. ingroup, outgroup) and the target behavior (i.e. cheating, trustworthy, or irrelevant), while ingroup biases remained. The results support the hypothesis that group members remember uncooperative ingroup members (i.e., those who violate norms for personal benefit) better than uncooperative outgroup members. Uncooperative ingroup members violate positive ingroup expectations (e.g., cooperativeness) and are more relevant than other ingroup or outgroup members. In Study 1 and 2, we found that memory for uncooperative targets exceeds memory for cooperative (and neutral) targets, but only for ingroup members. Furthermore, uncooperative ingroup members were remembered better than all other ingroup and outgroup targets. Study 3 demonstrated that not social categorization, but a meaningful group context altered reputational memory for relevant ingroup members. Moreover, ingroup biases were preserved and showed in overall group impression (Study 1 and 3) and in participants' guessing behavior (Study 2).

We tested our hypotheses by creating experimental groups with a minimal group paradigm (Study 1 and 2; e.g., Tajfel et al., 1971). Study 1 demonstrated enhanced reputational memory for uncooperative ingroup members when fairness was violated. Enhanced reputational memory for uncooperative ingroup members was also observed when uncooperative behavior was relatively frequent (i.e., occurring half the time) and manipulated in a highly abstract way (i.e., fair vs. unfair resource distributions). Study 2 replicated the effects of Study 1. Here, behavioral descriptions indicated uncooperative, neutral, and cooperative behavior in social exchange situations. One third of the group members were uncooperative. Despite better reputational memory for uncooperative ingroup members, participants showed ingroup bias. They indicated a more positive impression of the ingroup than the outgroup (Study 1), and more frequently guessed that an outgroup member was uncooperative compared to an ingroup member (Study 2).

Study 3 extended the studies by introducing a natural group context. In natural groups (own vs. other university) with no meaningful intergroup context participants did not generally remember ingroup cheaters over outgroup members. Uncooperative, cooperative, and neutral behaviors were described as student-relevant cheating, trustworthy, and neutral behavior. Cheating and trustworthy targets were remembered better than irrelevant behavior independent of group membership. As hypothesized, Study 3 shows that high identifiers remembered ingroup cheaters and trustworthy ingroup members significantly better than according outgroup members. This was not found for low identifiers for whom the group possesses are less relevant (Tajfel & Turner, 1979). Moreover, high authoritarians demonstrated a better memory for ingroup cheaters than outgroup cheaters. Low authoritarians who are less prone towards ingroup threat (Altemeyer, 1996) did not show such memory advantage. This indicates that motivational differentiation influences reputational memory for ingroup and outgroup members. As in Study 1 and 2, the positive ingroup impression was not altered throughout the experiment in Study 3. An ingroup bias showed before and after the experiment. The results suggest that normativity of ingroup behavior was more salient to highly identified than the overall group positivity.

The findings extend previous studies on memory for uncooperative individuals in social exchanges (e.g., Bell, Buchner, Erdfelder, et al., 2012; Buchner et al., 2009), as well as memory for group-related information (e.g., Brewer et al., 1995; Howard & Rothbart, 1980).

We demonstrated that memory for social behavior is modulated by the target's group membership (i.e., ingroup, outgroup) and behavior (i.e., uncooperative/deviant, or neutral), whilst retaining positive ingroup biases. In line with previous research, memory for faces was not affected by the target's uncooperative behavior (e.g., Buchner et al., 2009). Our results are in line with prior findings on enhanced memory for descriptions of negative ingroup behavior (Schaller & Maass, 1989; Gramzow et al., 2001). They add to the literature that not only the recognition of group-related information, but memory for behavior of particular targets and reputation guessing are modulated by common group membership.

2.5.1 Memory advantages for inconsistencies and ingroup differentiation

The present results are consistent with prior findings showing that advantages in reputational or person behavior memory are driven by general memory mechanisms. Specifically, enhanced reputational memory for uncooperative ingroup members can emerge, first, as a consequence of violations of expected ingroup positivity, and second, through the increased relevance of ingroup members.

First, reputational memory has been shown to be enhanced when target behavior violates prior expectations of the targets (e.g., Bell et al., 2015; Kroneisen & Bell, 2012). Studies in group contexts have found that memory performance for schema-incongruent information is higher than it is for schema-congruent information. Schemata such as stereotypes manifest in guessing biases (Ehrenberg & Klauer, 2005; Stangor & McMillan, 1992). Uncooperative ingroup members violate general assumptions of ingroup trustworthiness and cooperativeness. Expectations of cooperativeness of novel ingroups have been suggested to derive from a naïve understanding of groups being a container of helping and fairness (Yamagishi et al., 1999). Previous research has shown that ingroup biases might emerge through ingroup elevation, but not necessarily from outgroup derogation (Brewer, 1999; L. Gaertner, Iuzzini, Witt, & Oriña, 2006; Perdue et al., 1990). We found that even novel and experimentally created ingroups elicit expectations of positive ingroup behavior. Unlike cooperative ingroup members (or any outgroup member), uncooperative ingroup members are inconsistent with these expectations. The findings in Study 1 and Study 2 indicate that these expectations were strong, as only strong stereotypes or schemas elicit inconsistency effects in memory and expectancy-based guessing (Gawronski et al., 2003; Küppers & Bayen, 2014).

A positive ingroup bias, but no memory bias for cheaters, also shows in Study 3. However, incongruity can also emerge through rarity or atypicality of the target behavior, which evoke a better reputational memory (Bell et al., 2010). As reported, cheating and trustworthy behavior were perceived less realistic than neutral behavior in the pre-test, and remembered best.

Second, our results are also in line with the finding that memory is better for (self-) relevant information. Self-relevance has been shown to enhance reputational memory for others' uncooperativeness (Bell, Giang, et al., 2012). The ingroup is part of the self. Thus, ingroup members and their behavior are more important than outgroup members. In other words, ingroup members are more likely to be perceived and remembered on a person level (Boldry et al., 2007; Brewer et al., 1995). However, not all ingroup members are remembered individually. Reputational memory is particularly good for uncooperative ingroup members, but less so for cooperative ingroup members (and worst for neutral targets, Study 2). Moreover, participants had less reputational memory for individual outgroup members than for individual ingroup members. They had to rely more heavily on expectancy-based guessing when indicating outgroup members' reputations.

Moreover, memory in intergroup contexts differs by the meaningfulness of category distinction (Brewer et al., 1995). In Study 1 and Study 2 the minimal information was the only basis for creating group meaning. Hence, only group membership of the targets indicated their relevance to the self. In Study 3 the meaningfulness of the groups was represented by higher ingroup identification. Possibly, lower identified participants considered the common group of students as more important than the subgroup differentiation of university affiliation. Additionally, the behavioral descriptions captured inclusive standards of students at both institutions. Hence, the low salience of intergroup differences could account for the finding that University affiliation did not generally modulate the memory for cheating and trustworthy students.

Aside from expectation-incongruity and relevance, valence (e.g., Bell & Buchner, 2010; 2011; Bell, Buchner, Erdfelder, et al., 2012) and rarity (Bell et al., 2010; Volstorf et al., 2011) also enhance memory for others' behavior. Indeed, our results show that reputational memory is better for emotional content (cooperative and uncooperative behavior) than non-emotional content (neutral behavior). However, the observed effects were not solely driven by valence, as uncooperative ingroup and outgroup members did not differ in fairness and likability ratings. Moreover, even though uncooperativeness in the real world might be rather rare or uncommon (because cooperation is the norm), it was not in the present studies. The prevalence of uncooperative (or cheating) group members was equal in both groups.

Prior findings demonstrated that ingroup faces are also individualized more, and thus are recognized better than outgroup faces (e.g., Bernstein et al., 2007; Hugenberg et al., 2010). In contrast, the present studies show that reputational memory – but not face recognition – was sensitive to the ingroup context. Participants were less likely to recognize faces associated with neutral behavior than faces associated with either uncooperative or cooperative behavior. This suggests that target reputation and group membership might have been more relevant than the facial appearance of targets.

In sum, our results indicate that individual-based memory is applied primarily to targets that violate (positive) ingroup expectations. It has been argued that negativity within the group is attributed to a selected few individuals in order to preserve positive ingroup expectancies (Gramzow et al., 2001; Sherman et al., 1998). Thus, the individual memory seems to enable differentiation between uncooperative group members and others.

2.5.2 Reputational memory facilitates ingroup cooperation

The present findings suggest that simple memory mechanisms provide adaptive advantages in social contexts. Having an enhanced memory for uncooperative ingroup members may be useful, as most cooperation takes place within groups (Brewer & Kramer, 1986; Yamagishi et al., 1999), and remembering that a person has behaved uncooperatively is important for future encounters with them. This accounts for groups in which members withhold frequent reciprocal interactions, but also reputational networks (Nairne & Pandeirada, 2008). For example, reputational memory is better when people interact with others (even virtual targets) than after mere observation (Bell, Buchner, Erdfelder, et al., 2012; Bell, Buchner, Kroneisen, et al., 2012; Wilkowski & Chai, 2012). In addition, when compared with uncooperative outgroup members, uncooperative ingroup members are punished more frequently by fellow group members. Third-party punishers sometimes even invest personal costs in order to administer punishment (Fehr & Gächter, 2002). A better memory for uncooperative ingroup behavior enables avoidance and/or punishment and enhances the adherence to norms. By focusing on those uncooperative individuals with whom cooperation is more likely, group members may avoid future disappointments. Similarly, rewards for positive actions (respect, incentives, etc.) foster pro-social tendencies and moral behavior towards others (Balliet et al., 2011; Pagliaro et al., 2011). Attending to fellow group members' behavior that deviates from the group norm could enhance future ingroup interactions, and guide individuals to either cooperate or defect. This accounts for groups in which members withhold frequent reciprocal interactions, but also reputational networks (Nairne & Pandeirada, 2008). Thus, a group of highly identified individuals is able to maintain high levels of cooperation despite the possibility that some individuals may take a free ride.

In contrast, people cooperate less with outgroup members, and have fewer reasons to remember their behavior. However, this does not necessarily indicate specific cognitive modules for cheater memory (e.g., Cosmides & Tooby, 1992). The flexibility of memory for relevant expectation-incongruent information and schematic guessing allows coordinating behavior in changing environments, and across a variety of situations (e.g., Bell & Buchner, 2012).

2.5.3 Limitations and future research

The present research has limitations. First, the information about group membership may have been less salient relative to the amount of individual information about the presented target (i.e., individual ingroup and outgroup members). Thus, the influence of group membership might be weaker than other minimal group studies have shown. Despite this unevenness of individual and group information, we still found that group membership modulates reputational memory. While the present effects of group membership seem robust, future studies may try to enhance the salience of group categorization by introducing intergroup competition or conflict.

Second, identification and RWA have been interpreted as influencing reputational memory in Study 3, without demonstrating an empirical causality. The direction of the effect was derived from former research and theoretical assumptions (e.g., Hastie & Kumar, 1979).

Third, although we used a scale designed to disentangle facets of identification in Study 3 (Leach et al., 2008), our sample collapsed over the different facets. It would be interesting to explore whether the facets (e.g., self-investment, self-definition) affect memory for ingroup cheating and trustworthiness in distinct ways. Future studies could modulate and compare expectations about the groups. We would expect that, for example, an uncooperative group elicits better reputational memory for cooperative group members.

Fourth, we supposed that prospective cooperation could enhance reputational memory advantages for ingroup members. Ingroup uncooperativeness in Study 2 and 3 did not threaten the group members economically, but the group's moral standing. Moreover, trustworthiness of fellow students in Study 3 might have threatened the participants' personal moral standing (Brewer, 1991; Pagliaro et al., 2011; Parks & Stone, 2010). Here, the results are in line with research showing that cheater memory is driven by moral concerns and only biased by self-interest (Bell et al., 2014). We suppose that uncooperativeness is a special case of deviance that occurs in social interactions. The 'Black Sheep Effect' states that any deviant ingroup member is regarded less favorably than a deviant outgroup member, even beyond social exchange situations (Abrams et al., 2000; Branscombe, Wann, Noel, & Coleman, 1993; Marques & Paez, 1994). Considering that memory is better for unusual target behavior (Kroneisen & Bell, 2013; Volstorf et al., 2011), there should be a reputational memory advantage for ingroup deviants. Reputational memory should be observed in instances that do not include social exchanges or potential exploitation, and perhaps not even negativity – for example, walking barefoot on city streets.

Finally, we clarified the influence of the uncooperative target's group membership on memory. However, a common identity with victims also causes strong reactions towards perpetrators (Bernhard, Fehr, et al., 2006; Gordijn et al., 2001; McAuliffe & Dunham, 2016). Since interaction mostly takes place within groups, further investigations could extend the design to include the victim's identity. This could reveal whether it is the act of uncooperativeness or the exploitation of particular others (e.g., ingroup, outgroup victims) that determines the memory advantage.

2.5.4 Conclusion

It is helpful to remember when others behave uncooperatively, especially when they indicate cooperation by virtue of a shared group membership. Mutual cooperation requires a common set of norms that facilitates trust in others and the coordination with interaction partners (Brewer, 2007). This is more important within the ingroup, where most interactions take place. The presented studies extend prior findings on memory for uncooperative person behavior by specifying the social context: others' deviant behavior is remembered better when they belong to the same social group as us. The findings demonstrate that mere group membership (i.e., of minimal groups) triggers the enhanced memory for uncooperative ingroup members, as it elicits expectations of ingroup cooperativeness. In natural groups differential concerns influence in reputational memory for targets. With high relevance of the ingroup a better reputational memory for relevant ingroup members, compared to according outgroup members emerged. RWA led to an enhanced memory of uncooperative ingroup compared to uncooperative outgroup members.

The present studies support the notion that people strongly expect fellow group members to behave cooperatively in unspecific groups and normatively in natural groups. Moreover they show that people differentiate ingroup members' behavior more than that of outgroup members. Despite the novel ingroups and/or the high frequency of uncooperative group members, positive ingroup impressions were preserved. Subsequently, people trust in successful coordination and cooperation with fellow group members, but deviant ingroup members become infamous.

Research line II

3. A Tort is not a Crime: Moral Outrage versus Empathic Anger

3.1 Introduction

Research Line I showed that uncooperative ingroup members are remembered especially well, supposedly because they violate expectations of other's behavior. It remains unclear whether this serves the protection of common norms or potential victims. Bell and colleagues (2014) showed that the wrongfulness of moral violations rather than the implications for the self trigger memory for deviants. *Research Line II* examines which property of moral violations triggers emotional responses, the wrongfulness or the harmfulness.

The news headline: "Bus driver murders 46 children" shocks and angers us. We might even feel morally outraged. But what are we outraged about when we are confronted with such moral violations? Anger at the bus driver could emerge, because murdering children "just feels morally wrong". Alternatively, anger at the bus driver could emerge, because we feel for the children. This admittedly wicked example illustrates that enraging moral violations usually coincide with their consequences (i.e., suffering victims). Two competing suggestions currently have been raised about the elicitor of anger about moral violations. Haidt (2003) states that violations of moral standards trigger such anger (moral outrage). Batson and colleagues counter that moral outrage is disguised anger about consequences for the self or cared-for-others (empathic anger). They have shown that implications for the self increase anger about moral violations (Batson et al., 2009; Batson et al., 2007; O'Mara, Jackson, Batson, & Gaertner, 2011). However, it requires more evidence to conclude that moral outrage does not exist. To the best of our knowledge, these assumptions have not yet been tested in a design that puts them in competition. Section 3 aims at differentiating moral outrage and empathic anger by orthogonally crossing their appraisal situations. We suppose that wrongfulness (perpetrator's intentions) triggers moral outrage, whereas the harmfulness (suffering experienced by a cared-for-other) triggers empathic anger.

3.1.1 Antecedents and functions of anger

The current work draws on the notion that changes in the appraisal situation alter emotional responses. Appraisal theory defines emotions as a process that links the evaluation of a situation with arousal (i.e., affective response), the subjective experience of a feeling, and a motivational component with action tendencies (Moors, Ellsworth, Scherer, & Frijda, 2013). Anger is the negative feeling elicited by harmful moral violations. It is accompanied by the impulse to punish the perpetrator. Recent scientific approaches distinguish different kinds of anger about moral violations in terms of their elicitors. The recognition that behavior is unacceptable has been suggested to trigger *moral outrage.* Moral outrage thus is a disinterested moral emotion, triggered exclusively by the wrongfulness of the deed. It emerges even in situation in which the observer has no "reason" to react. This assumption amongst others is based on the finding that victimless moral violations are predicted by affective reactions (Haidt, Koller, & Dias, 1993). Similarly, moral outrage has been found to emerge even when the self has no stake in the situation, for example when reading about unjust events (e.g., Batson et al., 2007; Haidt, 2003; Montada & Schneider, 1989). Moral outrage motivates punishment in order to get back at the perpetrator. The perpetrator's suffering expresses condemnation of the perpetrator's action and restores just balance (e.g., Carlsmith et al., 2002; Darley, 2002; Darley & Pittman, 2003; de Rivera, Gerstmann, & Maisels, 2002; Tangney, Stuewig, & Mashek, 2007).

While pursuing moral outrage, Batson and colleagues (Batson et al., 2009; Batson et al., 2007) suggested that not the wrongfulness, but the recognition that the self or a cared-for-other suffer from undeserved harm elicits anger. *Personal anger* evolves when the consequences of a moral violation disadvantage or harm the self (Batson et al., 2007). Subsequently, people engage in revenge when they feel offended to protect themselves (e.g., Carlsmith, Wilson, & Gilbert, 2008; Gollwitzer, Meder, & Schmitt, 2011). Self-involvement can also emerge indirectly, for example via identity-relations (Batson et al., 2009) or empathy with the victims (Batson et al., 2007). *Empathic anger* is the response to another person's suffering out of concern for the other's wellbeing. This vicarious emotion requires an accurate perception and understanding of the other's experience. In the current work empathy is conceptualized as the perception of other's suffering, even though there are many different approaches to define empathy (Batson, 2009; Cuff, Brown, Taylor, & Howat, 2016). Empathic anger is an interested emotion. It motivates the observer to protect the interest of the victim by undoing the harm, compensating the victim and/or punishing the harm-doers

(Batson et al., 2007; Darley & Pittman, 2003; Hoffman, 1990, 2000; Vitaglione & Barnett, 2003). Thus, the harmfulness of a moral violation, and to some extend personal involvement with victims, triggers empathic anger, whereas its wrongfulness elicits moral outrage.

Moral violations trigger affective responses (i.e., "this just feels wrong") to victimless offenses (Haidt et al., 1993). However, those enraging immoral actions are usually located in the domains of fairness, justice, or harm (Haidt, 2003; Mikula, Scherer, & Athenstaedt, 1998; Rozin, Lowery, Imada, & Haidt, 1999; Russell & Giner-Sorolla, 2011). They include a perpetrator who commits the moral violation and a victim who suffers from the consequences. Some researchers suggest that all moral violations necessarily induce the perception of a blameworthy perpetrator and a suffering victim. For example, disobeying leaders must not have harmful consequences, but under certain circumstances it could cost a soldier's life in the battle field (Gray, Waytz, & Young, 2012). Consequently, it is hard to determine what anger about moral violations really is about. The identification of features that cause moral outrage independently from the perception of suffering victims might clear ambiguities about moral outrage and empathic anger.

3.1.2 Intentions and harm in moral judgment

I argue that findings in moral psychology uncover elicitors of moral outrage, as affective reactions predict moral judgment, blame and punishment (Goldberg et al., 1999; Greene & Haidt, 2002; Haidt, 2007; Nelissen & Zeelenberg, 2009; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). The perpetrator's intention is a key aspect of moral condemnation and punishment. It has received considerable attention. Intentionality refers to the desire, believe, and initiative to take action in order to produce a certain outcome (e.g., Cushman, 2008; Malle & Knobe, 1997). Adults (in contrast to children) give more weight to the perpetrator's intentions than the consequences of her action while rendering moral judgments (for a review, see Darley & Shultz, 1990; Piaget, 1932). Legal systems distinguish criminal offenses and accidents by the perpetrator's intentions (Mikhail, 2007). Lay people also rely more on their judgments of wrongfulness than on the harmfulness of an event when suggesting legal sanctions for criminal or immoral behavior (Alter, Kernochan, & Darley, 2007). The perpetrator's actual responsibility for the harm plays a minor role for the assignment of blame and punishment. Intentions to cause damage are sufficient to elicit moral disapproval, even in the absence of harmful consequences. In contrast, a person who causes an accident is judged leniently despite the harmful consequences (Cushman, 2008; Young &

Saxe, 2011). Bad intentions even increase blame on a person even when she did not cause the harm (Alicke, 1992; Woolfolk, Doris, & Darley, 2006). The perpetrator's intentions also magnify the perceived damage of harmful actions (Ames & Fiske, 2013; Darley & Huff, 1990).

Anger and blame are bi-directionally connected. Anger provokes a stronger tendency to attribute blame, and blame increases anger at perpetrators (Keltner, Ellsworth, & Edwards, 1993; Quigley & Tedeschi, 1996). People report higher anger when an actor is aware of breaking a taboo than when she is ignorant to her wrongdoing (Russell & Giner-Sorolla, 2011; Young & Saxe, 2011). Persons in an angry state assume more intentionality and apply more punishment when criminal intent is ambiguous (Ask & Pina, 2011). These findings suggest that the perception of the bad intentions is crucial for moral outrage, but not the actual harmful consequence of an action.

3.1.3 Pursuing moral outrage

What is called moral outrage has been suggested to be anger about actions that affect the self negatively (e.g., Batson et al., 2009; Batson et al., 2007; Hoffman, 1990; O'Mara et al., 2011). In a series of studies, Batson and colleagues showed the importance of self-involvement in consequences over wrongfulness for anger about moral violations. They did so by manipulating empathy (2007) or indicating a common group membership with the victim (2009). However, both approaches only provide limited conclusions about the existence of moral outrage. In the first article (2007), the authors showed that unfair distribution, which affected participants' personal outcomes elicited more anger compared to those affecting others instead. Only when empathy with the victim was induced, participants reported anger on behalf of others. However, the study did not take into account the negative interdependence between participants and victims: Unfairness towards others went along with a higher outcome for the self. The relief that unfairness did not hit them could have buffered the emergence of moral outrage in the "no empathy"-condition. In a subsequent study (2009) the authors described ingroup or outgroup victims who were tortured by hostile outgroups. Whereas participants judged both incidents of torture as equally wrong, the torture of an ingroup member elicited more anger than torture of an outgroup members. As anger still emerged on behalf of outgroup victims (only less), self-involvement fails to fully explain anger about torture.

3.1.4 Hypotheses and studies overview

The main aim of *Research Line II* is to determine whether the immorality of an action or the consequent suffering of moral violations elicit anger. While one major elicitor of anger may be empathy with the victim, we argue that the wrongfulness provides additional explanatory value to the emergence of anger about moral violations. Prior studies on empathic anger and moral outrage implied that moral violations produce harmful outcomes (Batson et al., 2009; Batson et al., 2007; Montada & Schneider, 1989). We suggest that varying the preconditions of moral outrage and empathic anger can shed more light on the issue. The *main hypothesis* states that perpetrator's intentions will trigger anger and punishment independently of the perception of the victims' suffering. Suffering victims will primarily elicit other emotions, such as sadness and discontent.

Three studies orthogonally crossed the occurrence of perpetrator's intentions (yes/no) and consequences (harm/no harm) to test these hypotheses. Main dependent variables were anger and punishment. Study 1 tested whether intentions or actual harm elicits anger and punishment tendencies in an ingroup context. Participants were members in sport teams, in which one member pushed a teammate. Study 2 and 3 aimed at distinguishing moral outrage and empathic anger in a situation of serious harm doing, similarly to torture (Batson et al., 2009). They describe the story of a school bus driver who missed out on taking care of his protegées. Whereas in Study 2 the storyline told about the intentionality before the consequences, we interchanged the order in Study 3. Other negative feelings (e.g., sadness, discontent) were measured to discriminate anger as unique response to intentions.

3.2 Study 1: (Un)fairness in sports teams: intentions count

Study 1 tested which feature of intentional harmful acts (i.e., moral violations), the wrongfulness of the act compared to its harmfulness, triggers more anger and subsequent punishment in an ingroup context. Thus, the participants were involved with perpetrator and victims, as they belonged to the same group. Anger about intentions and harmful consequences of fair-play violations within sports teams were assessed independently. Participants were active members in sport teams. In the scenario, one of the team players pushed a fellow member. The perpetrator's intentions were to injure a rival team member in order to get his position in the line-up. In the no intentions conditions, the perpetrator accidentally bumped into the victim. Consequently, the victim was injured and dropped out of several upcoming league games. In the no harm conditions, the victim was not hurt.

Moral outrage and empathic anger are not distinguishable, when the perpetrator harmed the rival on purpose (intentional harm). When the perpetrator accidentally injured the rival (accidental harm), anger expresses concern for the victim, and/ or for the team. Thus, anger about accidental harm demonstrates interested, empathic or personal concern. When the perpetrator intends to injure the rival, but does not achieve his goal (attempted harm), only the moral violation occurs, but no harm. As there were no consequences for the victim or the team, any reported anger illustrates moral outrage. We assume that anger only displays in regard of the perpetrator's intentions to harm the rival. Moreover, we suggest that anger triggers punishment tendencies. Less anger may emerge in response to the harmful consequences.

3.2.1 Method

Participants. A sample of 120 German hobby team sport players, mostly soccer players, was conducted. Five participants were excluded from the main analysis, because they failed to fully complete the questionnaire. A sample of 115 participants (52 female; M_{age}= 25.51, *SD*= 7.22) remained after exclusion. Participants were randomly assigned to the four conditions. 30 participants were in the condition attempted harm (intention/ no harm), 30 in the condition accidental harm (no intention/harm), 28 in the condition intentional harm (intention/ harm) and 27 in the neutral condition (no intention/no harm). They have played in their current team since *M*=5.94 years (*SD*= 6.62). 34.8 % indicated that their last received degree was the general qualification of university entrance, 23.7 % possessed a university degree.

Design. A 2 (intentions: yes/ no) x 2 (consequence: harm/no harm) between-participants analyses of variance was applied to test the hypotheses. With a power of $1\text{-}\beta = .80$ and significance level of α= .05, the sample size enabled to reliably discover main effects η^2= .06 and interaction effects of η^2= .09 (Faul et al., 2007).

Procedure. Participants were approached before or after training sessions in their sports clubs. A 10-Euro voucher was raffled between participants for incentivizing. After signing written formed consent, participants filled out the questionnaires. They indicated the sport they played, for how long they have played in their current team, and their team identification (eight items; e.g., "I see myself as part of this team.", "I enjoy being a member of this team."…; α= .80). Then participants received a vignette containing the manipulation. The victim was described as the best player on his position. The incident was located in a training session of the team. The intentional harm condition would describe that player C pushed player B intentionally in order to get an advantage in the line-up. Thereby player B was hurt and would be disabled for the upcoming training sessions and league games. The attempted harm condition contained the same text with the difference that player B was not hurt and able to continue the training session. Accidental harm was indicated through accidental pushing that injured payer B. In the neutral condition player B accidentally was pushed and continued the training session (see Appendix C for wording).

After reading the vignette, participants were asked to answer questions about each player: A (observer), B (victim) and C (perpetrator). The order of questions about player A, B and C was randomized. Only those about the perpetrator were considered with regard to the research question. First, participants reported their perception of the player's fairness ("Player B behaved fairly"; "Player B abided by the fair play rules"; α= .92).[16] Second, they indicated emotional reactions towards the player on eight adjectives describing anger and nine distractor adjectives (see Batson et al., 2007; 2009). The anger adjectives included angry, shocked, indignant etc. The index "anger" represents the anger-related items (α= .94). Third, participants reported their action tendencies towards the perpetrator on three punishment items ("should not participate in training sessions anymore"; "should be excluded from team"; "should be punished"; α= .85) and one reward item.

Finally, participants indicated how important team fairness is to them ("It is important for me that: …all players stick to the rules of the team"; "…the players in our team respect each other"; "…adhere to fair play"; α= .67). They once more filled out the identification scale to make sure that participants would not distance themselves from the team after reading about an unfair team member (Eidelman & Biernat, 2003; Marques et al., 1998). All dependent variables were measured on 7-point scales (1= *does not apply*; 7= *applies completely*). Finally, participants indicated personal data, were thanked for participation and debriefed.

3.2.2 Results

Manipulation Check. The overall identification with the team before (*M*= 6.36, *SD*= .68) and after the experiment (*M*= 6.39, *SD*= .66) was equally high, *t*(114)= -.83, *p*= .41, *d*= -

[16] Participants also reported on moral wrongness and perceived damage. Since both variables were not of primary interest they will not be reported here.

.04. Before the experiment, identification did not differ between conditions, *F*s$\leq$ 1.13, *p*s$\geq$ .29. Participants reported that the fairness norms in the team were important to them, *M*= 6.50 (*SD*= .57). This did not differ between conditions, *F*s$\leq$ 1.08, *p*s$\geq$ .26.

Fairness perceptions towards the perpetrators' behavior revealed that, perpetrator's intentions (*M*= 1.40, *SD*= .97) increased unfairness ratings compared to no intentions (*M*= 5.57, *SD*= 1.51), *F*(1,111)= 311.00, *p*< .01, η^2= .73. The fairness ratings were not influenced by the actual harm the perpetrator caused (harm: *M*= 3.56, *SD*= 2.40; no harm: *M*= 3.37; *SD*= 2.51), *F* < .01. There was no interaction effect on fairness ratings, *F*(1,111)= 1.57, *p*= .21, η^2< .01.[17] Thus, the perpetrator's behavior was only perceived unfair, when he intentionally pushed the rival (see also Table 4).

Anger and punishment. The subsequent analyses focused only on the perpetrator. Participants reported significantly more anger when the perpetrator pushed the victim intentionally (*M*= 4.63, *SD*= 1.42) in contrast to accidentally (*M*= 2.22, *SD*= 1.11), *F*(1,111)= 102.83, *p*< .01, η^2= .48. Anger did not differ in terms of the consequences (harm: *M*= 3.49, *SD*= 1.89; no harm: *M*= 3.39; *SD*= 1.62), *F*(1,111)= .73, *p*= .40, $\eta^2\leq$ .01. There was no interaction effect on anger, *F*(1,111) $\leq$ 1.11, *p*= .29.

Willingness to punish differed in terms of perpetrator's intention and slightly in terms of harm. Participants were significantly more willing to punish the perpetrator when he intentionally pushed (*M*= 3.18, *SD*= 1.70) than when the pushing was an accident (*M*= 1.13, *SD*= .52), *F*(1,111)= 78.96, *p*< .01, η^2= .41. In contrast to anger, more willingness to punish was reported, when the victim was injured (*M*= 2.35, *SD*= 1.98) in contrast to when he was not injured (*M*= 1.98, *SD*= 1.48), *F*(1,111)= 3.96, *p*= .049, η^2= .02. There was no interaction effect on willingness to punish, *F*(1,111)= .98, *p*= .33, η^2= .01. The willingness to reward the perpetrator did not differ between conditions, *F*s$\leq$ 1.50, *p*s$\geq$ .31.

[17] A 2 (intentions: yes/ no) x 2 (harm: yes/ no) x 3 (player: A/ B/ C) analysis of variance with player as within-participant factor on the fairness ratings of the players was conducted. A main effect of player revealed that the perpetrator's behavior overall was rated less fair than that of the others', *F*(2,222)= 137.90, *p*< .01, η^2= .40. An interaction effect of player and intentions showed that participants perceived the perpetrator's behavior least fair, when he intentionally pushed the rival, *F* (2,222)= 95.70, *p*< .01, η^2= .28. There was no interaction effect of player and harm on fairness ratings, *F*(2,222)= 1.05, *p*= .35, $\eta^2\leq$ .01.

Table 4. Means and standard deviations of fairness ratings, anger and willingness to punish for each condition, Study 1 (*N*= 115), Section 3

	Intention		No intention	
	Harm	No harm	Harm	No harm
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Fairness ratings	1.55 (1.27)	1.25 (.54)	5.43 (1.52)	5.72 (1.51)
Anger	4.87 (1.48)	4.41 (1.34)	2.20 (1.19)	2.25 (1.04)
Willingness to punish	3.54 (1.75)	2.84 (1.60)	1.24 (.69)	1.01 (.06)

Note: M=*mean,* SD= *standard deviation*; *all variables varied from* 1 = *"not at all" to* 7 = *"very much"*

It was assumed that intentional pushing would elicit moral outrage, which motivates punishment (Carlsmith et al., 2002; Cushman, 2008). Anger about harm should not be related to punishment tendencies towards the perpetrator. A bootstrapping procedure (Preacher & Hayes, 2008) tested whether the relation between perpetrator's intentions and punishment tendencies towards the perpetrator was mediated by participants' anger. Confirming our prior results, the path from intentions (yes = 1; no = 0) on anger was positive, indicating that intentional pushing elicited greater anger than accidental pushing (a-path), β= .69, *t*(111)= 10.13, *p*< .01. Anger was positively related to punishment tendencies (b-path), β = .45, *t*(111)= 4.90, *p*< .01. Intentionality of the action explained a significant part of the variance of punishment tendencies (c-path), β= .63, *t*(111)= 8.70, *p*< .01. This relation remained significant, when entering the mediation through anger (c'-path), β= .32, *t*(111)= 3.53, *p*< .01. The indirect effect of anger as a mediator was significant, since the bias-corrected confidence interval estimated through bootstrapping did not include zero, β= .31, 95%CI = [.16, .49]. Thus, anger mediated the relationship between the perpetrators' intention and punishment of the perpetrator. In contrast, a mediation model with the independent variable consequences (Harm=1; no Harm=0) showed little direct effect on punishment, β= .23, *t*(111)= 1.24, *p*= .21, and no mediation by anger, β= .04, 95%CI= [-.21, .32].

3.2.3 Discussion

The aim of Study 1 was to determine which feature of moral violations trigger anger, the wrongfulness or the harmfulness within groups. Perpetrator's intentions and the harmful consequences were independently manipulated. A member of a sports team pushed a rival, either to benefit in the line-up or accidentally. Consequently, he either injures the rival or no damage occurred. The results show that anger is sensitive to perpetrator's intentions (i.e., the wrongness of the harmful act), but not the actual harm.

In the harm conditions, the victim was not able to support the team in the next games. As he was introduced as valuable team member, the whole team suffered from this consequence. The findings indicate that this was not crucial for anger. Anger about intentional harm only slightly exceeded anger about attempted, but not achieved, harm. Thus, anger was not driven by the victim's suffering, which is crucial for empathic or identity-related anger (Batson et al., 2009; Batson et al., 2007). In line with prior suggestions (e.g., Darley & Pittman, 2003), the relation between the intentional pushing and punishment tendencies was mediated by anger. The relation between actual damage and punishment, however, was independent of anger.

Alternative explanations could account for these findings. First, the fairness violation threatened the moral validity of the team. Even though the perpetrator did not succeed in the attempted harm condition, he symbolically harmed the group by disregarding common values (Marques et al., 2001; Okimoto & Wenzel, 2010). Thus, anger about moral violations was not really disinterested, but had important implications for the ingroup. Second, the perpetrator's intentions imply a disposition to inflict harm on others. People's intentions are used as reliable predictors of future behavior (Waytz et al., 2010). Anger about intentions (and subsequent punishment or avoidance of the perpetrator) could protect from potential personal or team harm in the future.

3.3 Study 2: Shocking news: bad intentions, no damage done

In Study 2, the scenario represented a severe and shocking incidence. Reactions towards shocking moral violations usually are described as quick affect-laden processes (Haidt, 2007; Monin, Pizarro, & Beer, 2007). Empathy with the victims was measured as a proxy for the reaction to the suffering of others. Thus, Study 2 aims at emphasizing the differences between emotional reactions to moral violations and victims' suffering. As in Study 1, the occurrence of perpetrator's intention to inflict harm (intention: yes/no) and the consequence (consequence: victim harm/ no victim harm) were independent variables. Moreover, Study 2 addressed limitations of Study 1 by the following modifications: there was no identity relation between participant and victim or perpetrator; the harmful consequences did not hinder ingroup goals; and the perpetrator died at the end of the story. The latter was implemented to control for potential repetition of the deed. The scenario presented in Study 2 described a school bus driver who killed or almost killed his protégées, either intentionally or by accident. We assume that more anger emerges as response to the bus driver's intentions than to the actual suffering of the victims. Moral outrage is anger about moral violations that emerges independently from empathic involvement. To indicate the existence of moral outrage, more anger than empathy should emerge in the attempted harm condition, and simultaneously empathy should exceed anger in the accidental harm condition. Moreover, willingness to punish should be more sensitive to intentions than the victims' suffering, as it is the motivational component of moral outrage.

3.3.1 Method

Participants. The sample consisted of 88 students at the Polytechnical University Jena, Germany. One participant did not fully complete the questionnaire and was therefore excluded from analysis. 87 participants (43 female; M_{age}= 23.81, *SD*= 2.69) remained in the sample. Mean religiousness of participants was 2.49 (*SD*= 1.81), mean political interest was 4.25 (SD= 1.59) and mean political orientation was 3.31 (*SD*= 1.02). There were 22 participants in each condition, except of 21 in the neutral one (no intention/ no victim harm). Political interest, political orientation, and religiousness did not differ per conditions.

Design. We applied a 2 (intentions: yes/ no) x 2 (consequence: victim harm/ no victim harm) between-participant analyses of variance to test the hypotheses. With a power of $1\text{-}\beta$ = .80 and significance level of α= .05, the sample size enabled to reliably discover main effects η^2= .08 and interaction effects of η^2= .12.

Procedure & Material. Data collection took place in the entrance hall of the university. After signing written informed consent, participants were handed the situations. We claimed to investigate reactions to varying text styles. Then participants read one of four scenarios (see Appendix D for wording). In the intention conditions, the school bus driver "Ulrich H." fantasized about driving off a cliff to kill the school children. One day he decided to act out his fantasy and loosened the bus's breaks. In the no intention conditions, Ulrich H. did not notice that the breaks loosened by themselves. In the victim harm conditions, the bus fell off the cliff and the 46 children and the bus driver died. In the no victim harm condition, Ulrich suddenly died of a heart attack before he started to work that day. The school bus received a general overhaul.

After reading the scenario participants first indicated the intensity they experienced seven anger-related and other emotion-related adjectives concerning the scenario. We combined adjectives to an index "anger" (angry, indignant, etc.; α= .92). Additionally, we explored responses to other emotions: "fear" (afraid, intimidated; α= .69), "sadness" (sad, depressed, uninvolved (-), etc.; α= .67), and "content" (satisfied, calm; α=.70; see Appendix E for correlations). Participants reported how realistic they perceived the situation (plausibility) and the degree of empathy they feel towards the children and their parents. Empathy with the children and their parents correlated highly (*r* = .89) and were therefore indexed.[18] All items were completed on a seven-point scale (1 = *not at all*, 7 = *very much*). Last, participants provided some personal information (religiousness, political interest, political orientation, age, gender). Finally, they were thanked, debriefed and incentivized with a chocolate bar.

3.3.2 Results

Preliminary Results. The conditions did not significantly differ in terms of how realistic they were perceived, *F*s≤ 2.30, *p*s≥ .13. The overall mean was 4.25 (SD= 1.58).

We assessed empathy with the victims in order to determine whether their actual participants' react on the actual suffering. In contrast to our expectations, participants empathized more with the victims, when the perpetrator demonstrated bad intentions (*M*= 5.84, *SD*= 1.62) compared to no intentions (*M*= 4.86, *SD*= 2.01), *F*(1,83)= 9.21, *p*< .01, η^2= .07. A second main effect showed that empathy with the victims was higher when they were actually harmed (*M*= 6.14, *SD*= 1.55) than when victims were not harmed (*M*= 4.56, *SD*= 1.87), *F*(1,83)= 23.10, *p*< .01, η^2= .18. A significant interaction effect showed that participants reported least empathy with victims in the neutral condition, *F*(1,83)= 10.94 , *p*< .01, η^2= .09 (see Table 5). Thus, the perpetrator's intentions elicited empathy with the victims even when no actual harm occurred.

Emotional Reactions. A MONAVA showed a marginal multivariate main effect of intention on all four emotions: anger, fear, sadness, and content (-), *V*= .10, *F*(4,78)= 2.13, *p*= .09. There was also a significant multivariate main effect of harm on all emotions, *V* = .16, *F*(4,78)= 3.65, *p*= .01. There was no interaction of the two dependent variables across emotions, *V*= .03, *F*(4,78)= .54, *p*= .70. Thus, main effects were determined for emotional responses.

[18] We also measured perceived damage, wrongness, guilt and empathy with the perpetrator. Since they were not of primary interest, results are not reported here.

As predicted, anger increased with the perpetrator's intentions (*M*= 4.57, *SD*= 1.61) compared to no intentions (*M*= 2.61, *SD*= 1.46), *F*(1,83)= 21.34, *p*< .01, η^2= .20. Anger did not differ for actual victim harm (*M*= 4.09, *SD*= 1.59) and no victim harm (*M*= 3.56, *SD*= 1.81), *F*(1,83)= 2.84, *p*= .10, η^2= .03. There was no interaction effect on anger, *F*(1,83)= 1.04, *p*= .31, $\eta^2 \leq$ .01. Fear, as only other emotional reaction, showed a significant main effect of intentions, *F*(1,83)= 6.36, *p*= .01, η^2= .07. Participants reported slightly more fear in the intention condition (*M*= 2.88, *SD*= 1.41) than in the no intention condition (*M*= 2.14, *SD*= 1.31). There was no main effect of harm, *F*(1,83)= 2.39, *p*= .13, η^2= .03. Thus, fear did not differ between actual victim harm (*M*= 2.28, *SD*= 1.40) and no victim harm (*M*= 2.74, *SD*= 1.39). In contrast, sadness and content did not differ due to perpetrator's intentions, *F* < 1, and *F*(1,83)= 1.39, *p*= .02, η^2= .01. However, participants reported more sadness, when the victims were harmed (*M*= 5.33, *SD*= 1.12) in contrast to when they were no harmed (*M*= 4.85, *SD*= 1.09), *F*(1,83)= 4.27, *p*= .04, η^2= .05. They were less content when harm actually occurred (*M*= 1.34, *SD*= 2.07) than when it was prevented (*M*= 2.07, *SD*= 1.14), *F*(1,83)= 9.68, *p*< .01, η^2= .10. Table 5 displays the descriptive statistics of all emotion indices. In sum, anger and fear were sensitive to the perpetrator's intentions, whereas sadness and content responded only to the harmful consequences.

Table 5. Means and standard deviations of dependent variables for each condition, Study 2 (*N*= 88), Section 3

	Intention		No intention	
	Harm	No harm	Harm	No harm
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Plausibility	4.41 (1.50)	4.22 (1.56)	4.62 (1.66)	3.67 (1.57)
Empathy with victims	6.09 (1.80)	5.52 (1.41)	6.19 (1.32)	3.48 (1.68)
Anger	4.68 (1.59)	4.46 (1.67)	3.50 (1.39)	2.61 (1.46)
Fear	2.55 (1.40)	3.20 (1.39)	2.02 (1.37)	2.26 (1.25)
Sadness	5.19 (1.29)	5.00 (.95)	5.48 (.93)	4.68 (1.21)
Content	1.45 (.94)	2.22 (1.62)	1.23 (.51)	1.90 (.98)
Willingness to punish	6.58 (1.17)	4.52 (2.16)	1.95 (1.57)	2.05 (1.63)

Note: M=*mean,* SD= *standard deviation*; *all variables varied from* 1 = *"not at all" to* 7 = *"very much"*

Willingness to punish. Participants would punish the perpetrator more severely when he intentionally inflicted harm (*M*=5.64, *SD*= 1.98) compared to accidental harm (*M*=2.12, *SD*= 1.67), *F*(1,83)= 93.09, *p*< .001, η^2= .05. As in Study 1, they assigned slightly more punishment to the perpetrator, in the harm conditions (*M*=4.41, *SD*= 2.67) compared to the no harm conditions (*M*=3.37, *SD*= 2.31), *F*(1,83)= 8.55, *p*= .004, η^2= .04. An interaction effect demonstrates that punishment tendencies were stronger when intentional actions actually lead to harm, *F*(1,83)= 6.53, *p*= .001, η^2= .03.

Anger vs. empathy. The hypothesis suggested that anger is sensitive to intentions, and empathy with victims to victim harm. Therefore we first conducted a mixed ANOVA with the additional within-participant factor reaction (empathy/anger) to find out whether the covariate empathy reduces the effect of intentions on anger noticeably. A main effect of reaction showed that participants overall reported significantly more empathy than anger, *F*(1,83)= 80.38, *p*< .001, η^2= .45. In contrast to our expectations, empathy and anger did not differ in terms of the perpetrator's intentions, *F*(1,83)= 2.21, *p*= .14, η^2= .01. As the within-participant factor did not interact with perpetrator intentions, anger and empathy were similarly affected by perpetrator's intention. An interaction effect of harm and reaction demonstrates that empathy, but not anger, increased with actual harm of the victims, *F*(1,83)= 9.57, *p*= .003, η^2= .05. The different effects of the manipulations on empathy and anger show in a significant three-way interaction, *F*(1,83)= 6.41, *p*= .03, η^2= .03.

A second test was conducted in order to find possible differences in anger and empathy. If victim empathy and anger were two distinct states, distinct relations would show in the incongruent conditions (accidental harm: high empathy, low anger; attempted harm: low empathy, high anger). Across the congruent conditions, anger and empathy should be positively correlated. Anger and empathy correlated positively in the incongruent conditions (attempted harm, accidental harm), *r* = .50, as well as in the congruent conditions (intentional harm, neutral) *r* = .59. In sum, the two tests show that anger and empathy were not elicited independently from each other. Moreover, participants perceived the suffering of the victims as a consequence of perpetrator's intentions even when no actual harm occurred.

3.3.3 Discussion

Study 2 replicated the results of Study 1, showing that anger emerges due to the perpetrator's intentions rather than due to harmful consequences for the victims. However, anger about moral violations and the perception of suffering victims (i.e., empathy with the victims) congruently emerged in attempted harm (intentions/ no victim harm) and accidental harm (no intentions/ victim harm) conditions. The scenario described in a story about a school bus driver who desired and planned (or not) to kill school children. Thereby he succeeded or died right before achieving his goal. Two features of harmful acts were separately manipulated, the perpetrator's intentions (intention) and the victims' suffering (harm).

Prior research has shown that moral outrage is the emotional response to intentional moral violation that motivates punishment (e.g., Darley & Pittman, 2003). In Study 2, the perpetrator's intentions elicited more anger and punishment than the harmful consequences. Moreover, we found that the emotions react differentially on the features intention and harm. The perpetrator's intentions elicited anger and fear, but not the harmful consequences; whereas the harmful consequences evoked sadness and dissatisfaction independent of intention.

We hypothesized that moral outrage is a reaction to moral violations (intentional harmful acts), whereas empathy is elicited by suffering of the victims. On the one side, anger was strongest, when the perpetrator displayed the intention to murder the children. On the other side empathy was strongest when the victims were harmed, but also reacted on perpetrator's intentional action. Moreover, anger and empathy did not diverge for moral violations without harmful consequence (intention, no harm) or accidental harm (no intention/ harm). Thus, participants reported to feel sorry for the victims even when the perpetrator did intend, but not succeed in killing the children. This indicates that empathic anger contributes in explaining the emergence of anger about moral violations.

The results of Study 2 show that empathy was observed irrespectively of anger, anger however, was not observed without the prevalence of empathy. There are two possible explanations for this incident. First, empathy is a necessary condition for anger about moral violations, but does not necessarily lead to it. This is in line with the suggestion that moral outrage does not exist (Batson et al., 2007). Second, the perpetrator's intentions imply suffering of the victims, because both concepts are closely connected. The surprising non-occurrence of harm could have initiated counterfactual thinking; that is the thought of "what might have happened" or could have been. Counterfactuals have been shown to elicit affective reactions regardless of the facts of an event (Roese, 1997). In line with this assumption, Gray and colleagues (2012) suggested that all moral violations are understood as dyads composed of perpetrator's agency and victim's suffering. When people experience anger, they have been found to also increasingly presume harmful consequences (Gutierrez & Giner-Sorolla, 2007). Thus, it might be difficult to separate moral violations from their harmful consequences.

3.4 Study 3: Shocking news: no damage done, but bad intentions

Study 3 extended Study 2 by important changes in order to answer the remaining question: Can intentions to harm construed independently from the perception of actual harm, and if so, do they elicit anger independently of empathy with the victims? The main purpose of Study 3 was to avoid that perpetrator's intentions imply or foreshadow victims' harm. In Study 3, the same material as in Study 2 was used. However, the order of consequences and intentions in the vignette description were interchanged. This procedure emphasized that there were no harmful consequences before the perpetrator's intentions were mentioned. Additionally, we added one sentence to the no harm condition: "The children and parents never find out about the loosened breaks on the bus (and Ulrich's evil intentions)." Apart from that, design, wording and dependent variables remained identical to those in Study 2. If this procedure separates the perception of both appraisals, we may differentiate between moral outrage and empathic anger. Moreover, we increased the sample size to obtain higher statistical power.

3.4.1 Method

Participants. 210 participants (118 female; M_{age}= 21.18, *SD*= 3.15) took part during the introductory social psychology lecture. Mean religiousness was 2.46 (*SD*= 1.73), mean political interest was 4.24 (*SD*= 1.48), and mean political orientation was 2.85 (*SD*= 1.00). 55 participants were each in the intentional harm and the attempted harm conditions, 51 in the accidental harm, and 49 in the neutral condition. Political interest, political orientation, and religiousness did not differ per conditions.

Design. As in Study 1 and 2, the effects of intention and harm on the dependent variables were determined through 2 (intention: yes/ no) x 2 (consequence: victim harm/ no victim harm) between-participants analyses of variance. The sample size enabled to detect small effects of the manipulations on dependent variables. It provided the sensitivity to detect main effects of η^2= .06 and interaction effects η^2= .08 of with a power of 1-β= .95 and a significance-level of α= .05.

Procedure & Material. As mentioned, the scenario in Study 3 resembled that of Study 2, but participants were informed about consequences before the perpetrator's intentions. We created an index for each emotion: "anger" (α= .91), "fear" (α= .77), "sadness" (α= .64), and "content" (α=.70; see Appendix F for correlations). All other dependent measures were identical to those in Study 2. Empathy with the children and their parents (*r* = .87) were averaged to indicate empathy with the victims.

3.4.2 Results

Preliminary. Participants indicated that the story was equally plausible in terms of intention (intention: *M*= 4.11, *SD*= 1.43; no intention: *M*= 4.24, *SD*= 1.63) *F* < 1. When harm of the children occurred (*M*= 4.67, *SD*= 1.39), participants perceived the story as more plausible than when it was prevented (*M*= 3.67, *SD*= 1.50), *F*(1,204)= 24.69, *p*< .01, η^2= .11. There was no interaction effect on plausibility, *F*(1,204)= .21, *p*= .65, $\eta^2 \leq$ .001. Plausibility was entered as a covariate in the following analyses, as emotions are sensitive to the degree their appraisal are perceived as real (Frijda, 1988).

Empathy with victims was higher when the perpetrator showed intention (*M*= 5.59, *SD*= 1.79) in contrast to no intention (*M*= 4.89, *SD*= 2.13), *F*(1,206)= 9.46, *p*< .01, η^2= .03. Participants reported to feel more empathy with victims when they were actually harmed (*M*= 6.17, *SD*= 1.35) in contrast to when their harm was prevented (*M*= 4.32, *SD*= 2.10), *F*(1,206)= 53.91, *p*< .01, η^2= .20. An interaction on empathy with victims indicated that empathy increased with intention even when no harm occurred, *F*(1,206)= 6.76, *p*= .01, η^2= .02.

Emotional reactions. We entered all four emotions into a MANCOVA. There was a significant effect of intention on the emotions, *V*= .41, *F*(4,200)= 34.00, *p*< .01, and a significant effect of harm on emotions, *V*= .10, *F*(4,200)= 5.61, *p*< .01. However, there was no interaction of intention and harm on all emotions, *V*= .03, *F*(4,200)= 1.65, *p*=.16. Thus, we report on the univariate results. Table 6 displays the descriptive statistics of different emotional reactions.

Participants reported significantly more anger when the perpetrator displayed intentions (*M*= 4.11, *SD*= 1.63) compared to no intentions (*M*= 2.18, *SD*= 1.12), *F*(1,203)= 98.16, *p*< .01, η^2= .32. As in the previous studies, there was no effect of actual harm (*M*= 3.30, *SD*= 1.65) compared to no harm (*M*= 3.09, *SD*= 1.77), *F*(1,203)= .77, *p*= . 38, η^2< .01, and no interaction effect on anger, *F*(1,203)= 1.06, *p*= .31, η^2< .01.

Fear was higher when the deed was intentional (*M*= 2.38, *SD*= 1.49) than when it as unintentional (*M*= 1.95, *SD*= 1.02), *F*(1,203)= 6.01, *p*= .01, η^2= .03. Fear did not differ in terms of consequences (harm: *M*= 2.34, *SD*= 1.34; no harm: *M*= 2.01, *SD*= 1.24), *F*(1,203)= 1.95, *p*= .16, η^2< .01. There was no interaction effect on fear, *F*< 1. Sadness and content were not affected by the perpetrator's intention, *F*s≤ 2.72, *p*s≥ .10. In contrast to anger and fear, sadness was higher when the children were actually killed (*M*= 3.57, *SD*= 1.30) compared to when no harm occurred (*M*= 2.99, *SD*= 1.28), *F*(1,203)= 6.07, *p*= .02, η^2= .03. There was no interaction effect on sadness, *F*< 1. Content increased when the harm was prevented (*M*= 1.29, *SD*= .64) compared to the children being killed (*M*=1.79, *SD*= 1.04), *F*(1,203)= 18.23, *p*< .01, η^2= .08. An interaction effect shows that participants were most content in the neutral condition, *F*(1,203)= 5.78, *p*= .02, η^2= .03. In sum, the perpetrator's intention to kill the children explained 32 % of the variance in anger. In the intention condition more anger and fear emerged, whereas sadness and content only reacted on actual harm.[19]

Table 6. Means and standard deviations of dependent variables for each condition, Study 3 (*N*= 210), Section 3

	Intention		No intention	
	Harm	No harm	Harm	No harm
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Plausibility	4.56 (1.26)	3.65 (1.46)	4.78 (1.53)	3.69 (1.56)
Empathy with victims	6.25 (1.30)	4.92 (1.97)	6.08 (1.40)	3.64 (2.05)
Anger	4.13 (1.63)	4.08 (1.64)	2.40 (1.10)	1.94 (1.10)
Fear	2.55 (1.52)	2.21 (1.45)	2.12 (1.10)	1.78 (.90)
Sadness	3.43 (1.29)	2.92 (1.39)	3.73 (1.31)	3.07 (1.15)
Content	1.33 (.73)	1.57 (.94)	1.24 (.51)	2.03 (1.10)
Willingness to punish	6.76 (.88)	6.85 (.70)	2.78 (1.51)	2.77 (1.63)

Note: M=*mean,* SD= *standard deviation*; *all variables varied from* 1 = *"not at all" to* 7 = *"very much"*

Willingness to punish. In line with outrage, participants willingness to punish the perpetrator increased substantially with the intention to kill his protégées (*M*=5.38, *SD*= 1.79) compared to no intention (*M*= 1.69, *SD*= 1.21), *F*(1,202)= 294.48, *p*< .01, η^2= .59. In contrast to Study 1 and 2, the actual harm inflicted on the children did not influence the punishment tendencies (harm: *M*= 3.80, *SD*= 2.42; no harm: *M*= 3.50, *SD*= 2.39), *F*(1,202)= 1.13, *p*= .29, η^2< .01. Intention and harm did not display an interaction effect on punishment, *F*< 1.

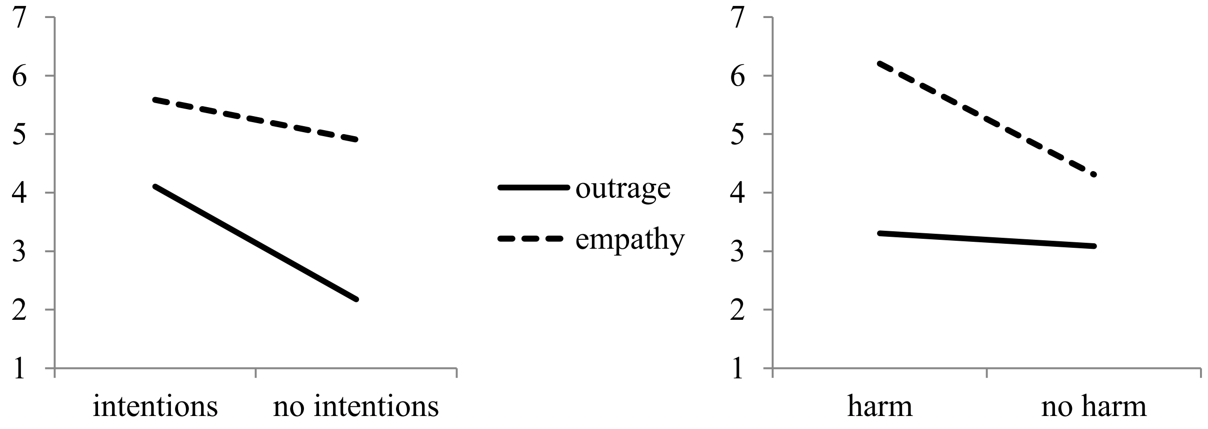
Anger vs. empathy. In order to disentangle empathic anger from moral outrage, three different methodological approaches were used. First, we looked at whether the effects of our

[19] Effects did not change in comparison no covariate (see Appendix K for details)

manipulations on anger were explained by empathy. A 2 x 2 ANCOVA with anger as dependent variable and empathy as covariate showed a main effect of empathy on anger, *F*(1,202)= 14.33, *p*< .01, η^2= .07. The covariate empathy did not change the increased anger due to intentionality, *F*(1,202)= 84.61, *p*< .01, η^2= .26. There were no other significant effects, *F*s≤ 1.06.

Second, we entered all anger and empathy as a within-participants factor in a 2 (intention) x 2 (harm) x 2 (reaction: anger/ victim empathy) mixed ANCOVA with realism as covariate. There was a significant main effect of reaction that emphasized generally higher rates of empathy than anger, *F*(1,203)= 21.59, *p*< .01, η^2= .08. A significant interaction effect of intention and reaction demonstrated that the difference between anger and empathy increased, when intentions were implied, *F*(1,203)= 21.55, *p*< .01, η^2= .08. Whereas victim empathy did not change with the perpetrator's intentions, anger increased substantially when intentions where described. There was a substantial difference between empathy and anger due to actual harm, *F*(1,203)= 34.31, *p*< .01, η^2= .12. In opposition to no harm, actual harm elicited higher rates in empathy, but not in anger (see Figure 7). There was no three-way interaction, *F*(1,203)= 2.38, *p*= .13, η^2= .01.

Figure 7. Means of anger and empathy with victims in terms of intention and harm, Study 3, Section 3



Third, correlations among congruent and incongruent conditions were tested. We combined the conditions in which we expect that moral outrage and empathic anger would collapse, and those they would differ for. A common correlation coefficient for the two conditions would be zero if empathy and anger diverge. As expected, in the intentional harm and neutral conditions, higher anger accompanied higher empathy, *r* = .56; whereas in the incongruent conditions the relationship vanished, *r* = .02. The correlation coefficients differed significantly, *z* = 4.38, *p*< .01. Hence, empathy and anger did not collapse, when either intentionality of the perpetrator (indicating moral outrage) or the victim harm (indicating empathic anger) were absent, but both related positively in the congruent conditions.

3.4.3 Discussion

Study 3 replicated the findings of Study 1 and 2 that perpetrator's intention to commit a moral violation (i.e. kill school children) elicits anger, but not the consequences of the deed. In contrast to the two previous studies, we ensured that participants noticed that the potential victims remained save in the "no harm"-conditions. As a result, anger emerged independently from empathy with the victims. Moreover, distinct emotions reacted differentially at intentions to harm others and actual harm. Only fear was slightly higher with perpetrator's intention. Actual harm increased sadness and dissatisfaction. In line with the notion that anger motivates punishment, participants were willing to punish when the perpetrator displayed a bad intention, independently of actual harm.

The assumption that moral outrage exists apart from empathy was tested through various statistical procedures (analyses of variance, correlations). The results show that empathy and anger have different antecedents (perpetrator's intentions and victim harm). The inclusion of empathy as covariate did not eliminate the strong effect of intentionality on anger. Further, a mixed ANOVA with two within-participant measurements anger and empathy showed that empathy and anger differentially react on our manipulations of intention and harm. The perpetrator's intentions triggered anger, but not empathy. In contrast actual harm triggered empathy with the victims, but not anger. The findings further showed that congruent (intentional harm, neutral) and incongruent (attempted harm, accidental harm) conditions lead to different relations between anger and empathy. Across the congruent conditions anger and empathy correlated positively. Across the incongruent conditions, empathy and anger had a null-relation.

3.5 General Discussion

Three studies tested the assumption that moral violations elicit anger and motivate punishment (i.e., moral outrage) independent from their consequences. Therefore, we orthogonally crossed the appraisals of emotional reactions to wrongfulness and harmfulness in three Studies. The perpetrator’s intentions (i.e., wrongfulness) consistently elicited more anger than harmful outcomes (i.e., harmfulness). This was the case for sports teams, where one team member pushed a fellow team member (Study 1), and also for the killing of children by a school bus driver (Study 2 and 3). The recognition of other’s suffering (i.e. empathy) was especially pronounced when actual harm was inflicted on the victims. Willingness to punish was motivated by the perpetrator’s intentionality, independent of the victims’ suffering, similarly to anger. Anger mediated the relationship between perpetrator’s intentions, and punishment, but not between harm and punishment. Punishment was only slightly affected by harm in Study 1 and 2, and not at all in Study 3. Anger differed from other negative emotional responses, such as sadness and dissatisfaction, which reacted more severely on the actual harm. Fear was also sensitive to the perpetrator’s intention to harm, but much less than anger.

Moreover, the studies showed that the perception of bad intentions and harm are closely related. In Study 1 and 2 intentions and harm were not perceived as independent. In Study 1 the team member with bad intentions harmed the team’s moral standing. In Study 2 the assessment of empathy with the victims indicated that participants perceived victims’ suffering when harm was only attempted, but not successful. Study 3 demonstrated that a clear separation of intentions and harm caused anger only to react on intentions and empathy only to react on harm.

Previous research doubted the existence of moral outrage that is an affective and motivational reaction to moral violations independent from self-involvement. Batson and colleagues’ (2007) found that anger about unfairness increases with an empathic reaction to harm inflicted on cared-for-others (i.e., empathic anger). Enraging moral violations in the fairness and care usually negatively affect victims’ interests. Indeed in many prior studies on moral outrage, the infliction of harm (or disadvantage) on others and moral violations overlap (e.g., de Rivera et al., 2002; Gutierrez & Giner-Sorolla, 2007; Haidt, 2001; Montada & Schneider, 1989). The present studies manipulated perpetrator’s intentions (precondition for moral judgments and blame, see Cushman, 2008; Darley & Pittman, 2003), independently from the actual harm inflicted on innocent victims (precondition for empathic reactions, see Cuff et al., 2016; Hoffman, 2000). The present studies show that anger is indeed closely connected to perceiving harm. They extend previous research to three insights in the relation between suffering victims and anger at perpetrators of moral violations. First and most importantly, moral outrage and empathic anger do not necessarily overlap. Second, perpetrator’s intentionality to trespass (but less the consequences) is a necessary precondition for anger, and thereby corresponds to moral judgment (Cushman, 2008; Greene, Cushman, Stewart, Lowenberg, Nystrom, & Cohen, 2009; Haidt, 2001; Russell & Giner-Sorolla, 2011). Third, the strong relationship of anger about intentions and punishment tendencies imply that anger about moral violations drives enforcement of moral standards more than anger out of concern about others' wellbeing.

3.5.1 Implications for the concept of "moral outrage"

Moral outrage is a response to moral violations in the harm and unfairness domain (Haidt, 2003; Rozin, Lowery, et al., 1999; Russell & Giner-Sorolla, 2011; Tetlock, Kristel, Elson, Green, & Lerner, 2000). People are supposed to have intuitive affective responses when someone commits moral violations, which is independent of harmfulness (e.g., Haidt, 2001; Haidt et al., 1993; Tetlock, 2002). Our results show that the same appraisals account for anger as for judgment of moral wrongness and for punishment (Cushman, 2008; Darley & Pittman, 2003). The definition of "moral emotions" often even contains concern for the wellbeing of others (Haidt, 2008; Hoffman, 1990; Tangney et al., 2007). This indicates that empathic anger and moral outrage respond to very similar appraisal situations. Prior studies have shown that egoistic, identity-related, and empathic concerns substantially increase anger about moral violations (Batson et al., 2009; Batson et al., 2007; Gordijn et al., 2001; O'Mara et al., 2011; Yzerbyt et al., 2003). Empathic anger is an interested and non-moral emotion because they respond to the disadvantage of cared-for-others. The present findings indicate that moral outrage exists by showing that failed attempts to harm elicit anger. Study 3 demonstrated that perpetrator's intentions elicit more anger than empathy with victims and conversely bad consequences elicit more empathy with the victims than anger. However, victims' suffering and anger about moral violations were difficult to disentangle. Study 2 demonstrated that failed attempts elicited more anger than achieved harm, but also empathic reactions. Moreover, in Study 3 anger and empathy with victims did not differ in terms of harm and intentions, as there was no three-way interaction. Thus, the findings should be interpreted with caution.

The dyadic model of the intentional moral agent and the suffering moral patient suggests that intentions and harm are necessarily interconnected and jointly produce moral condemnation and emotions (Gray & Wegner, 2011; Gray, Young, et al., 2012). In the presented studies this dyadic relationship of intention and harm is necessarily a part of the story. The scenarios indicated other's suffering, even if there was no real damage. Thus, at least the possibility of harm was present throughout the experiments. As other studies demonstrated before, we could also observe that the perceived suffering of the victim's increased with the perpetrators intentions in Study 1 and 2 (Ames & Fiske, 2013; Darley & Huff, 1990). Moreover, anger is a reaction towards the perpetrator, not the victim (this might be rather compassion or sadness). Thus, it was influenced stronger by the perpetrator's mental state than the victim's physical or mental state.

3.5.2 The social function of outrage at intentions

Moral emotions regulate standards which are generally relations (Haidt, 2008; Janoff-Bulman & Carnes, 2013; Rai & Fiske, 2011). Moral outrage is an other-directed emotion (in contrast to self-conscious moral emotions, such as shame and guilt) that controls others' adherence to moral standards as it motivates punishment of the perpetrators (e.g., Averill, 2001; Darley et al., 2000; Hutcherson & Gross, 2011; Nelissen & Zeelenberg, 2009). Our findings showed that intentions crucially predicted anger at the perpetrators even when they did not succeed in fulfilling their plans. Moreover, anger about intentions was a strong predictor of punishment, and less anger on behalf of victims. Intentions were suggested to signal repetition of an action (Caruso et al., 2010), and thus could directly affect the observer in the future. Accidental harms are neither repeated, nor seen as personal failure, meaning that there is no need to re-affirm moral standards. This idea is supported by the finding that perpetrator's intentions, but not the harm, elicited fear next to anger. Moreover, we would even expect a person who accidentally harms somebody else, for example in a car crash, to feel guilty and suffer him- or herself (Baumeister, Stillwell, & Heatherton, 1994; McGraw, 1987). Conversely, intentional moral violations are less likely to elicit negative moral emotions (guilt, shame) within the perpetrator. This indicates that an observer's anger might encourage the re-establishment of justice whenever the perpetrator's state itself does not satisfy just desert motives. Moral outrage has been shown to mediate the relation between perceived severity of a moral violation and punishment for just desert (Carlsmith et al., 2002). The willingness to punish is also reduced when the perpetrator signals understanding that he committed a moral violation (Funk, McGeer, & Gollwitzer, 2014; Goldberg et al., 1999). In contrast, empathy, or empathic reactions are directed at protecting the victims' interests (Darley & Pittman, 2003). Thus, trigger demands for compensation, but not the re-establishment of standards.

3.5.3 Limitations and future research

Future studies that address the limitations of the present research are recommended. First, distinct emotional reactions (i.e., subjective feeling to the scenarios) were assessed through indications on different adjectives that were presented successively. The emotional states were only conceptually distinguished, but emotional reactions might have overlapped. Further studies could focus on the differentiation of emotions in the context of moral violations (e.g., anger, fear, sadness). Second, empathy with victims was only assessed by a two-item measurement in order to control for perceived suffering. Further studies should investigate the role of empathy in regard to its complex conceptualization (Batson, 2009; Cuff et al., 2016). Third, the scenarios, especially Study 2 and 3 were designed as propriety standards, which clearly point out to wrong and right (Haidt, 2001), because those were expected to elicit most outrage. Batson (2008, p. 54) suggested that “mundane” stereotypical situations might provoke scripted responses. It remains a question for future research whether the strong relation of moral violations and outrage persist in conflict standards (e.g., fairness). Finally, we cannot conclude that moral outrage is truly morally motivated. The desirability to appear moral could contribute to reporting “moral emotions” (Batson, 2008). Taking the “moral” side in a conflict also contributes to bystanders’ reputation (DeScioli & Kurzban, 2013); that makes them a popular partner for cooperation. This instrumental use would predict the same reactions, however with severe implications for the self. Further studies investigate audience effects on anger about bad intentions.

3.5.4 Conclusion

Thus the current studies contribute to the scientific debate about moral emotions, as they tested two competing hypothesis on anger about moral violations. The very negative reactions towards bad intentions indicate that people are generally aversive towards moral violations per se, independent from the victims’ wellbeing. The news headline “Children murdered by bus driver” would elicit as much outrage as “bus driver tries to murder children”. Conversely, one could ask whether this is enough to improve social life. In Max Weber’s understanding an ethic of responsibility in sanctioning, that takes the severity of an action’s outcome into account, provides a more rational account to care for a society’s wellbeing.

Research Line III

4. Moral Outrage, Black Sheep, and their Victims

4.1 Introduction

So far, *Research Line I* demonstrated that group-based concerns enhance observer's memory for ingroup deviants. *Research Line II* found that moral violations elicit anger (and subsequent punishment) in observers independently of their consequences for the victims. However, dealing with deviance is more relevant within groups for two reasons. On the one hand, an ingroup perpetrator threatens ingroup norms and cooperation. On the other hand, ingroup perpetrators are likely to affect ingroup victims (see Section 1.2.3 and 1.2.4).

Research Line III extends this research by considering the role of wrongfulness of behavior, the perpetrators' group membership and simultaneously the victims' group membership in triggering punishment tendencies. In the subsequent studies, perpetrator and victim group membership in fair and unfair interactions were modified orthogonally (Bernhard, Fischbacher, et al., 2006). Differential reactions indicate that different sources account for anger and punishment in group contexts: First, reactions to immoral behavior in outgroup/outgroup interactions indicate a universal and impartial concern about moral adherence. Second, the concern about the maintenance of normative fit, ingroup cooperativeness, and moral standing is expressed in reactions to ingroup perpetrators. Third, reactions to ingroup victims indicate concerns about the wellbeing of fellow group members.

The present studies extend the previous research by investigating outrage and punishment in response to moral violations in dictator games with minimal groups (Study 1a, 1b, and 2), and in natural large-scale groups (Study 3 and 4). Moreover, the role of ingroup identification and intergroup relations for effects on perpetrator and victim group membership are considered.

4.1.1 Anger about and punishment of moral violations in group contexts

Anger and punishment tendencies are components of the typical emotional reaction to deviance from fairness or care standards (Mikula et al., 1998; Rozin, Lowery, et al., 1999; Tetlock et al., 2000). One significant current discussions in the study of reactions to deviants in group contexts, such as anger, are their specific appraisals: It has been questioned whether anger and punishment are triggered by moral violations (*moral outage*) or involvement with the victims (*group-based anger*; e.g., Batson et al., 2009). Moreover, a debate has emerged about whether ingroup victims or ingroup perpetrators (*black-sheep anger*) elicit punishment tendencies (Fehr & Fischbacher, 2004a; McAuliffe & Dunham, 2016). There is little published data that puts these hypotheses directly into competition (for two exceptions, see Bernhard, Fischbacher, et al., 2006; Goette et al., 2006). This introductory section provides a conceptualization of the three sources of anger and punishment.

Moral emotions, such as *moral outrage*, have been suggested to arise "just because something is wrong" (Greene, 2013; Haidt, 2001, 2003; Lerner, 1980; Skitka, 2010; Tetlock et al., 2000). They emerge with moral intuition, the sudden appearance of evaluative feeling (good or bad/ positive or negative) in response to a social situation (e.g., Greene & Haidt, 2002; Haidt, 2001). If this unspecific feeling is differentiated and powerful enough to motivate action, it is considered a moral emotion (Haidt & Kesebir, 2010). Thus, they are independent of self-involvement (Haidt, 2001, 2003; Montada, 1998; Tangney et al., 2007). In line with this argument, anger about moral violations motivates punishment of the perpetrator in order to provide just desert, and less to provide instrumental goals (e.g., Carlsmith et al., 2002; Darley, 2002; Darley & Pittman, 2003; Frank, 1988; Tangney et al., 2007).

In search for this "moral component" of anger about and punishment of moral violations, Batson and colleagues (Batson et al., 2009; Batson et al., 2007; O'Mara et al., 2011) have drawn on the importance of self-involvement. They showed in three articles that anger and punishment primarily emerges, when the self, a cared-for other or an ingroup member (*group-based anger*) is affected, rather than on behalf of strangers. Their findings suggest that the involvement with victims is crucial for eliciting anger and punishment. In Section 1.2.4 it was suggested that aversive group-based emotions are reactions towards perpetrators who harm or disadvantage ingroup members. Indeed, anger has been observed to increase for ingroup compared to outgroup victims. For example, a shared category with the victim of intentional unfair assignment to study load enhanced anger at the perpetrators (Gordijn et al., 2001). Ingroup identification is a principal determining factor of group-based emotions. Highly identified group members, and less lowly identified ones, experience anger on behalf of ingroup victims (Yzerbyt et al., 2003). Even when categorical standards[20], such as the application of torture, are violated, ingroup victims elicit more outrage than outgroup victims (Batson et al., 2009). Accordingly, punishment tendencies increase when unfair or immoral offenses affect ingroup members (Batson et al., 2009; e.g., Bernhard, Fischbacher, et al., 2006; Goette et al., 2006; Gordijn et al., 2006; Lieberman & Linke, 2007). These findings highlight that the primary source of anger and punishment is self-involvement with the victim through a shared group membership.

A third approach offers insight into the relevance of black sheep (ingroup perpetrators) in triggering anger and punishment (*black-sheep anger*). An ingroup perpetrator threatens coordination and cooperation within groups (Rabbie & Horwitz, 1988). Moreover, group members validate their moral norms and values by the perceived consensus with the ingroup and by the perceived ingroup positivity (S. A. Haslam, McGarty, & Turner, 1996). The Black Sheep Effect describes that ingroup members are differentiated from fellow group members. Whereas ingroup members are usually regarded more favorably than outgroup members, ingroup deviants are disliked and derogated harsher than outgroup deviants (e.g., Marques et al., 1998; Marques et al., 2001). In addition to harsher derogation of ingroup deviants, people report more anger and punishment tendencies towards ingroup perpetrators than outgroup perpetrators (e.g., Abrams et al., 2000; Braun & Gollwitzer, 2012; van Prooijen, 2006). Anger at ingroup perpetrators might emerge to preserve ingroup norms (see Section 1.2.3) and a positive ingroup identity (see Section 1.2.4). Most previous studies on the role of perpetrator group membership for anger about and punishment of norm violations did not specify group membership of the victim (Pinto et al., 2010), declared victims as ingroup members (Abrams et al., 2000; Chekroun & Nugier, 2011; Okimoto & Wenzel, 2010), or declared the victims as group members of the perpetrator group (Gollwitzer & Keller, 2010; Shinada et al., 2004). This bears possible confounds with the meaning of victim group membership for reactions to moral violations.

4.1.2 Intergroup relations and moral violations

Intergroup contexts influence the differential punishment of ingroup and outgroup perpetrators. Research has shown that the perpetrator's group membership can change the

[20] The basic principles for ethical behavior which should be endorsed by all humans, according to Kant (1785 /2007)

meaning of an offense towards the ingroup: Whereas offenses within the group undermine norms and values, outgroup perpetrators who offend the ingroup are perceived as threat to the ingroup's power and status (Okimoto & Wenzel, 2010). Punishment that demonstrates power is preferred when an outgroup perpetrator harmed the ingroup (Okimoto & Wenzel, 2010; Wenzel, Okimoto, Feather, & Platow, 2008). Ingroup members experience more anger against outgroup perpetrators, and are willing to take action against the offending outgroup when they perceive their ingroup as powerful (Mackie et al., 2000). An offense of the ingroup against the outgroup, however, threatens the ingroup's image, and elicits anger that predicts intentions to confront the ingroup perpetrators (Iyer, Schmader, & Lickel, 2007).

This malleability of punishment might depend on the intergroup relations. In competitive intergroup relations, group memberships become especially salient and ingroup bias enhances (Balliet et al., 2014; Brewer, 1979). A salient group context also enhances the Black Sheep Effect, and increases the tendency to punish ingroup perpetrators to re-establish ingroup positivity (Hutchison et al., 2008; Marques et al., 1988; see also Section 1.2.3). Conversely, the favorable ingroup treatment over outgroup treatment diminishes in cooperative intergroup relations (Bettencourt, Brewer, Croak, & Miller, 1992). Cooperative intergroup relations reduce intergroup biases because they encourage one-group representations of the befriended groups (S. L. Gaertner, Mann, Dovidio, Murrell, & Pomare, 1990).

4.1.3 Hypotheses and Studies Overview

The following studies address the question whether the aversion to moral violation, the norm diffusion through ingroup perpetrators, and/or the wellbeing of fellow group members trigger more anger and subsequent punishment of perpetrators. To discriminate moral outrage, group-based and black-sheep anger, we applied a full design that modulated fairness of the behavior, victim, and perpetrator group membership orthogonally. To the best of our knowledge, only two studies examined punishment of moral violations while considering different perpetrator and victim group memberships (Bernhard, Fischbacher, et al., 2006; Goette et al., 2006). They found that altruistic punishment gradually differed relative to the degree of unfairness, and was harsher when ingroup members were affected than when outgroup members were affected (see Section 1.2.3). Both studies included natural groups, with longstanding relationships and positive interdependence between ingroup members.

The present studies extend this research by the application of minimal groups and the consideration of ingroup identification. Moreover, anger as the “intuitive reaction” to moral violation is investigated, additionally to punishment tendencies. As a third novel factor two intergroup relations (cooperative/competitive) are distinguished. The scenarios in Study 1a, 1b and Study 2 described a dictator game among minimal group members. They manipulated moral violations as unequal sharing. In Study 3 and Study 4 vignette descriptions about natural groups were used. They consider intergroup relations. The moral violations described violations of the rights to freedom and safety through national authorities. Study 1a and 1b main dependent variable was (altruistic) punishment, whereas the main focus in Studies 2 to 4 was anger.

The *first hypothesis* states that moral violations elicit moral outrage regardless of self-involvement. This becomes obvious outgroup/outgroup interactions in which group-related explanations for anger are unsustainable. The *second hypothesis* proposes that a shared group membership with the victim is an important elicitor of anger and punishment. This should be dependend on ingroup identification, as especially highly identified group members experience anger on behalf of fellow group members. As *third hypothesis*, it is assumed that ingroup perpetrators elicit harsher anger and punishment compared to outgroup perpetrators. The *fourth hypothesis* suggests that perpetrator and victim group membership elicit negative reactions particularly in competitive intergroup encounters, and are less relevant in cooperative relations.

4.2 Study 1a: Punishment of ingroup and outgroup (un-)fairness (between design)

Study 1a aimed at replicating the results of Bernhard et al (2006) in a student sample with minimal groups (Tajfel et al., 1971), and considered the role of ingroup identification. Even in minimal group, favorable treatment of ingroup members validates ingroup superiority, as it is the only means to positively differentiate the ingroup from the outgroup (Spears et al., 2009). Perpetrator and victim group membership were operated as between-participant factors. Hence, each participant was confronted with either ingroup/ingroup, ingroup/outgroup, outgroup/ingroup, or outgroup/outgroup interactions. Consequently they could not compare behavior of or towards targets with other group memberships.

Unequal distributions of resources (Euros) in anonymous one-shot situations represent fair and unfair treatment. The methods ensured observations in isolated experimental conditions which minimized the amount of content information. The minimal group procedure guaranteed that any group-based differences emerged through the categorization and its subjective meaning. Participants assigned punishment to perpetrators by spending personal outcome. The main dependent variable was altruistic punishment. Additionally, participants' anger was assessed, as altruistic punishment has been suggested to be driven by negative emotions (Fehr & Gächter, 2002). People's satisfaction with their punishment and their willingness to gossip about the perpetrators were assessed. This later indicated whether establishing a bad reputation of perpetrators is desired in addition (or instead) to reducing perpetrator's outcome (Dunbar, 2004; M. Feinberg, Willer, & Schultz, 2014).

4.2.1 Method

Participants. 100 participants took part in the experiment, which was conducted at the University of Jena. Seven participants were excluded, because they were familiar with the minimal group paradigm. One reported troubles with the instructions due to low language skills. The remaining sample consisted of 92 (50 female) students from diverse study fields at the University of Jena (M_{age}= 22.13; *SD*= 2.46). There were 23 participants in each of the four conditions.

Design. We applied 2 (treatment: fair/ unfair) x 2 (perpetrator group membership: ingroup/ outgroup) x 2 (victim group membership: ingroup/ outgroup) mixed analyses of variance with treatment as only within-participants factor. The significance level in planned contrasts analyses was adjusted to multiple testing through Bonferroni-Holmes corrections. The sample size enabled the detection of small effects of η^2= .03 (within-participants) and η^2= .04 (between/ within-between participants), with a power of 1-β= .80 and a significance level of α= .05 (Faul, Erdfelder et al., 2004).[21]

Procedure. Participants signed written informed consent, and then sat at one of 10 computers. All instructions were presented on the screen. Participants read that they will complete various reading, speech and perception tasks in the following experiment. First, they completed a fictitious perceptual task by estimating the number of dots in a series of eight subsequent pictures. Supposedly based on their perception, participants were arbitrarily

[21] computed for punishment with highest correlations among repeated measures *r* = .276

assigned to the group of “under-” and “over-estimators” (see Tajfel et al., 1971). Salience of the experimental groups was increased by the statement that the differentiation could explain basic psychological differences between people (e.g., Forgas & Fiedler, 1996).

Second, we explained the rules of the dictator game, informing participants that they will observe interactions between two persons. They were informed about the group membership of the persons. The dictator (perpetrator: either ingroup or outgroup) was granted 100 Euros. The dictator was free to distribute the money between herself and the receiver (victim: either ingroup or outgroup). Participants then read about six interactions between dictator and receiver. The victims’ shares varied from 0, 10, 20, 30, 40 to 50 Euros (treatment: unfair/ fair), which appeared in random order. After each treatment, participants received 50 Euros to invest in punishment. For each Euro participants invest, the perpetrator lost 3 Euros (punishment). Then participants reported how angry they felt about the treatment (anger), how satisfied they were to have reduced the perpetrator’s outcome (satisfaction), and how much they would be willing to spread negative information about the perpetrator (willingness to gossip). Last, participants completed six items on ingroup identification (“How much do you see yourself as an over-/ under-estimator?”, “How much do you identify as an over-/ under-estimator?”, “How much do you perceive yourself and the other group members as a group?”…). All dependent variables, except of punishment, were measured on a seven-point scale (1 = *not at all*; 7 = *very much*). Finally, participants were debriefed, thanked, and received coffee and a chocolate bar as incentive for participation.

Material. In a pre-test we evaluated perceptions of fairness in dictator game decisions (*N* = 100). A principal component analysis with varimax rotation showed that fair and unfair distributions loaded on two separate factors with the Eigenvalues 2.59 (factor 1: unfair distribution) and 1.10 (factor 2: fair distribution). For analyzing results of the main study, unfair distributions were summarized in an index including all shares from 10 to 30 units (α= .89; *M*=1.93, *SD*= .79).[22] The fair distribution was 50-50 (*M*=6.50, *SD*= 1.11). The pre-test showed that unfair distributions were rated significantly less fair than fair distributions, *t*(1,99)= 29.43, *p*= .001, *d*= 4.79. In the main study, mean reactions towards unequal treatments (0-100, 10–90, 20–80, 30-70) indicate punishment (*α*= .83), anger (*α*= .89), satisfaction (*α*= .88), and willingness to gossip (*α*= .93) of unfair treatment.

[22] Donating a share of 40 to the other player loaded on both factors (loadings: unfair = .47, fair = .74), but correlated low with the 90-10 (*r* = .19) and the 50-50 distribution (*r* = .06). Hence, distributing 40 out of 100 units seems to be an ambiguous case and was excluded from further analyses in the main experiments.

4.2.2 Results & Discussion

Perliminary analyses. The mean ingroup identification after indexing all six items (α=.80) was 3.22 (*SD*= .98). Participants identified with their ingroup below midpoint, *t*(91)= -7.65; *p*= .01; *d*= -1.47.

As only the victim and perpetrator group membership, but not fairness, were between-participants factors, we applied a 2 (perpetrator group membership) x 2 (victim group members) analysis of variance to determine differences in ingroup identification between conditions. There were no differences in ingroup identification due to group memberships of victim and perpetrator, *F*s≤ 2.25, *p*s≥ .14.

Generally, participants punished more the more the perpetrators' share deviated from the fair 50-50 distribution. Overall, only 25% of the participants punished fairly behaving perpetrators. 71.5% of the participants reduced the perpetrator's outcome when she kept everything to herself. Mean satisfaction with punishing unfair perpetrators was 4.21 (*SD*= 1.41).[23]

Reactions to unfairness. The result section is organized in order of the hypotheses (moral outrage, group-based anger, and black-sheep anger), and thus does not display effects on one dependent variable sequentially. Descriptive statistics are displayed in Table 7. It was hypothesized that unfairness elicits altruistic punishment (and anger) independent of victim and perpetrator group membership. In line with the hypothesis, participants invested significantly more money to punish perpetrators in the unfair treatment (*M*= 11.69, *SD*= 8.90) than in the fair treatment condition (*M*= 3.67, *SD*= 8.27), *F*(1,88)= 53.94, *p*< .01, η^2= .38. They also reported significantly more anger in the unfair treatment (*M*=3.17, *SD*= 1.51) than in the fair treatment condition (*M*= 1.54, *SD*= 1.03), *F*(1,88)= 85.87, *p*< .01, η^2= .49. Planned contrasts[24] revealed that even in the outgroup/outgroup interaction unfair treatment elicited more punishment and anger than fair treatment (see Table 7), *p*< .01, *d*= 4.98; *p*< .01, *d*= .69. Similarly to monetary punishment, participants were more willing to gossip about the unfair treatment (*M*= 2.53, *SD*= 1.50) than the fair treatment (*M*= 1.60, *SD*= 1.28), *F*(1,88)= 40.20, *p*< .01, η^2= .30. Thus, results are in line with the first hypothesis. They correspond to previous studies that demonstrate third-party punishment of unfairness (e.g., Fehr,

[23] Satisfaction with punishment was only considered, when participants actually punished. In the further analysis, satisfaction with punishment of unfair versus fair treatment was not determined, because only 23of 92 participants punished fair treatment. 15 participants did not punish any unfair treatment.

[24] Simple effects in SPSS were determined from estimated marginal means of the 2x2x2 analysis of variance, *p*-values were Bonferroni-Holmes corrected

Fischbacher, & Gächter, 2002; Henrich et al., 2006). Bernhard and colleagues (2006), for instance, found that the tribe members in Papua New Guinea punish fair sharing significantly less than unfair sharing independently of the interaction partners' group memberships.

Reactions to victim and perpetrator group membership. There were no significant effects of victim or perpetrator group membership on punishment, anger, or satisfaction with punishment of unfair treatment, *F*s≤ 2.40, *p*s≥ .13. An interaction of victim group membership and treatment showed that participants were more willing to gossip about unfair treatment of outgroup victims (*M*= 3.05, *SD*= 1.50) than ingroup victims (*M*= 2.36, *SD*= 1.71), *F*(1,88)= 5.10, *p*= .03, η^2= .04. There was no other effect on willingness to gossip, *F*s≤ 1.68, *p*s≥ .20.

In sum, Study 1a replicated the gradual punishment of unfairness, but not the importance of victim group membership that has been reported in other studies (Bernhard, Fischbacher, et al., 2006; Goette et al., 2006). We expected punishment and anger to vary in regards to group memberships. However, group membership did not influence the participants' reactions. This lack of effects might be caused by lower salience and/or importance of the group affiliations than in previous studies.

Table 7. Means and standard deviations of dependent variables for each condition, Study 1a (*N*= 92), Section 4

		Perpetrator			
		Ingroup		Outgroup	
		Victim		Victim	
		Ingroup	Outgroup	Ingroup	Outgroup
	Treatment	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Ingroup Identification		3.15 (.97)	3.29 (.94)	3.45 (.99)	2.97 (1.02)
Altruistic Punishment	Unfair	11.09 (1.92)	10.26 (1.61)	13.18 (2.06)	12.24 (1.87)
	Fair	2.87 (1.38)	3.61 (1.48)	5.26 (2.55)	2.96 (1.23)
Anger	Unfair	3.25 (1.55)	3.37 (1.54)	2.93 (1.15)	3.14 (1.81)
	Fair	1.62 (1.20)	1.52 (.85)	1.78 (1.24)	1.26 (.75)
Satisfaction with Punishment	Unfair	4.21 (1.27)	4.54 (1.28)	3.98 (1.65)	4.11 (1.43)
Willingness to gossip	Unfair	2.24 (1.25)	2.48 (1.37)	2.36 (1.71)	3.05 (1.60)
	Fair	1.57 (1.16)	1.52 (1.04)	1.83 (1.56)	1.48 (1.38)

Note: M=*mean,* SD= *standard deviation; Altruistic punishment = personal contribution of virtual 0 to 50 Euros which were tripled and reduced from perpetrator outcome; other dependent variables varied from* 1 = *"not at all" to 7 = "very much"; satisfaction with punishment was only determined, when punishment was applied. Therefore, different N emerged per condition: ingroup perpetrator/ ingroup victim:* N= *18; ingroup perpetrator/ outgroup victim:* N= *19; outgroup perpetrator/ ingroup victim:* N= *20; outgroup perpetrator/ outgroup victim:* N= *20*

4.3 Study 1b: Punishment of ingroup and outgroup (un-)fairness (within design)

Therefore, Study 1b aimed to replicate Study 1a, but applied a within-participants design instead of varying different victim and perpetrator group membership between participants. In Study 1a, participants did not differentiate between ingroup and outgroup in terms of perpetrator behavior, or victim outcome. A within-participant design was supposed to increase comparative effort and category salience, because participants can compare behaviors and outcome of ingroup and outgroup members. Each participant saw six interactions that varied step-wise from fair to unfair in four blocks. Each block contained a different combination of victim and perpetrator group membership. Additionally, ingroup identification was assessed after the experiment.

4.3.1 Method

Participants. 31 students took part in the experiment, which was again contucted at the University of Jena. One participant was excluded, because he 50 points in punishment in every single interaction he observed. The remaining sample consisted of 30 participants (14 female; M_{age}= 22.06; *SD*= 6.44).

Design. We applied 2 (treatment: fair/ unfair) x 2 (perpetrator group membership: ingroup/ outgroup) x 2 (victim group membership: ingroup/ outgroup) within-participants analyses of variance to test the hypotheses. Ingroup identification was added as a continuous between-participants moderator for testing the group-related hypothesis. For example, an interaction effect of ingroup identification and victim group membership would indicate that reaction to victim group membership was influenced by identification. The sample size enabled the detection of small interaction effects of η^2= .05 with a power of 1-β= .80 and a significance level of α= .05.

Procedure. The procedure resembled that in Study 1a. Participants were randomly assign to minimal groups (figure- and ground-seers), based on a fictitious perceptual task (see also Study 1 & 2, Section 2, Wentura & Otten, 2002). In contrast to Study 1a, we introduced the study as an investigation of how group membership affects attitudes in social exchanges. Participants read about the six subsequent interactions in four blocks that differed in the combination of victim and perpetrator group membership. The order of the four blocks was randomly assigned. The victims' shares varied step-wise from 0, 10, 20, 30, 40 to 50 Euros. The four unequal treatments (0-100, 10–90, 20–80, 30-70) were reduced to one dimension of unfair treatment in each block. Indices had high reliability for punishment (all α's≥ .89), anger (all α's≥ .95), satisfaction (all α's≥ .94)[25], and willingness to gossip (all α's≥ .91). Ingroup identification was measured at the end of the experiment (α=.87; see Study 1a).

4.3.2 Results & Discussion

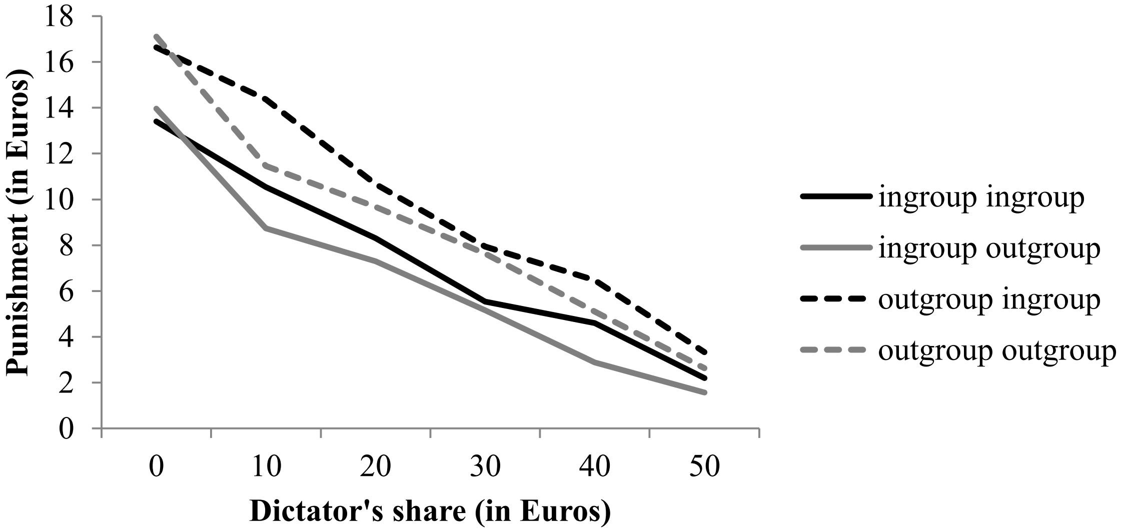
Preliminary analyses. Across the whole span of distributions, participants punished perpetrators' gradually less with increasing equality of the distribution (see Figure 8). Overall, participants inflicted more punishment the more dictator shares undercut the 50-50 distribution (see also Study 1a). With equal sharing, 18% of the participants reduced the perpetrator's outcome. Considerably more participants punished when the perpetrator kept 90

[25] As in Study 1a, satisfaction with punishment was only considered when the participants actually punished. Per condition, only 4 to 8 participants punished fair behavior. Therefore treatment effects could not be determined.

out of 100 Euros (71% on average). In ingroup/ingroup interactions only 63% of participants punished the unequal sharing.

Overall, participants were satisfied with their punishment (*M*= 4.46; *SD*= 1.52).[26]

Figure 8. Mean punishment across distributional decisions for each combination of interaction partners, Study 1b (*N*= 30), Section 4



Note: The graph specifies mean investment for punishment in the four interactions. The investment options varied from 0 to 50 Euros. The first group membership stands for the perpetrator and the second group membership identifies victims (ingroup ingroup= ingroup perpetrator & ingroup victim, ingroup outgroup= ingroup perpetrator & outgroup victim, outgroup ingroup= outgroup perpetrator & ingroup victim, outgroup outgroup= outgroup perpetrator & ingroup victim)

Reactions to unfairness. In accordance with our hypothesis, participants invested more money to punish perpetrators in the unfair treatment (*M*= 10.53, *SD*= 8.70) than in the fair treatment condition (*M*= 2.43, *SD*= 8.14), *F*(1,29)= 23.51, *p*< .01, η^2= .31 (see Table 8). Like-wise participants reported significantly more anger in the unfair treatment (*M*= 3.06, *SD*= 1.61) than in the fair treatment condition (*M*= 1.66, *SD*= 1.13), *F*(1,29)= 23.10, *p*< .01, η^2= .32. Unfair treatment elicited more punishment and anger than fair treatment even in the outgroup/outgroup interaction (see Table 8), *p*< .01, *d*= .93; *p* < .01, *d*= .75. Participants were also more willing to gossip about unfair perpetrators (*M*= 2.48, *SD*= 1.44) than fair perpetrators (*M*= 1.42, *SD*= 1.13), *F*(1,29)= 20.69, *p*< .01, η^2= .25. As in Study 1a, the stronger punishment and anger at unfairness than fairness across conditions indicates that moral standards elicited negative reactions independently of group membership.

[26] 15 participants did not punish any unfair treatment, hence satisfaction was not taken into further consideration.

Reactions to victim and perpetrator group membership. The results did not show increased negative reactions to unfair treatment towards or by ingroup members. There were no main or interaction effects of victim group membership on punishment, *F*s$\leq$ 1.82, *p*s$\geq$ .19. Conversely, punishment significantly increased when the perpetrator was an outgroup member (*M*= 7.46, *SD*= 8.67) compared to an ingroup member (*M*= 5.50, *SD*= 5.96), *F*(1,29)= 5.40, *p*= .03, η^2= .02. This overall harsher punishment of outgroup perpetrators was higher in the unfair treatment condition (*M*= 11.93, *SD*= 10.11) than in the fair treatment condition (*M*= 2.98, *SD*= 9.56). There was an interaction effect of perpetrator group membership and fairness, *F*(1,29)= 4.96, *p*= .03, η^2= .004. The result indicates that group membership gained importance compared to Study 1a. In Study 1a harsher punishment of outgroup compared to ingroup perpetrators showed as a descriptive trend. The same effect was reported by Bernhard et al. (2006). The reduction of outgroup perpetrators' outcome caused the total ingroup outcome to exceed the outgroup outcome. Other research has shown that people treat outgroup perpetrators harsher than ingroup perpetrators in competitive intergroup encounters (Goette, Huffman, Meier, & Sutter, 2012; Lieberman & Linke, 2007; Schiller, Baumgartner, & Knoch, 2014). Hence, punishment of outgroup perpetrators could have served to maximize outcome differences between groups (see also Tajfel et al., 1971).

In contrast to punishment, there was no effect on anger due to victim or perpetrator group membership (see Table 8), *F*s$\leq$ 2.10, *ps*$\geq$ .16. However, participants were more willing to gossip when an ingroup victim was affected (*M*= 2.03, *SD*= 1.23) in contrast to an outgroup victim (*M*= 1.87, *SD*= 1.06), irrespective of treatment. The difference was marginally significant, *F*(1,29)= 3.72, *p*= .06, η^2= .007. There were no other effects of group membership on the willingness to gossip, *F*s< 1. Thus, participants would rather spread reputations of a perpetrator who affected an ingroup victim than one who affected an outgroup victim, regardless of treatment. However, these results have to be interpreted with caution as they differed between Study 1a and 1b.

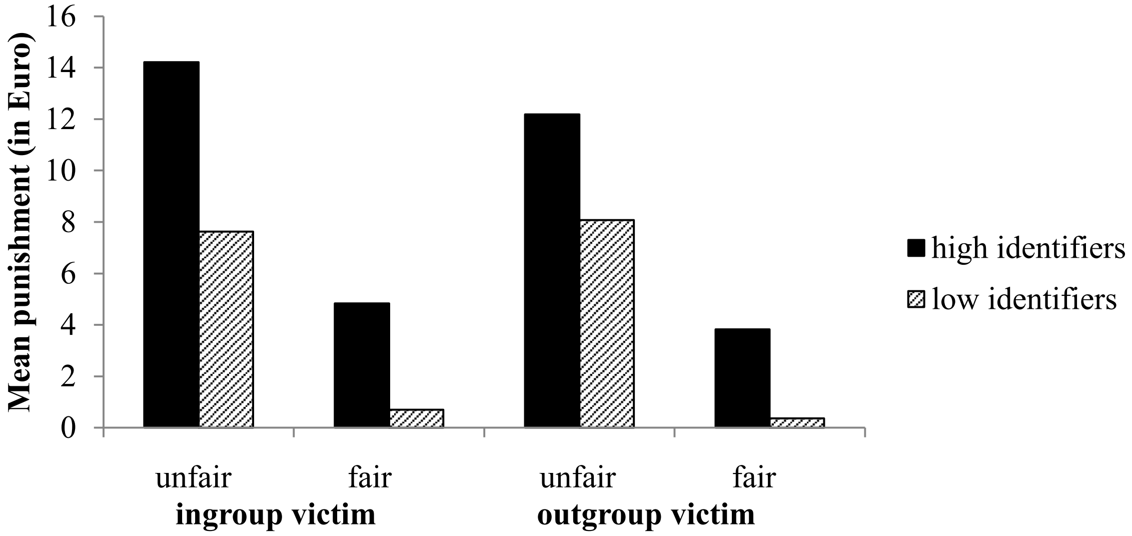
Table 8. Means and standard deviations of dependent variables for each condition, Study 1b (*N*= 33), Section 4

		Perpetrator			
		Ingroup		Outgroup	
		Victim		Victim	
		Ingroup	Outgroup	Ingroup	Outgroup
	Treatment	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Altruistic Punishment	Unfair	10.75 (10.82)	10.12 (12.29)	13.61 (13.07)	12.71 (11.78)
	Fair	3.74 (11.39)	3.13 (11.29)	4.84 (12.86)	4.16 (12.51)
Anger	Unfair	3.27 (1.76)	2.78 (1.60)	3.07 (1.81)	3.01 (1.75)
	Fair	1.68 (1.35)	1.58 (1.26)	1.58 (1.34)	1.81 (1.49)
Satisfaction with punishment	Unfair	4.63 (1.36)	4.45 (1.43)	4.56 (1.31)	4.68 (1.27)
Willingness to gossip	Unfair	2.47 (1.49)	2.21 (1.38)	2.52 (1.80)	2.53 (1.80)
	Fair	1.52 (1,39)	1.23 (.80)	1.48 (1.34)	1.39 (1.28)

Note: M=*mean,* SD= *standard deviation*; *Altruistic punishment = personal contribution of virtual 0 to 50 Euros which were tripled and reduced from perpetrator outcome; other dependent variables varied from 1 = not at all to 7 = very much; satisfaction with punishment was only determined, when punishment was applied. Therefore, different N emerged per condition: ingroup perpetrator/ ingroup victim:* N= *24; ingroup perpetrator/ outgroup victim:* N= *30; outgroup perpetrator/ ingroup victim:* N= *25; outgroup perpetrator/ outgroup victim:* N= *23*

The role of ingroup identification. Mean identification with the minimal ingroup was 3.60 (*SD*= 1.47). We added ingroup identification as a continuous between-subject moderator to see whether it interacts with reactions to different group memberships. Overall, punishment tended to be higher, when participants identified stronger with their group, *F*(1,28)= 2.66, *p*= .11, η^2= .09. In contrast to before, there was no main effect of treatment, *F*(1,28)= 2.51, *p*= .13, η^2= .05. Participants punished equally strongly in the unfair treatment and in the fair treatment condition when ingroup identification was controlled for. There were no main effects of perpetrator or victim group membership on punishment, *F*s≤ 1.80, *ps*≥ .19. However, participants punished slightly more severely with increasing ingroup identification when the victim was an ingroup member, *F*(1,28)= 4.06, *p*= .05, η^2= .007. Figure 9 illustrates the effect of ingroup identification on punishment of ingroup offenders. It shows the means derived from a median split of ingroup identification. There were no other effects on identification on punishment, *F*s≤ 2.25, *p*s≥ .15.

Figure 9. High and low identifiers' mean punishment of fair and unfair treatment of ingroup and outgroup victims, Study 1b, Section 4



*Note: Mean punishment investment (0 to 50 Euros) according to victim group membership; high (*N*= 16) and low ingroup identification (*N*= 14) separated through median split (*Mdn *= 3.33)*

Anger did not generally differ with the participants' degree of identification, *F*(1,28)= .52, *p*= .48, η^2= .02. The main effect of treatment on anger remained, *F*(1,28)= 7.56, *p*= .01, η^2= .12. Thus, anger about unfair treatment was overall higher than anger about fair treatment. Participants with higher identification also showed a tendency to report more anger when the victim was an ingroup member. The difference did not reach significance, *F*(1,28)= 2.55, *p*= .12, η^2= .01. A three-way interaction with the dependent variable anger indicated that this was mainly the case, when an ingroup victim was treated unfairly. This interaction did not reach significance, *F*(1,28)= 3.18, *p*= .09, η^2= .01. In sum, after entering ingroup identification as a between-participants moderator, the higher punishment of outgroup perpetrators was no longer meaningful. With increasing ingroup identification victim group membership gained importance for punishment and anger about unfairness. This indicates that ingroup identification enhances negative reactions to unfairness on behalf of fellow ingroup members (Brewer, 2007; Gordijn et al., 2001).

4.4 Study 2: Anger about ingroup and outgroup (un-)fairness

Study 2 was designed to replicate and extend Study 1a and 1b. The aim was to rule out prior limitations of the studies, increase group salience, and moreover focus on negative emotional reaction (i.e., anger) in more detail. Victim and perpetrator group membership varied between participants, fairness was a within-participants factor (see also Study 1a).

We particularly addressed the following limitations of the previous studies. First, we focused on anger as a dependent variable, instead of (altruistic) punishment. The punishment option in Study 1a and 1b were not truly altruistic, as they did not affect participants' real incentives. This might have increased the temptation of outgroup discrimination (Schiller et al., 2014). Negative emotions capture the concept of moral outrage better than punishment (e.g., Greene & Haidt, 2002; Haidt, 2003; Tetlock et al., 2000). Moreover, previous research has shown that anger fosters altruistic norm enforcement (e.g., Fehr & Gächter, 2002; Nelissen & Zeelenberg, 2009), and expresses concern for fellow ingroup members (Gordijn et al., 2001; Yzerbyt et al., 2003). Second, Study 2 aimed at increasing the salience and importance of minimal group membership, as the influence of group membership in Study 1a and 1b was low. Study 2 was conducted in a paper-pen format. After the perceptual task, the experimenter "evaluated" the participants' answers before handing them a note about their group membership. We supposed that this procedure increases the salience and credibility of group memberships. Third, fairness perceptions were assessed after each interaction to control for fairness biases (e.g., Tarrant, Branscombe, Warner, & Weston, 2012). Fourth, we observed one fair (50-50) and one unfair treatment (10-90). The greater distance between both treatments should increase differences in reactions to them.

4.4.1 Method

Participants. 81 students filled out a questionnaire in a separated area in the foyer of the Technical University Jena. 11 participants reported to perceive the unequal treatment as more or equally fair as the equal treatment. Thus, in these participants the manipulation of fair versus unfair treatment was unsuccessful, and they were hence excluded from the analysis. The remaining sample consisted of 70 participants (34 female, 3 without indication of gender; M_{age}= 26.07; *SD*= 6.10).

Design. We applied 2 (treatment: fair/ unfair) x 2 (perpetrator group membership: ingroup/ outgroup) x 2 (victim group membership: ingroup/ outgroup) mixed analyses of variance to test the hypotheses. In a subsequent ANCOVA, we controlled for ingroup identification for testing the identity-related hypothesis with fairness as within-participant factor. The elimination of an existing victim effect would indicate that ingroup victims elicit outrage depending on ingroup identification. The sample size enabled the detection of small interaction effects of η^2= .03 (between/ within-between participants) with a power of 1-β= .80 and a significance level of α= .05.

Procedure. After signing written informed consent, participants received a task for minimal group assignment (see Study 1b). All participants were assigned to be "figure-seers". A note informed them about alleged psychological differences between figure- and ground-seers. Then two subsequent interactions between a perpetrator (ingroup/outgroup) and a victim (ingroup/outgroup) were presented to each participant. Participants read about one equal (50-50: fair treatment) and one unequal distribution (10-90: unfair treatment). The order was randomized between participants.

After each interaction, participants rated the fairness of the distribution. They indicated how much they experience certain emotions regarding the interaction. Six emotional adjectives described anger (angry, mad, furious,…), whereas the other six were distractor adjectives (confused, ignorant, content, …; for details on the procedure see: Batson et al., 2007). The anger-items were combined to two indices of "anger". One represented anger in the fair treatment (α= .93), the other in the unfair treatment condition (α= .90). Participants also stated how much they would be willing to punish the perpetrator and their degree of identification with their ingroup (five items; α=.82). Subsequently, they read about the second interaction and answered the same questions. All items were measured on seven-point scales (1 = *not at all*; 7 = *very much*). Afterwards the participants were debriefed and incentivized with a chocolate bar.

4.4.2 Results

Preliminary Analyses. Table 9 displays descriptive statistics of all dependent variables. As intended, participants rated the equal sharing to be significantly more fair (*M*= 5.79, *SD*= .54) than the unequl sharing (*M*= 1.56, *SD*= 1.18), *F*(1,66)= 602.62, *p*< .01, η^2= .90. There were no effects of perpetrator or victim group membership on fairness ratings, *F*s≤ 1.44, *p*s≥ .24.

Mean identification was 3.57 (*SD*= 1.36). A 2 x 2 between participants indicated that perpetrator group membership did not significantly affect ingroup identification, *F*< 1. Participants numerically identified higher when ingroup victims were affected (*M*= 2.76, *SD*= 1.35) compared to outgroup victims (*M*= 3.89, *SD*= 1.34). This difference in identification was not significant, *F*(1,64)= 1.40, *p*= .24, η^2= .02.

Reactions at unfairness. The unfair treatment elicited more anger (*M*= 2.71, *SD*= 1.64) than the fair treatment (*M*= 1.17, *SD*= .61), *F*(1,66)= 60.57, *p*< .01, η^2= .46. Participants were also more willing to punish perpetrators who treated victims unfairly (*M*= 2.64, *SD*= 1.83) than those who treated victims fairly (*M*= 1.14, *SD*= .69), *F*(1,66)= 78,36, *p*< .01, η^2= .38. Planned contrasts showed outrage at unfairness independently from common group membership with either perpetrator or victim. In the outgroup/outgroup interaction participants reported more anger about and more willingness to punish unfair treatment than fair treatment, *p*= .02, *d*= 1.21; *p*= .01.[27] None of the participants was willing to punish outgroup perpetrators who applied fair treatment (see Table 9).

Reactions to victim and perpetrator group membership. Participants reported more anger when the victim was an ingroup member (*M*= 2.20, *SD*= 1.03) compared to an outgroup member (*M*= 1.69, *SD*= .71), *F*(1,66)= 5.58, *p*= .02, η^2= .08. Specifically, participants reported more anger when ingroup victims were treated unfairly (*M*= 3.17, *SD*= 1.81) compared to fairly (*M*= 2.26, *SD*= 1.33). This showed by an interaction effect of victim group membership and treatment, *F*(1,66)= 4.07, *p*= .05, η^2= .03. There was no interaction effect of the factors on anger, *F*< 1. The perpetrator group membership did not affect anger (ingroup: *M*= 2.08, *SD*= .98; outgroup: *M*= 1.79, *SD*= .82), *F*(1,66)= 1.69, *p*= .10, η^2= .02, and no interaction effect of perpetrator group membership and treatment on anger (see Table 9), *F*< 1. In contrast to anger, neither perpetrator nor victim group membership influenced participant's willingness to punish the perpetrator, *F*s≤ 1.82, *p*s≥ .18.

In line with the second hypothesis, ingroup victims elicited more anger when treated unfairly. In opposition to the third hypothesis, participants were not willing to punish perpetrators more harshly when they affected ingroup compared to outgroup victims.

[27] The planned contrast determined a significant difference in punishment of unfair versus fair treatment in outgroup/outgroup interactions, *p*= .01. However, the effect size could not be determined, because correlation coefficients were not definable (see Table 9: participants did not punish fair outgroup perpetrators).

Table 9. Descriptive statistics of control and dependent variables for each condition, Study 2 (*N*= 70), Section 4

		Perpetrator			
		Ingroup		Outgroup	
		Victim		Victim	
		Ingroup	Outgroup	Ingroup	Outgroup
	Treatment	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Ingroup Identification		3.87 (1.12)	3.28 (1.34)	3.65 (1.65)	3.48 (1.17)
Fairness Ratings	Unfair	2.39 (1.04)	2.44 (1.29)	2.44 (1.03)	2.94 (1.30)
	Fair	6.67 (.97)	6.89 (.32)	6.87 (.50)	6.72 (.57)
Anger	Unfair	3.40 (1.83)	2.42 (1.37)	2.92 (1.82)	2.11 (1.31)
	Fair	1.29 (.90)	1.20 (.70)	1.15 (.38)	1.04 (.28)
Willingness to punish	Unfair	3.39 (1.06)	3.11 (1.53)	3.44 (1.90)	3.61 (1.72)
	Fair	1.22 (.94)	1.33 (.97)	1.00	1.00

Note: M=*mean,* SD= *standard deviation*; *all variables varied from* 1 = *"not at all"* to 7 = *"very much"*

The role of ingroup identification. It was hypothesized that ingroup identification promotes the tendency to protect fellow ingroup members. In Study 1b, the covariate ingroup identification indicated that identification enhances anger when the ingroup is affected. If identification triggered the effects of victim group membership on anger in Study 2, the effect would vanish when the influence of ingroup identification is controlled for. We entered ingroup identification as continuous covariate to partial out its effects on anger from the effects of perpetrator and victim group membership. Overall, participants reported significantly more anger with increasing identification, $F(1,60)= 5.18$, $p= .03$, $\eta^2= .08$. Unfair treatment still tended to elicit more anger than fair treatment, but the main effect did not reach significance, $F(1,60)= 3.74$, $p= .06$, $\eta^2= .05$. As hypothesized, neither the main effect of victim group membership nor the interaction effect of treatment and victim group membership emerged when ingroup identification was controlled for. The ANCOVA showed no effects of victim or perpetrator group membership on anger, $F\text{s} \leq 2.60$, $p\text{s} \geq .11$.

In sum, participants reported more anger and punishment in the unfair treatment compared to the fair treatment condition, even in outgroup/outgroup interactions. There were no effects of perpetrator group membership. However, the victim's group membership substantially enhanced anger at unfairness independently from perpetrator group membership. Participants' willingness to punish increased for unfair treatment compared to fair treatment, but was not modified by group memberships. An ANCOVA controlling for ingroup identification showed that the enhanced anger on behalf of ingroup victims was dependent ingroup identification.

4.4.3 Discussion

Studies 1a, 1b, and Study 2 investigated reactions to fairness violations in a dictator game with minimal group memberships. It has been suggested that anger and subsequent altruistic punishment of immoral behavior evolves without self-involvement (*moral outrage*; e.g., Haidt, 2003). Other researchers argued that anger at moral violations only emerges on behalf of ingroup victims (*group-bound anger*; e.g., Batson et al., 2009), or in response to ingroup perpetrators (*black-sheep anger*; e.g., Abrams et al., 2000). The present studies orthogonally manipulated perpetrator and victim group membership as between-participant conditions (Study 1a and 2) and as within-participants condition (Study 1b). Additionally, participants indicated their reactions to each fair and unfair treatment.

Results show harsher altruistic punishment (Study 1a and b) and anger (Study 2) towards unfair compared to fair perpetrators across all three studies, even for outgroup/outgroup interactions. These findings are inconsistent with the notion that negative reactions to unfairness are exclusively elicited by self-involvement (Batson et al., 2009). Interactions between outgroup strangers explicitly do not affect oneself. Unfairness between outgroup members neither affected the ingroup nor the group norms of participants. The results support to the notion that moral violations (i.e., unfairness) elicit moral outrage (Haidt, 2003; Montada & Schneider, 1989), independently of offenders or victims.

Study 1b and Study 2 demonstrate enhanced negative reactions in respect to ingroup victims, especially when treated unfairly. This was triggered by identification with the ingroup. Previous studies have shown that people altruistically punish moral violations on behalf of fellow group members but not on behalf of outgroup members in longstanding natural groups (Bernhard, Fischbacher, et al., 2006; Goette et al., 2006). Ingroup identification represents personal attachment to the ingroup and its members (Brewer, 2007). In minimal groups, ingroup identification indicates salience and importance of the novel groups. As ingroup identification elicits group-based emotions (e.g., Yzerbyt et al., 2003), highly identified group members showed stronger altruistic punishment (Study 1b) and more anger (Study 2) towards ingroup offenders compared to outgroup offenders.

In contrast to the suggestion of *black-sheep anger*, altruistic punishment of outgroup perpetrators exceeded that of ingroup perpetrators in Study 1b (and a tendency in Study 1a). Anger did not differ in respect to perpetrator group membership in Study 2. This is in line with other, who found that infliction on punishment is more leniently for ingroup than outgroup members (Bernhard, Fischbacher, et al., 2006; Lieberman & Linke, 2007; Schiller et al., 2014). Especially competitive intergroup relations foster punishment of outgroup perpetrators (Goette et al., 2012). Participants therefore might minimize outgroup perpetrators' outcome, so that the ingroup obtains an overall higher (monetary) outcome. Conversely, other research showed that ingroup perpetrators are punished harsher than outgroup perpetrators to maintain ingroup positivity (Abrams et al., 2000; van Prooijen, 2006) or cooperative tendencies within groups (Mendoza et al., 2014; Shinada et al., 2004).

The presented studies showed that unfair sharing, and shared group membership with the victim increase negative reactions to distributors in dictator games with minimal group contexts. In order to test whether the same effects emerge in natural groups with reaction to moral deviance, Study 3 and 4 were conducted. The subsequent studies extend the previous examinations, as contexts particularly about intergroup relations (cooperation/competition) was provided. At the time of data collection ingroup and outgroup were positively interdependent (cooperative relation) in Study 3, and negatively interdependent (competitive/ conflictive relation) in Study 4. We chose groups whose relation was well-known, for example because of media reports.

4.5 Study 3: Reactions to treatment by the police in France and Germany

The main aim of Study 3 was to investigate effects of unfairness and varying ingroup and outgroup perpetrators in a natural and cooperative intergroup context. We applied the same design as in Study 1a, 1b, and 2. Participants' ingroup were "Germans" and their outgroup "French". All conditions were implemented between participants, implying that participants were confronted with the intergroup relation only in the ingroup/outgroup and outgroup/ingroup conditions. Thus, the observed reactions to each condition were independent from each other.

In a scenario German or French police officers treated German or French tourists fairly or unfairly during a passport control. A cooperative relation is defined by positive interdependence, such as sharing resources (e.g., economic standing, social system…) and culture (e.g., values, traditions…). Germany and France have a strong geographical, economic, cultural and historical connection. Germany and France are related though common economic and political efforts in the European Union. They share cultural experiences, for example through common cultural institutions (e.g., ARTE) and town twinning projects. The friendly relationship was emphasized through the example of tourism, which highlights economic and cultural exchange between the two groups. As in the previous studies, we expected that *moral outrage* emerges in unfair treatment conditions. A moral violation between groups threatens positive intergroup relations (Okimoto & Wenzel, 2010). Futher, anger on behalf of ingroup victims (*group-bound anger*) may emerge in this study as well, especially for participants who highly identify with "the Germans". Otherwise, negative reactions provide the possibility of stabilizing cooperative relations through signaling and promoting mutual fair treatment. No differences in anger and punishment tendencies may emerge between conditions, or even in stronger negative reactions to unfairness in intergroup encounters. Additionally, we measured other emotions (content, irritation, and disinterest) to explore whether they are affected differently than anger.

4.5.1 Method

Participants. We collected data of 216 participants to obtain high statistical power for detecting small effects. Data collection took place in the Foyer at the University of Jena in November and December 2014 (*N* = 108) and February 2015 (*N* = 108). Eleven participants took part twice, and were therefore eliminated from the sample. Three more were removed, due to language difficulties. The remaining sample consisted of 202 students from various disciplines (106 female; 2 without indication of gender; M_{age}= 22.85; *SD*= 4.97).

Design. We applied 2 (treatment: fair/ unfair) x 2 (perpetrator group membership: ingroup/ outgroup) x 2 (victim group membership: ingroup/ outgroup) between-participants analyses of variance to test the hypotheses. The sample size of 202 participants enabled the detection of main effects of η^2= .06 and two-way interaction effects of η^2= .07 with a probability of α= .05 and a power of 1-ß= .95.

Procedure. Participants first signed written informed consent. Then they were handed one of eight scenarios. The descriptions of the scenarios were simple to avoid confounds and to give only essential information: "An occurrence in a German/French city (anonymous): A German /French tourist group during a city trip in Germany/France was stopped in the streets and controlled for passports by the police, because of confusions." The manipulation of treatment ended in: "After the tourists identified themselves, the police officers apologized for the hassles and wished them a pleasant trip (fair treatment)." or "After the tourists identified themselves, they were taken to the police station for a thorough interrogation. Without obvious reason they had to wait for several hours in a locked room, until the police officers allowed them to leave the station (unfair treatment)."

After reading the scenario, participants indicated to what extent they experienced seven anger-related and other emotions that were presented as adjectives (see Study 2). We explored relations between the emotional adjectives with an explorative factor analysis (see Appendix H for factor loadings). On this basis we created the indices "anger" (α= .93), "content" (satisfied, calm, α= .70), "irritation" (confused, surprised, puzzled; α= .73), and "disinterest" (bored, uninvolved, α= .62). Afterwards, participants indicated their willingness to punish the police officers ("In your opinion, should the officers be punished?"), and rated fairness and morality of the police officers' behavior ("Is the officers' behavior fair"; "… immoral?"). Lastly, identification with Germans was measured by nine items (α= .75; "I see myself belonging to the Germans"; "I identify with the Germans"; "My personal fortune is more important for me than the fate of the Germans" (-); "I often regret to be German" (-) etc.). All dependent variables were measured on a seven-point scale (1 = *not at all*; 7 = *very much*). Participants were thanked, debriefed and received a chocolate bar as incentive.

4.5.2 Results

Preliminary analyses. Mean identification with the German ingroup was 5.01, *SD*= .92. Ingroup identification did not differ between experimental conditions, although a trend indicated that it was lower in the unfair treatment conditions (*M*= 5.13, *SD*= .88) than in the fair treatment conditions (*M*= 4.89, *SD*= .95), *F*(1,194)= 3.38, *p*= .07, η^2= .02. There were no further effects on ingroup identification, *F*s≤ 1.47, *p*s≥ .23.

Unfair treatment was rated as more unfair (*M*= 4.56, *SD*= 1.76) than the fair treatment (*M*= 1.90, *SD*= 1.44), *F*(1,194)= 139.90, *p*< .01, η^2= .41. Unexpectedly, fairness ratings were harsher when outgroup victims (*M*= 3.55, *SD*= 2.09) were affected compared to ingroup victims (*M*= 2.97, *SD*= 2.06), *F*(1,194)= 5.38, *p*= .02, η^2= .02. There were no other effects on fairness ratings, *F*s≤ 2.47, *p*s≥ .12. Participants rated the unfair treatment to be less moral (*M*= 2.87, *SD*= 1.45) than the fair treatment (*M*= 5.03, *SD*= 1.87), *F*(1,194)= 83.95, *p*< .01, η^2= .30. There were no further effects on morality ratings, *F*s≤ 2.45, *p*s≥ .12. As intended, behavior of the police was perceived as more unfair and morally wrong when they took the tourist group to the office, instead of releasing them with apology after the control.[28] All descriptive statistics are displayed in Table 10.

The study addressed whether the anger people experience when observing moral violations is dependent on the interaction partners' group memberships. We additionally tested when content (reversed), irritation, and disinterest are dependent on the group memberships to differentiate them from anger. Overall, there was a significant effect of treatment across emotions, V= .43, F(4,190)= 36.28, p< .01, and a significant interaction of treatment and perpetrator on emotions, V= .06, F(4,190)= 2.81, p= .03. There were no other effects across emotions, V= .04, F(4,190)≤ 1.80, p≥ .13. Emotional reactions were heterogeneous across the manipulations and thus univariate analyses were conducted.

Reactions to Unfairness. In line with the hypothesis, participants reported more anger when the police officers treated the tourists unfairly (M= 3.87, SD= 1.26) compared to fairly (M= 2.06, SD= 1.05), F(1,193)= 120.96, p< .01, η^2= .38. They also reported significantly less content with unfair treatment (M= 1.74, SD= 1.11) than fair treatment (M= 2.89, SD= 1.31), F(1,193)= 46.90, p< .01, η^2= .19. Irritation and disinterest did not differ in terms of treatment (irritation: unfair: M= 3.69, SD= 1.47; fair: M= 3.35, SD= 1.46; disinterest: unfair: M= 2.90, SD= 1.62; fair: M= 3.26, SD= 1.59), F(1,193)= 2.93, p= .09, η^2= .01; F(1,190)= 2.40, p= .12, η^2= .01. Participants were more willing to punish the police officers when they treated the tourists unfairly (M= 3.23, SD= 1.73) instead of fairly (M= 1.44, SD= .82), F(1,193)= 156.99, p< .01, η^2= .30 (see Table 10).

Reactions to victim and perpetrator group membership. There was no effect of victim group membership on anger, or any other emotion, Fs≤ 1.36, ps≥ .25. There were no effects of perpetrator group membership on anger (and disinterest), Fs≤ 2.61, ps≥ .11. Compared to the French police officers, participants were less content, when German police officers stopped the tourist group for a short time (fair ingroup perpetrator: M= 2.64, SD=

[28] During an interruption of data collection due to semester break an islamistic terrorist attack on the Parisian satirical magazine "Charlie Hebdo" occurred. This emphasized the commonalities between Germany and France, as well as their interdependence in defending their values of freedom of speech against outside threats (e.g., http://www.zeit.de/politik/deutschland/2015-01/mahnwache-berlin-charlie-hebdo-weltoffenheit-menschenrechte). Therefore, we tested whether the time of data collection actually changed the fairness and morality ratings towards in- and outgroup perpetrators and victims. There was only one significant interaction of measurement time on fairness ratings. The effect indicated that, after Charlie Hebdo, participants rated fair treatment as less fair and unfair treatment as fairer than before, F(1,186)= 4.67, p= .03, η^2= .01. Morality ratings did not differ before and after Charlie Hebdo, F(1,186)= .44, p= .51, η^2< .001. Since the effect did not affect the overall main effect of treatment on fairness perception, and did not have other influences, it was not considered in further data analyses.

1.21; fair outgroup perpetrator: *M*= 3.12, *SD*= 1.71). An interaction effect showed less differences in content when officers kept the tourist group locked in (unfair ingroup perpetrator: *M*= 1.83, *SD*= 1.37; unfair outgroup perpetrator: *M*= 1.71, *SD*= .83), *F*(1,193)= 4.33, *p*= .04, η^2= .02. Participants were more irritated by an outgroup perpetrator (*M*= 3.75, *SD*= 1.48) than an ingroup perpetrator (*M*= 3.30, *SD*= 1.43), *F*(1,193)= 5.11, *p*= .03, η^2= .02. Controlling for ingroup identification did not change any effects on anger.

Participants reported numerically more punishment tendencies, when outgroup victims (*M*= 2.52, *SD*= 1.72) were affected compared to ingroup victims (*M*= 2.18, *SD*= 1.52). The difference did not reach significance, *F*(1,193)= 4.76, *p*= .07, η^2= .01. An interaction effect indicated that participants were more willing to punish perpetrators in intergroup encounters, *F*(1,193)= 5.94, *p*= .02, η^2= .02. This was pronounced when police officers treated the tourists unfairly, *F*(1,193)= 5.28, *p*= .01, η^2= .02.

Table 10. Means and standard deviations of control and dependent variables for each condition, Study 3 (*N*= 202), Section 4

		Perpetrator			
		Ingroup		Outgroup	
		Victim		Victim	
		Ingroup	Outgroup	Ingroup	Outgroup
	Treatment	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Ingroup Identification	Unfair	4.78 (1.08)	5.10 (.79)	4.88 (1.01)	4.81 (.93)
	Fair	5.19 (1.04)	5.00 (.86)	5.13 (.78)	5.19 (.93)
Fairness	Unfair	4.44 (1.81)	5.12 (1.45)	4.27 (2.03)	4.35 (1.86)
	Fair	1.62 (.75)	1.91 (1.24)	1.56 (1.26)	2.04 (1.76)
Moral Wrongness	Unfair	3.18 (1.61)	2.75 (1.17)	3.42 (1.53)	3.35 (1.35)
	Fair	5.48 (1.14)	5.59 (1.45)	5.94 (1.29)	5.26 (1.66)
Anger	Unfair	3.57 (1.33)	3.99 (1.14)	4.05 (1.34)	3.88 (1.23)
	Fair	2.12 (.90)	2.09 (1.21)	1.81 (.79)	2.24 (1.24)
Content	Unfair	2.00 (1.45)	1.67 (1.30)	1.60 (.71)	1.71 (.83)
	Fair	2.56 (1.68)	2.74 (1.16)	3.12 (1.42)	3.17 (1.36)
Irritation	Unfair	3.57 (1.51)	3.64 (1.18)	3.67 (1.44)	3.88 (1.74)
	Fair	3.26 (1.57)	2.67 (1.27)	3.63 (1.42)	3.80 (1.37)
Disinterest	Unfair	2.86 (1.64)	2.54 (1.14)	2.92 (1.88)	3.29 (1.74)
	Fair	3.58 (1.66)	3.22 (1.49)	3.48 (1.87)	2.74 (1.23)
Willingness to punish	Unfair	2.75 (1.29)	4.15 (1.69)	3.19 (1.91)	2.77 (1.66)
	Fair	1.50 (.58)	1.61 (.94)	1.28 (1.02)	1.40 (.71)

Note: M=*mean,* SD= *standard deviation*; *all variables varied from* 1 = *"not at all to"* 7 = *"very much"*

4.5.3 Discussion

Study 3 was designed to test the emergence of outrage and punishment in natural cooperative intergroup contexts. In a scenario German or French police stopped German or French tourists on the street. Similarly to the results in Study 1a, 1b and 2, anger about moral violations emerged independently of shared group memberships with victim or perpetrator. Releasing them after passport control (fair treatment) elicited less anger and punishment than holding the tourists back for several hours without explanation (unfair treatment). A shared group membership with officers and tourists did not have any significant effect on anger about moral violations, even though the statistical power of the study was high. Furthermore, there were no effects of ingroup identification on the relevance of group membership.

The same amount of anger was reported in unfair outgroup/outgroup and ingroup/ingroup interactions. Effects of victim and perpetrator group membership showed for minimal groups (Study 1b, Study 2), but not in the present cooperative intergroup context. One could argue that observers were involved in both groups because of their cooperative relationship. Nevertheless, the results support the idea that morality, and reactions to moral violations, serve to maintain cooperation and positive relationships (see Section 1.2.1; e.g., Haidt & Kesebir, 2010; Rai & Fiske, 2011). They moreover suggest that the psychological underpinnings of morality operate not only on an interpersonal, but also on an intergroup level.

Participants reported less satisfaction and punishment with unfair intergroup encounters. Intergroup interactions, in contrast to within-group interactions, are not necessarily perceived cooperative (e.g., Balliet et al., 2014; Kramer & Brewer, 1984). Restoring positive intergroup relations after an intergroup offense might thus be especially important. In particular, participants' willingness to punish unfair ingroup perpetrators in particular indicates the "expressive functions of punishment" (J. Feinberg, 1965): punishment can be used as a mean to demonstrate that the group member distances herself from the perpetrator, and discharges the victim from potential blame.

It is important to bear in mind the possible bias in the observed responses. Fairness ratings were particularly harsh when ingroup perpetrators affected outgroup victims. Judgments about moral violations are influenced by deliberate reasoning, as well as spontaneous affective responses (Greene & Haidt, 2002). The affective response to a moral violation did not show as anger, however, other emotions could elicit strong negativity that influences unfairness perceptions. Next to anger, shame is an emotional response to ingroup moral failure that also triggers punishment tendencies to restore the group's moral image (Chekroun & Nugier, 2011; Gausel, Leach, Vignoles, & Brown, 2012; Iyer et al., 2007; Shepherd, Spears, & Manstead, 2013).

Moreover, it cannot be ruled out that the group context were not salient, especially in the ingroup/ingroup and outgroup/outgroup condition. Group salience is an important feature to elicit group-based emotions (Kuppens & Yzerbyt, 2012). The inclusive self-categorization (e.g., European) might have been more salient than the national groups because of their cooperative group relation and the tourist context (S. L. Gaertner et al., 1990).

4.6 Study 4: Reactions to torture by Secret Service in the "Western Society" and the "Islamic State"

As expected, Study 3 demonstrated that, in cooperative intergroup contexts, anger at moral violations emerges independently of group memberships. It remains unclear, whether this is due to the general nature of moral concern or self-involvement with the outgroup. Moral violations should motivate moral outrage even despite of group-, empathic, or personal consequences (Batson et al., 2009; O'Mara et al., 2011). Thus, moral outrage is a response to the deed itself (see *Research Line II*).

Therefore, Study 4 investigated reactions to moral violations within the ingroup, in hostile intergroup conflicts, and within the competing outgroup. Ingroup biases and outgroup derogation increase when the ingroup competes with the outgroup (see Section 1.2.3). Sometimes outgroups are even dehumanized after being offended by the ingroup (Čehajić, Brown, & González, 2009). Consequently, they are not worth of solidarity and empathy (for an overview, see N. Haslam & Loughnan, 2014). Group members even experience pleasure when confronted with a negative event that affects the outgroup, and simultaneously heightens ingroup status (Leach, Spears, Branscombe, & Doosje, 2003). In intergroup competitions harm inflicted on the outgroup may even help the ingroup to gain power over the outgroup. This setting provides a strong test for moral outrage, as the concept predicts anger even at moral violations within a despised outgroup. Moral outrage has been suggested to emerge most obviously in cases of fundamental and intuitive moral violation (Batson et al., 2009; Haidt, 2001). Thus, we described a case of torture (vs. a case of a fair trial).

We set the stage in a protracted and violent conflict, which was very present by the time the study was conducted: Western states (West) against the so called "Islamic State" ("IS"; time of data collection: June 2015).[29] Western media frequently reported of torture and public killings of Western prisoners by the "IS". Additionally, some reports of captures of "IS"-fighters appeared, although less frequently. The "IS" was famous for treating their own deviants very harshly. The conflict was additionally shaped by media reports of cruelty and inhumanity of "IS"-members. According to Western perceptions, the two parties fight for resources (i.e., Syrian territory) and possess competing value systems.

[29] For illustration: A Google-search in June 2016 on "Islamic State June 2015" would indicate over 6 Mio. hits; the same search on German still provided 109.000 hits; even though the state is not recognized (sometimes the group is referred to as Al Nusra), we'll keep at the common term here

If *moral outrage* is a universal reaction to moral violations, it should evolve in offenses within the outgroup. Moreover, the differential hypothesis that empathy with victims accounts for anger and punishment (Batson et al., 2007) was tested. Empathy (i.e., distress on behalf of another person's suffering) has been found to be increased for ingroup members, but still accounts for reactions to outgroup suffering (Dovidio et al., 2004; Stürmer, Snyder, & Omoto, 2005; Tarrant, Dazeley, & Cottom, 2009).

It is further hypothesized that torture of an ingroup member elicits *group-based anger*. Additionally to anger on behalf of fellow group members, the threat towards the Western society might increase anger about ingroup victims (Batson et al., 2009; Okimoto & Wenzel, 2010). Moreover, violations of general norms may elicit harsh anger towards ingroup perpetrators (*black-sheep anger*), because they threaten the ingroups' moral superiority over the outgroup (Marques et al., 2001). Outgroup perpetrators should elicit less anger because they accentuate intergroup differences in moral virtue. Additionally, other emotional reactions (i.e., content (-), fear, shame) were measured and distinguished from anger.

4.6.1 Method

Sample. 104 students of the Polytechnical University in Jena took part in the study. We excluded one participant because he gave the same answer to every question, and one participant because she already participated in Study 3. The remaining sample consisted in 102 students from different study fields (48 female; M_{age}= 24.00; *SD*= 3.49). Participants mean religiousness was 2.64, *SD*= 1.94 (46.1% Christians, 52.0% no affiliation). Mean political interest was 4.48, *SD*= 1.49, and mean political orientation was 3.36, *SD*= 1.13.

Design. We applied 2 (treatment: fair/ unfair) x 2 (perpetrator group membership: ingroup/ outgroup) x 2 (victim group membership: ingroup/ outgroup) between-participants analyses of variance in order to test the hypotheses. The sample of 102 participants allowed determining main effects of η^2= .07 and interaction effects of η^2= .10 with α= .05 and 1-β= .80.

Procedure. After signing written informed consent, participants read one of eight scenarios. Instructions stated that the study inspected newspaper articles. It was claimed that the main research question would address emotional reactions to the reports. Participants then read about the treatment of a prisoner who was either European (ingroup victim) or from the so called "Islamic State" ("IS"; outgroup victim). The area in which the scenario took place and the Secret Service who dealt with the prisoner was either European (ingroup perpetrator) or from the “IS” (outgroup perpetrator). Half of the participants read that the prisoner was receiving a court trial watched by a human rights organization (trial). The other half read that the army member died while being imprisoned because of brutal interrogating methods (torture; see Appendix I for wording).

The main dependent variable was anger. Further, we included items describing fear, shame and content for differentiating outrage from other negative emotions. We combined emotional adjectives in indices (“anger”: seven items, α= .94; “fear”: two items: α= .66; “content”: three items, α= .70; “shame” was treated as a single item dependent variable; see Appendix J for correlations of emotions). Participants then rated how much they were willing to punish the secret service and to what extent they found the treatment immoral. A variety of control variables was measured, in order to explore: how strongly participants perceived the “IS” as a threat to Europe and European values; how severe they perceived the conflict between Western states and the “IS”; and whether they feel empathic towards the prisoner (“Mitgefühl”).[30] Then, they indicated their degree of identification with Western society and values (α= .74). Finally, they reported how plausible the article was. All dependent variables were measured on a seven-point scale (1 = *not at all*; 7 = *very much*). Participants were thanked, debriefed and received a chocolate bar as incentive.

4.6.2 Results

Preliminary analyses. Mean identification with Western society and its values was 4.67 (*SD*= .96). Participants identified with Western society stronger when the tortured victim was an ingroup member (*M*= 5.04, *SD*= .64) compared to an outgroup member (*M*= 4.37, *SD*= 1.15). This effect of victim group membership was not observed in case of a fair trial (ingroup victim: *M*= 4.54, *SD*= .90; outgroup victim: *M*= 4.75, *SD*= .98). This expressed in a significant interaction effect of treatment and victim group membership, *F*(1,93)= 5.36, *p*= .02, η^2= .05. There were no other effects on ingroup identification, *F*s≤ 2.22, *p*s≥ .14.

The perceived threat to Europe and European values did not differ between conditions (overall: *M*= 5.02, *SD*= 1.81), *F*s≤ 1.66, *p*s≥ .20. Likewise, the conflict between Western States and the “IS” was perceived as equally severe across conditions (overall: *M*= 5.78, *SD*= 1.34), *F*s≤ 3.94, *p*s≥ .09.

[30] We additionally measured victim deservedness and victim blame. They will not further be reported here, as they were not of primary interest.

Participants rated torture as significantly more immoral (*M*= 6.59, *SD*= .80) than the fair trial (*M*= 4.52, *SD*= 1.59), *F*(1,94)= 76.46, *p*< .01, η^2= .41. Ingroup and outgroup perpetrators' actions were judged equally immoral in the torture conditions (ingroup: *M*=6.84, *SD*= .37; outgroup: *M*= 6.35, *SD*= 1.02). An interaction effect of perpetrator and treatment indicated that the "IS" holding a fair trial for a Western prisoner was perceived as less moral than other trials (see Table 11), *F*(1,92)= 9.70, *p*< .01, η^2= .05.[31]

In line with the predominant opinion in the current Western media, participants perceived "IS" perpetrators in the fair trial and Western perpetrators in the torture condition as less plausible than reversed (see Table 11). This was expressed in a significant interaction effect of treatment and perpetrator group membership, *F*(1,93)= 8.28, *p*< .01, η^2= .08. All other conditions were perceived as similarly plausible, *F*s≤ 1.14, *p*s≥ .29. Plausibility was added as a covariate in the main analyses to control for its buffer effects on emotional experience (Frijda, 1988). Table 11 displays descriptive statistics of all dependent variables per conditions.

A MANCOVA[32] (covariate plausibility) combined all emotional reactions towards torture compared to a court trial. There was a significant multivariate effect of treatment on the combination of anger, fear, content (reversed), and shame, *V*= .47, *F*(4,91)= 20.46, *p*< .01. There was an interaction of treatment and perpetrator group membership in the multivariate analysis, *V*= .11, *F*(4,91)= 2.74, *p*= .03. The analysis did not reveal other effects on the emotions, *V*s≤ .05, *F*s≤ 1.09, *p*s≥ .39.

Reactions to Unfairness. The univariate analyses revealed that anger was significantly higher in the torture (*M*= 4.57, *SD*= 1.62) than in the trial condition (*M*= 2.44, *SD*= 1.28), *F*(1,92)= 66.10, *p*< .01, η^2= .39. Likewise there were significant, but smaller, main effects of treatment on shame (torture: *M*= 3.00, *SD*= 1.96; trial : *M*= 1.92, *SD*= 1.35), *F*(1,92)= 11.98, *p*< .01, η^2= .10, fear (torture: *M*= 2.89, *SD*= 1.41; trial : *M*= 1.95, *SD*= 1.01), *F*(1,92)= 13.36, *p*< .01, η^2= .12, and content (torture: *M*= 1.30, *SD*= .62; trial : *M*= 1.95, *SD*= 1.17), *F*(1,92)= 12.41, *p*< .01, η^2= .11. As in the prior studies, the pairwise comparison in the "IS" perpetrator and "IS" victim conditions shows that participants reported more anger about torture than about the fair trial (see Table 11). The difference did not reach significance, *p*= .12, *d*= .57. Shame did not differ if outgroup perpetrators interrogated an outgroup victim for a fair trial or tortured him, *p*= .81, *d*= .02. Fear and discontent were

[31] The effects remained after controlling for plausibility.

[32] Effects did not change in comparison no covariate (see Appendix K for details).

slightly higher in the "IS"/"IS" torture condition than in the "IS"/"IS" trial condition, *p*= .38, *d*= .26; *p*= .13, *d*= .05.

As expected, the main effect of treatment on punishment was also significant, *F*(1,90)= 58.04, *p*< .01, η^2= .36. Participants were more willing to punish perpetrators in the torture condition (*M*= 4.82, *SD*= 2.01) than in the trial condition (*M*= 2.20, *SD*= 1.25).

Reactions to victim and perpetrator group membership. We further looked at how the different emotions were affected by perpetrator and victim group membership. Participants reported more anger, when ingroup perpetrators applied torture (*M*= .81, *SD*= 1.21) compared to a fair trial (*M*= 2.02, *SD*= .99). This difference was smaller for outgroup perpetrators (torture: *M*= 4.33, *SD*= 1.62; trial: *M*= 2.86, *SD*= 1.28). The two-way interaction was significant, *F*(1,92)= 7.47, *p*< .01, η^2= .04. There were no other effects on anger, all *Fs*≤ 1.52, *ps*≥ .22. Shame was significantly higher, when ingroup perpetrators applied torture (*M*= 3.56, *SD*= 2.14) compared to a trial (*M*= 1.48, *SD*= .87), but did not differ for outgroup perpetrators (torture: *M*= 2.46, *SD*= 1.63; trial: *M*= 2.36, *SD*= 1.60), *F*(1,92)= 6.07, *p*= .02, η^2= .05. There were no other effects on shame, *F*s≤ .79, *p*s≥ .38. Fear and (dis)content did not differ for perpetrator and victim group membership, *F*s≤ .57, *p*s≥ .45; *F*s≤ 1.67, *p*s≥ .20.

Table 11. Means and standard deviations of control and dependent variables for each condition, Study 4 (*N*= 102), Section 4

		Perpetrator			
		Ingroup		Outgroup	
		Victim		Victim	
		Ingroup	Outgroup	Ingroup	Outgroup
	Treatment	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Ingroup Identification	Torture	4.95 (.48)	4.47 (.98)	5.11 (.76)	4.27 (1.34)
	Trial	4.82 (.94)	4.65 (.82)	4.26 (.90)	4.86 (1.17)
Moral Wrongness	Torture	6.83 (.39)	6.85 (.38)	6.23 (1.17)	6.46 (.88)
	Trial	4.08 (1.55)	4.15 (1.86)	5.62 (.96)	4.33 (1.50)
Plausibility	Torture	4.50 (1.88)	4.15 (2.04)	5.62 (1.12)	5.00 (1.73)
	Trial	5.00 (1.48)	5.23 (1.54)	4.00 (2.24)	4.25 (1.22)
Anger	Torture	4.80 (1.22)	4.81 (1.26)	4.80 (1.34)	3.86 (1.78)
	Trial	2.10 (.90)	1.96 (1.11)	2.76 (1.27)	2.96 (1.33)
Shame	Torture	3.92 (2.15)	3.23 (2.17)	2.38 (1.56)	2.54 (1.76)
	Trial	1.42 (.79)	1.54 (.97)	2.23 (1.64)	2.50 (1.62)
Fear	Torture	2.79 (1.41)	2.96 (1.59)	3.00 (1.21)	2.81 (1.58)
	Trial	2.00 (.88)	1.69 (.99)	1.85 (.94)	2.29 (1.23)
Content	Torture	1.25 (.38)	1.31 (.48)	1.21 (.40)	1.44 (1.02)
	Trial	2.25 (1.06)	2.15 (1.55)	1.41 (.75)	2.00 (1.13)
Willingness to punish	Torture	5.67 (1.57)	5.23 (1.88)	4.54 (2.07)	3.83 (2.21)
	Trial	2.00 (1.35)	1.67 (1.16)	2.92 (1.66)	2.17 (1.34)
Empathy with victims	Torture	5.08 (1.88)	5.00 (1.58)	5.15 (1.46)	4.31 (1.97)
	Trial	4.00 (1.73)	2.15 (1.14)	5.00 (1.41)	3.67 (1.78)

Note: M=*mean,* SD= *standard deviation*; *all variables varied from* 1 = *“not at all” to* 7 = *“very much”*

In line with anger, there was an interaction effect of perpetrator group membership and treatment on punishment, *F*(1,90)= 8.48, *p*= .005, η^2= .05. Participants indicated more willingness to punish, when ingroup perpetrators applied torture (*M*= 5.44, *SD*= 1.71) compared to a trial (*M*= 1.83, *SD*= 1.24). This difference was less pronounced for outgroup perpetrators (torture: *M*= 4.20, *SD*= 2.12; trial: *M*=2.56, *SD*= 1.53). Willingness to punish tended to be stronger, when the victim was an ingroup member (*M*= 3.78, *SD*= 2.16) compared to an outgroup member (*M*= 3.27, *SD*= 2.19). The victim main effect did not reach significance, *F*(1,90)=2.58, *p*= .11, η^2= .02. There were no other effects on punishment, *F*s≤ .66, *p*s≥ .42. Thus, anger, shame and punishment tendencies were higher when the torturing Secret Service was Western instead of from the "IS". While anger also increased with torture by outgroup perpetrators, shame did not.

Identification and anger. As mentioned above, identification was higher when ingroup victims were tortured. We applied analyses of variance on anger that included identification as a covariate. All beforehand reported effects remained significant, indicating that they were independent of ingroup identification.

Empathy and moral outrage. Moral outrage has been suggested to actually emerge on behalf of the suffering victims (Batson et al., 2007). Across all torture conditions, serious harm was inflicted on the victims. Therefore, a meditational model via bootstrapping (Preacher & Hayes, 2004, 2008) tested whether anger about torture was mediated by empathy with the victims. The model overall explained 52% of the variance of anger, *F*(2,99)= 53.09, *p*< .01. Results show a significant effect of empathy on anger (b-path), β= .37, *t*= 4.96, *p*< .01. Treatment also significantly affects anger (c-path), β= .63, *t*= 8.12, *p*< .01. This effect remains significant when introducing the mediator empathy (c' path), β= .51, *t*= 6.98, *p*< .001. The indirect effect of torture on outrage via empathy was also significant. The confidence interval includes zero, β= .12, 95%CI= [.06; .22]. This indicates that torture indeed elicits anger. The relationship is partly mediated by empathy with the victims.

4.6.3 Discussion

Study 4 aimed at investigating whether reactions to torture vary regarding different group memberships of victim and perpetrator. In a hostile intergroup context (Western states and "IS"), the Secret Service would either torture a prisoner to death or provide him with a court trial. If people generally experience anger about moral violations, this should emerge even when the moral violation affects a member of a despised outgroup.

The results show that participants generally experienced more negative emotions (anger, shame, fear), and less satisfaction, about torture than at a fair trial. Torture versus trial explained most variance in anger. This was also the case when "IS" captured an "IS"-soldier: anger, fear, and dissatisfaction, and not shame, tended to be higher in the torture compared to the trial condition. However, the differences did not reach significance. Negative emotional reactions to deviance were suggested to support cooperation, as they motivate norm-enforcement (Fehr & Gächter, 2002). In contrast, our results indicate that even moral violations that affect a competitive outgroup elicit strong emotional reactions and punishment tendencies. As in the previous studies in this Section, moral violations elicited outrage independently of group membership, indicating that it is generally despised.

In contrast to previous studies (Batson et al., 2009; Gordijn et al., 2001), no *group-based anger* (i.e., anger on behalf of ingroup victims) emerged in the present study. This finding was unexpected, and suggests that victim group membership was less relevant than the perpetrator group membership in the presented intergroup conflict. When contrasting merely the conditions in which the perpetrator is an outgroup members (similarly to Batson et al., 2009), the present data suggests that torturing a Western soldier elicits more anger than torturing an "IS"-fighter. However, this is not the case when the "IS"-Secret Service provides the prisoner with a fair trial.

The current findings does not support the idea that moral outrage is actually empathic anger (Batson et al., 2007). Empathy with victims (the perception of the victim's suffering; "Mitgefühl") contributed to the strength of outrage about torture, but did not fully explain it. In the present study, all moral violations elicited suffering of victims that were either ingroup and outgroup members. Perceiving the suffering of others, even outgroup members, may elicit empathic reactions, especially as the scenario activates the Western norm of maintaining human rights and wellbeing. Similarly, Tarrant and colleagues have shown that ingroup norms prescribing empathy increase empathy for outgroup members (Tarrant et al., 2009).

Anger at ingroup perpetrators resembles the Black Sheep Effect (Marques et al., 2001; Marques & Paez, 1994). Torturing ingroup perpetrators elicited more anger than torturing outgroup perpetrators, whereas ingroup members providing a fair trial elicited less anger than respective outgroup members. The notion that "IS"-supporters often apply torture is a western stereotype of the "IS", whereas Westerners expect their ingroup to hold up the values of human rights. Thus, Western torturers undermine ingroup norms and decrease positive distinctiveness from the outgroup (e.g., Marques et al., 2001).

Shame was highest for ingroup perpetrators applying torture, and did not emerge on behalf of outgroup perpetrators. Group-based shame derives from feelings of inferiority of the ingroup. It motivates distancing the self from the ingroup, avoidance of the appraisal situation (Iyer et al., 2007; Johns, Schmader, & Lickel, 2005; Lickel, Schmader, Curtis, Scarnier, &

Ames, 2005), and sometimes pro-social tendencies (Gausel et al., 2012). Thus, even though shame may have triggered anger at fellow ingroup members (Tangney, Wagner, Fletcher, & Gramzow, 1992), it cannot account for reaction to outgroup perpetrators. Fear and content, in contrast to anger and shame, did not differ in regards to perpetrator group membership.

In accordance with anger, participants were more willing to punish perpetrators who applied torture in contrast to those who provided a fair trial. Moreover, they were more willing to punish torturing Westerners than torturing “IS”-members. In intergroup offenses punishment demonstrates power, whereas it re-establishes group standards in within-group encounters (Okimoto & Wenzel, 2010; Wenzel et al., 2008). The current findings suggest that the perceived threat to ingroup values deserved higher punishment, as indicated by the high punishment tendencies towards ingroup torturers. In contrast, when “IS” perpetrators threaten Western power willingness less willingness to punish was reported.

The results must be interpreted with caution. Critically, moral judgments were biased in favor of the ingroup. Even though, we explicitly stated that the trial conditions were fair, participants found it to be an acceptable treatment applied by Western Secret Service, but not by the “IS”. Additionally, it is not clear how representative the ingroup (and outgroup) perpetrators and victims were in the current study. Full members are derogated and punished more harshly by fellow group members. Relatively less punishment is inflicted on new or marginal members (Pinto et al., 2010). Future studies should consider pre-tested representative group members as perpetrators and victims. Moreover, we recommend further research with increased sample sizes. This would enable determining whether the observed difference in emotional reactions to torture and a fair trial produces a significant change in the “IS”/”IS” condition.

4.7 General Discussion

The aim of *Research Line III* was to shed light on reactions to moral violations that emerge from different appraisal situations that often co-occur: the moral transgression, ingroup victims, and ingroup perpetrators (e.g., Batson et al., 2009; McAuliffe & Dunham, 2016). Differences in reactions to the situations indicate specific reasons why third-parties react on moral violations: an unspecific general aversion towards violations of moral standards, the protection of fellow group member’s wellbeing, and the preservation of normative consensus within groups. The present studies showed that anger about and punishment of moral violations even emerge irrespective of self-involvement (even in outgroup interactions; *moral outrage*). In salient minimal group contexts, unfair distributions triggered punishment and anger on behalf of ingroup victims (*group-bound anger*). In a cooperative intergroup context differences in respect to shared group membership with victim or perpetrator did not show. Reactions to torture increased when the perpetrator was an ingroup member, but not on behalf of ingroup victims (*black-sheep anger*).

Moral outrage has been considered a typical emotional reaction to transgressions that do not directly touch the self (Haidt, 2003). It motivates punishment in order to re-establish the moral balance that has been disrespected (Darley & Pittman, 2003). As people mostly interact within their social groups (Brewer, 2007), such anger might be group-specific. This implies that moral violations touch the self via shared group membership with perpetrators or victims. Moral violations within the ingroup are punished harsher than between groups, as they threaten normative consensus (Abrams et al., 2000; Shinada et al., 2004), and simultaneously the wellbeing of fellow group members (Batson et al., 2009; Gordijn et al., 2001). Typically, either differential perpetrator or victim group memberships have been studied. The presented studies orthogonally manipulated fairness, victim, and perpetrator group membership to properly disentangle the causes of anger about and punishment of moral violations (see also Bernhard, Fischbacher, et al., 2006). Studies 1a, 1b, and 2 investigated altruistic punishment (Studies 1a and 1b) and anger (Study 2) at unfair and fair dictators in minimal groups. Two scenario studies used a natural group context. Unfair treatment was described as deprivation of personal liberty in two befriended nations (Germany/France, Study 3), and torture to death in a violent intergroup conflict (Western States/"Islamic State", Study 4). Study 3 and 4 assessed other emotional reactions (irritation, shame, content, fear) additional to anger and punishment.

4.7.1 Moral outrage: the primary reaction to moral violations

The present findings indicate that moral violations elicit negative emotions and punishment that go beyond group-boundaries; suggesting that third-party reactions to moral violations indeed are independent from personal interest and group processes (Haidt et al., 2003; Montada & Schneider, 1989). Across all five studies, anger about moral violations emerged even when only outgroup members were involved in the interaction. Neither group-relations, nor egoistic interests could account for anger and punishment in outgroup/outgroup interactions. This difference was only descriptive for the competitive outgroup (Study 4). Negative other-directed emotions mediate the relationship between unfairness and punishment, even at personal costs (Fehr & Gächter, 2002; Nelissen & Zeelenberg, 2009). Moral outrage also includes an impulse to punish the perpetrators in proportion to their wrongdoing (i.e., just desert; Darley & Pittman, 2003). Bernhard et al (2006), as well as Study 1a and 1b, showed that altruistic punishment tendencies increase with growing degree of unfairness. Moreover, all studies showed that moral violation explained preferences for harsh punishment better than group memberships. Punishment in response to outgroup interactions satisfies the general aim of establishing just desert; but it cannot re-establish group norms (S. A. Haslam et al., 1996; Vidmar, 2001; Wenzel et al., 2008).

Two alternative explanations challenged the notion that moral outrage is elicited irrespective of self-involvement. They suggest that either group-related concerns for ingroup victims (Batson et al., 2009), or empathy with the victims (Batson et al., 2007) trigger anger about moral violations. The present results propose that group-bound anger caused by a suffering ingroup member, if it emerges, only adds to the anger about the moral violation. We additionally tested the mediating role of empathy on anger in Study 4. In line with the suggestion of Batson and colleagues (2007), empathy enhanced anger about moral violations in Study 4. However, the direct effect of the moral violation on anger remained when the mediation of empathy with the victims was taken into account.

Previous studies showed that anger is a distinct affective reaction to violations of individual rights and to harm compared to disgust and contempt (Gutierrez & Giner-Sorolla, 2007; Rozin, Lowery, et al., 1999). Reading about the moral violations decreased participants' satisfaction (Study 3 and 4), and increased fear (Study 4) independently of group memberships. Anger still reacted stronger to the moral violation than other emotions. However, it cannot be concluded that anger is a unique reaction to moral violations in the present cases.

4.7.2 Victim or norm protection?

Differential results emerged for the role of victim and perpetrator group membership. In Study 1b and 2, altruistic punishment and anger were higher when ingroup victims were affected in contrast to outgroup victims in salient minimal groups. Torture in the Western society/"IS" context elicited more anger about and punishment of the ingroup perpetrators than outgroup perpetrators, independently of victim group membership. In the German-

French context, group memberships did not modify the reported anger, but increased punishment in intergroup encounters (especially with ingroup perpetrators). The following section outlines the results in more detail, connects them to previous theories and studies, and suggests how to integrate the seemingly contradictory findings.

Ingroup victims have been shown to raise altruistic punishment of unfair dictators within groups whose members are very attached and interdependent (Bernhard, Fischbacher, et al., 2006; Goette et al., 2006). Our studies extended these findings and indicate that in minimal groups, ingroup identification promotes anger and punishment on behalf of fellow group members. Ingroup identification has been shown to elicit group-bound emotions, such as anger on behalf of a fellow group member (Gordijn et al., 2001; Gordijn et al., 2006; Yzerbyt et al., 2003). Those motivate others to take action in favor of the suffering other. However, the group-based anger (on behalf of ingroup victims) only emerged in Study 1b and 2. We suggest that ingroup identification illustrates salience of the group context in minimal group studies, and less in natural group contexts. This is also indicated by the measurements, that capture recognition and the "feeling" that one belongs to the group.

There were no effects of victim group membership in Study 3 and 4. Supposedly, the German-French tourist context in Study 3 did not contrast ingroup and outgroup, but pronounced their cooperative relationship. Cooperative relationships are typically within-group relations (Brewer, 2007). Therefore, ingroup categorization might not have elicited group-based emotions (see Section 4.5.3). Study 4 did not show such differences in terms of ingroup victims. Batson et al (2009) suggested that even anger at harsh moral violations, such as torture, preferably emerges with a shared group membership with the victim. In line with Batson et al. (2009), Study 4 demonstrates that an outgroup perpetrator torturing an ingroup victim elicited more anger than an outgroup perpetrator torturing an outgroup member (see Table 12). When considering the full design, there was no main effect of victim group membership.

Further, people treat ingroup perpetrators harsher than outgroup perpetrators when they threaten ingroup cooperation and positive distinctiveness from the outgroup (Abrams et al., 2000; Shinada et al., 2004). Such an effect of perpetrator group membership was only found in Study 4. Here, moral behavior represented an ingroup stereotype that sharply contrasts the outgroup stereotype (i.e., "We" value human rights, whereas they, the "IS", are cruel and immoral.). In line with the Black Sheep Effect (Marques et al., 2001), the harsh reaction towards ingroup perpetrators may maintain moral superiority of the ingroup over the outgroup. The negative reactions express that such a behavior is unacceptable for “us”. In Study 4, this is also emphasized by shame which was elicited by ingroup perpetrators. Shame on behalf of the ingroup also motivates punishment of deviants, similarly to anger (Chekroun & Nugier, 2011). In Study 3, less content, but not more anger at ingroup perpetrators was observed. Ingroup perpetrators, however, elicited harshest punishment tendencies when affecting the outgroup. This emphasizes the cooperative nature of the intergroup relation. Disapproval of ingroup members who compromise the intergroup relation re-affirms common standards across group boundaries (e.g., of the superordinate group) in order to uphold friendly intergroup relations (J. Feinberg, 1965; Mackie et al., 2000; Wenzel, 2009). Additionally, people punish outgroup members leniently, when low punishment contributes to the ingroup’s moral image (Braun & Gollwitzer, 2012).

In Study 1a and 1b, participants showed more altruistic punishment of outgroup perpetrators, regardless of fair or unfair treatment. When punishment includes the inflictions of cost on perpetrators, outgroup discrimination by removing more resources from outgroup members than ingroup members has been observed (Bernhard, Fischbacher, et al., 2006; Schiller et al., 2014). In Study 1a and 1b, punishment (i.e., reduction of the perpetrators’ amount of money) was the only means to influence the money distribution between groups (see Section 4.4.3).

Negative reaction on behalf of ingroup victims, and less outgroup victims, as well as more anger and punishment of ingroup perpetrators compared to outgroup perpetrators illustrate and preserve ingroup biases (i.e., favorable evaluation and treatment of ingroup over outgroup members; Hewstone et al., 2002; McAuliffe & Dunham, 2016). On the one hand, anger and punishment on behalf of ingroup members enforces that they are treated favorably, and in accordance with moral standards. On the other hand, harsh reactions to moral violations within the group enforce ingroup cooperativeness (Fehr & Fischbacher, 2004b) and positive ingroup biases (Marques et al., 2001). The current findings indicate that these processes are malleable, depending on the intergroup relation, salience of the group context, degree of identification, and the consequences of the moral violation.

4.7.3 Limitations and future research

More research on this topic could address the limitations of the present studies and contribute to clarify their results. First, in Study 4, the pairwise comparison of anger in the outgroup/outgroup interaction was not significant. Recent recommendations for experimental practice (Cumming, 2014; Simmons, Nelson, & Simonsohn, 2011) suggest interpreting effect sizes, which were moderate to high. However, with a small sample size as in the pairwise comparisons in Study 4, cautious interpretations must be applied. Second, participants could have perceived different situations in the between-participant compared to the within-participants designs. In intergroup encounters a power imbalance between groups emerges because the perpetrator's group has the power to affect the victim's group, but not vice versa. This is prevented by the natural group scenarios. Moreover, in within-group encounters the group context is less salient. A valuable extension of our finding could focus on the interpretations of offenses (e.g., Okimoto & Wenzel, 2010). Third, we supposed that group-based emotions on behalf of ingroup victims (and subsequent punishment desires) would mainly show for highly identified group members. Conversely, ingroup identification also increases with group-based emotions (Kessler & Hollbach, 2005), and with it the desire to punish an ingroup offender (Okimoto & Wenzel, 2011). Study 2 and 4 indicated a trend that identification increased when the victim was an ingroup member. Thus, further work is required to determine the relationship between ingroup identification and the concern for ingroup victims. Fourth, hedonistic or group-related concerns could not account for our results in outgroup/outgroup interactions. However, we cannot fully exclude the possibility that empathy triggered outrage and punishment, as empathy has also been observed for outgroup members (e.g., Tarrant et al., 2009).

4.7.4 Conclusion

Moral outrage is useful to (group) cooperation. On a group level, anger triggers punishment to enforce cooperative norms (Fehr & Gächter, 2002). On an individual level angry punishers signals that morality is important to them, which stabilizes their personal positive relationships (DeScioli & Kurzban, 2013; Kurzban, DeScioli, & O'Brien, 2007). Both processes are more important within groups than between groups. Although moral norms might be group-bound (Haidt & Kesebir, 2010), people perceive their moral convictions as universally valid (Skitka, 2010). The current findings suggest that people are concerned about immoral behavior even in outgroup interactions. Research has indicated that cooperative interaction partners are preferred as future interaction partners (Baumard et al., 2013) or even as a future group member (Efferson, Lalive, & Fehr, 2008). Thus, punishment on behalf of strangers, even outgroup members, might increase an individual's chances for successful cooperation in the future.

5. General Discussion

5.1 Summary of the present findings

The present dissertation investigated cognitive, emotional, and behavioral responses to deviance, and in particular moral violations. Moral violations, such as cheating, trigger social conflicts, and threaten the maintenance of cooperation. Psychological mechanisms facilitate cooperation and enduring positive relations: People react aversively to moral violations, even on behalf of others. Subsequently, cheaters and other perpetrators are avoided and punished. The current studies take into account that most cooperation happens within social groups. Psychological attachment to an ingroup (i.e., ingroup identification) increases the levels of cooperation within a group. Group-specific reactions to perpetrators or victims of moral violations can facilitate group life and ingroup cooperation. Consequently, the present work hypothesized that reactions to moral violations are stronger within groups than between groups.

Three lines of research investigated the antecedents of psychological mechanisms that promote cooperation: memory for uncooperative and deviant group members, anger at moral violations on behalf of victims, and the influence of victim and perpetrator group membership on anger and punishment. In *Research Line I*, it was assumed and found that memory for uncooperative individuals evolves in ingroup but not outgroup contexts. Moreover, particularly highly identified group members remember deviant ingroup members. *Research Line II* tested whether anger about moral violations is moral outrage, or empathic anger. The results show that anger about moral violations emerges in response to the perpetrators intentions, and thereby reacts to moral violations independently from harmful consequences for cared-for-others. *Research Line III* combined the notion that involvement with perpetrator or victim accounts for anger at moral violations. It was found that moral outrage emerges irrespectively of shared group memberships, whereas reactions to ingroup victims and ingroup perpetrators are malleable.

The research represents a novel step for connecting knowledge in cognitive, moral, and group psychology. It shows that reactions to moral violations often generalize across contexts, and nevertheless are influenced by group processes.

An enhanced cheater memory was found in interpersonal encounters (Bell & Buchner, 2012; Buchner et al., 2009). As group contexts are important for successful cooperation, *Research Line I* pursued their integration with memory for uncooperative individuals. The results show that social categorization elicits memory advantages for uncooperative group members. Uncooperative members (i.e., violating fairness, norms in social exchanges) of novel experimental groups were remembered better, when they belonged to the ingroup compared to an outgroup (Study 1 and 2). A meaningful ingroup context (indicated by identification with a natural group) elicited enhanced memory for ingroup deviants (i.e., trustworthy and cheating; Study 3). Ingroup favoring biases remained stable throughout the experiments. The results support the notion that general memory mechanisms influence reputational memory for deviant group members.

First, incongruity with expectations towards targets and target behavior improves reputational memory (Bell, Buchner, Kroneisen, et al., 2012; Bell et al., 2015). People expect positive behavior within their groups, even in minimal groups (Balliet et al., 2014; Perdue et al., 1990; Tajfel et al., 1971). Study 1 and 2 showed that uncooperative ingroup members, who violate positive expectations towards the ingroup, are remembered best. An overall positive ingroup image, was expressed in group evaluations and guessing biases. Moreover, an ingroup is characterized through a common set of norms, that elicit expectations of ingroup behavior (Terry & Hogg, 1996). Group member who violate group norms (positively and negatively) were remembered especially well in Study 3, by those who identified with the natural ingroup.

Second, memory for ingroup members or behavior on an individual level increases with meaningfulness of the groups (Brewer et al., 1995). The presented studies show that relevant group members, such as uncooperative or deviant ingroup members were remembered better than respective outgroup members. Thus, group members do not generally remember any ingroup information especially well, but they remember the behavior of particular ingroup members (reputational memory). Study 3 showed that differential concerns account for enhanced reputational memory: Highly identified group members had improved memory for ingroup deviants compared to outgroup deviants. Similarly, authoritarians remembered negative ingroup members best, as they are highly concerned about antinormative behavior and ingroup threats (Kessler & Cohrs, 2008). In conclusion, memory might not be tuned to moral violations within groups, but general processes elicit enhanced memory for ingroup deviants. Remembering ingroup deviants ultimately facilitates coordination and cooperation among group members.

Interactions mostly happen within social groups. Thus, biased reactions to moral violations may illustrate a concern for the wellbeing of cared-for-victims (i.e., empathic anger; Batson et al., 2007). Conversely, anger at moral violations is assumed to emerge independently of self-involvement (i.e., moral outrage; Haidt, 2003; Montada & Schneider, 1989). *Research Line II* disentangled moral outrage and empathic anger by orthogonally manipulating the perpetrator's intentions (moral violation) and their consequences (suffering of the victim). It was found that intentions to harm a group member elicited more anger and punishment than the actual damage among active sport club members (Study 1). Anger and punishment emerged at perpetrator's intentions in severe moral violations, such as killing of children (Study 2). However, empathy with the victims also increases in responds to bad intentions, indicating presumed harm (see Gutierrez & Giner-Sorolla, 2007). An emphasis on the consequences before mentioning intentions revealed that killing of children elicits most anger when it was intentional, but only little anger when the victims were accidentally harmed (Study 3). Anger and empathy with the victims clearly diverged for the different appraisal situations. In line with anger, participants' punishment tendencies are also crucially influenced by the perpetrator's intentions, and less by the consequences of his actions. Moreover, the perpetrators intentions elicited fear. Sadness and dissatisfaction in contrast increase with the actual harm inflicted to the victims.

The current findings show that the wrongfulness (i.e., the perpetrator's intentions) and not the harmfulness (i.e., the victims' suffering) of moral violations elicit anger. They differ from prior suggestions that implications for the self trigger anger about moral violations via involvement with the victims (Batson et al., 2009; Batson et al., 2007). Instead, the results suggest that people indeed experience moral outrage and willingness to punish because of an immoral deed (Darley & Pittman, 2003; Haidt, 2003). Similarly, moral judgment and blame were found to be sensitive to bad intentions irrespectively of their consequences (Cushman, 2008; Haidt et al., 1993). Most moral violations that elicit anger and subsequent punishments affect the wellbeing of victims (Rozin, Lowery, et al., 1999). Indeed, a distinction between intentionality and harmful consequences was only observed in Study 3. To sum it all up, the present findings indicate consequences for the victims, and involvement with them, alone do not fully account for anger at moral violations.

Involvement with the victims might contribute to anger and punishment, as people despise negative treatment of ingroup members. Additionally, punishing the “right” perpetrator (i.e., ingroup perpetrators) may protect group norms. *Research Line III* tested effects of fairness, victim, and perpetrator group membership in a complete factorial design (see also Bernhard, Fischbacher, et al., 2006). Unfair treatment, in contrast to fair treatment, explained most variance in anger and punishment in each of the five studies. In fact, moral violations elicited anger even in outgroup interactions, and despite a competitive intergroup relation (Study 4). Enhanced anger about unfair sharing on behalf of ingroup victims showed in two minimal group studies and was triggered by ingroup identification (Study 1b and Study 2). In a natural and cooperative intergroup context, participants reported more punishment of, but not more anger towards perpetrators that offended outgroup victims (Study 3). More anger at and punishment of torturing ingroup perpetrators than outgroup perpetrators emerged in a natural intergroup conflict (Study 4).

The findings confirm that moral violations elicit moral outrage and subsequent punishment even independently of implications for the ingroup, as suggested before (see *Research Line II*; Darley & Pittman, 2003; Haidt, 2003). Increasing altruistic punishment to increasing deviations from fairness emerged in minimal group contexts (Study 1a and 1b). This extends previous findings from studies on tribal members, who punished unfairness irrespectively from ingroup involvement (Bernhard, Fischbacher, et al., 2006). Moreover, they contradict the conclusions by Batson and colleagues (2009), who proposed that reactions to moral violations are elicited by shared group membership with the victims.

Most studies investigating reactions to moral violation and their dependence of victim group membership did not systematically vary the group membership of perpetrators (Batson et al., 2009; Gordijn et al., 2001). The current results confirm that people experience emotions on behalf of fellow group members irrespectively of perpetrator group membership. Furthermore, this was shown to be triggered by ingroup identification (Yzerbyt et al., 2003). People also applied more punishment to ingroup perpetrators than outgroup perpetrators as response to a severe moral violation (i.e., torture). These results are in line with the Black Sheep Effect that predicts such reactions, when perpetrators threaten the ingroup’s moral standing and value consensus (Abrams et al., 2000; Marques et al., 2001). In a cooperative intergroup scenario, group membership did not modify anger, thereby supporting the idea that cooperation between groups expands perceived group-boundaries to include both groups (S.

L. Gaertner et al., 1990). The principal implication of *Research Line III* is that anger about and punishment of moral violations emerge despite of group boundaries. Group processes influence the reactions in two different ways: to protect ingroup victims and stabilize ingroup norms. The malleability of victim and perpetrator effects encourages further studies to examine their boundaries.

5.2 Implications for morality, cooperation and group life

The current work is based on theories which emphasize that the social function of morality is to foster cooperation (Greene, 2013; Haidt & Kesebir, 2010; Leach et al., 2015; Tomasello & Vaish, 2013). This derives from the assumption that personal investment in social groups ultimately leads to benefits for the self. Strong tendencies to cooperate within social groups facilitate the successful election of cooperation partners (i.e., fellow ingroup members; Balliet et al., 2014; Kramer & Brewer, 1984). In line with this argument, morality may be more important within than between groups. The current dissertation tested and extended this view by suggesting that basic group processes account for reactions to cheating and moral violations. Those group processes also apply to reactions to deviance from any group norm (see also: Harms & Skyrms, 2008; Kessler & Cohrs, 2008).

The mere acknowledgement that there are social norms does not imply that they are there to achieve a social function. However, much research suggests that adherence to group norms facilitates group functioning by, for example, increasing cohesiveness, coordination and cooperation within groups (Brewer, 2007; Chudek & Henrich, 2011; Fehr & Fischbacher, 2004a; Oakes et al., 1991; Wilson, Ostrom, & Cox, 2013). Even though most people adhere to common norms, some group members might threaten the group by deviation (Jetten & Hornsey, 2014). Psychological mechanisms, a “moral machinery”, facilitates dealing with such deviants in interpersonal encounters and in social groups, so that the norms are maintained (e.g., Baumard et al., 2013; Bell & Buchner, 2012; Bernhard, Fischbacher, et al., 2006; Darley & Pittman, 2003; DeScioli & Kurzban, 2013).

The present work extends previous research on reactions to moral violations by three advances:

First, the findings show how group processes influence cognitive and emotional reactions to deviants. Memory for uncooperative individuals emerges within social groups, whereas positive ingroup biases remain stable. Consequently, the ingroup may be approached in search of cooperation partners, even though some group members may be uncooperative. Moreover, only those perpetrators who are remembered can subsequently be punished to re-enforce ingroup cooperation. Similarly to memory, harsh emotional reactions to moral deviants emerge primarily in cooperative relationships within or between groups. Ingroup perpetrators elicit more anger than outgroup perpetrators when important ingroup values are at stake. This was not the case when moral violations in intergroup encounters burden friendly intergroup relations. Thus, psychological mechanisms that enforce cooperation are more sensitive to moral violations that threaten existing positive relationships, which are heuristically believed to exist within social groups (e.g., Brewer, 1999; Yamagishi et al., 1999).

Second, the current research highlights the importance of ingroup identification to reactions to deviance within groups. Even though it has been discussed that punishment illustrates group biases by its aim to protect ingroup members or/and group norms (e.g., Bernhard, Fischbacher, et al., 2006; McAuliffe & Dunham, 2016), the link to ingroup identification has been often overlooked. The present studies demonstrated that, on the one hand, a salient and meaningful group context triggers memory for ingroup deviants, and thereby facilitates partner choice in a pre-selected set of interaction partners. On the other hand, a salient ingroup context provokes harsh reactions to moral violations that affect ingroup members. Highly identified group members react harshly to those who offend their fellow group members. The present findings support the notion that ingroup identification is a psychological template that promotes cooperation beyond ingroup interdependence (Brewer, 2007; Seewald et al., 2016), because it elicits memory for and protection of potential interaction partners.

Third, in spite of the group-based nature of cooperation, most people perceive their moral agenda as impartial and universal, and are intolerant to any moral violations (Haidt et al., 2003; Skitka, 2010). The current research strengthens this idea, as emotional reactions and punishment tendencies towards perpetrators emerge even across group boundaries. In particular, moral outrage emerges regardless of the consequences of moral violations, and without implications for close others.

In sum, the principal theoretical implication is that cognitive reactions to moral violations are emphasized within social groups, which enables the maintenance of cooperation. Emotional reactions to moral violations operate universally, as similar moral violations elicit similar reactions across group boundaries. Such reactions are irrespective of consequences for the victims, indicating that punishing perpetrators is a goal in itself. Adding an ingroup context accentuates these emotional reactions in favor of the ingroup, indicating that group protection plays a role in reactions to moral violations.

5.3 Future directions

The presented research gives raises continuative and novel questions. To begin with, the studies showed that a salient group context provokes stronger reactions towards ingroup deviance. Group members, however, enforce not only moral norms, but any central norm (Terry & Hogg, 1996). Even small deviations from group-central (minimal) norms are punished harshly (Kessler, Neumann, Mummendey, Berthold, Schubert, & Waldzus, 2010). This suggests that any central group norm might elicit moral outrage. For example, failing to sacrifice to the gods can be considered “bad” in some cultures. Future research could focus on disentangling morality from other (social) norms.

Further, it was suggested that ingroup identification fosters reactions to moral violations beyond interdependence. Real interdependence and perceived interdependence are hardly distinguishable on a psychological level. For example, being a group member elicits perceived interdependence (Platow, Grace, & Smithson, 2012), expectations of reciprocity (Yamagishi et al., 1999), and mutual trust (Foddy et al., 2009). However, such reactions were mostly investigated by game theoretical approaches that translate group situation into experimental framings. Alternatively, one could determine the direct and indirect aims of reactions to moral violations in natural occurrences to specify the link between group situation and psychological processes. The importance of (central) group norms and wellbeing might exceed the longing to produce high (material) outcome in conflict situations. One might have to elaborate in each situation what people consider to be the greatest subjective threat for the group’s wellbeing, and if reactions to this threat facilitate group cooperation.

Moreover, people select those interaction partners who behave morally, and build sustainable relationships based on successful cooperation (Baumard et al., 2013). Prior research has shown that groups emerge on the basis of cooperation (Efferson et al., 2008). It would be interesting whether groups form in prospect of personal advantages through successful social exchanges or as a response to positive evaluations of interaction partners.

It was suggested that punishment re-enforces social norms. However, reward also fosters acceptable behavior (Balliet et al., 2011). Is harsh treatment of perpetrators even needed, when the social functions can be fulfilled by other means, such as programs of resocialization, or self-directed emotions such as shame and guilt? In contrast to reward, punishment additionally expresses disapproval of the punished behavior. This disapproval has been suggested to also discharge victims from guilt, distance the punisher from the deed, and illustrate that societal rules are valid (J. Feinberg, 1965). Further research should explore these "expressive" functions of punishment.

Last but not least, in the course of this work it was promoted that reactions to moral violations is a pro-social behavior. However, morality also has a dark side, which is also illustrated by reactions to moral violations. First, different groups often have diverging moral codes and beliefs (Haidt et al., 1993; Henrich et al., 2006). Moral conflicts could not only result from moral violations, but different applications of morality. However, as individuals regard their own moral convictions as impartial, true, and universal (Skitka, 2010), they are intolerant towards different sets of morality. In intergroup (and interpersonal) encounters, different moral agendas can result in harsh conflicts over what is right and what is wrong (Mikula & Wenzel, 2000). Second, moral arguments must not be rational, but are substantially influenced by a gut-feeling (Greene & Haidt, 2002; Tetlock et al., 2000). Thus, reactions to morality can conflict with the most rational solution for mutual benefit (Greene, 2013). Third, morality is a normative expression that is connoted with positivity and goodness, independently of its content. This implies that morality can be used instrumentally to legitimate any opinion, value, attitude, and behavior. For example, selfish aims are often legitimated by or disguised as moral aims (Batson, 2008). Consequently, reactions to moral violations might even hinder reconciliation and provoke vicious circles of punishment and revenge. On the other hand, reward for positive interactions following normative or "moral" behavior also foster cooperative behavior (Balliet et al., 2011). This implies that strong negative reactions to moral violations might not even be necessary to maintain cooperation.

5.4 Conclusion

The present dissertation shows that reactions to moral violations are influenced by psychological group processes, protecting group norms and fellow group members. This particular focus may encourage positive and sustainable relationships with other group members, for example with perpetrators we rehabilitated or victims we defend. Moreover, it supports the maintenance of mutual trust and care within the group. However, anger at moral deviants is a matter of the wrongfulness and not to the harmfulness of moral violations, and to a lesser degree emerges irrespectively of group boundaries. As anger drives the desire to punish perpetrators, reactions to moral deviance seem not to be applied to reach these utilitarian goals.

Returning to the example from the beginning of this dissertation, we can illustrate the findings in the case of Susan pushing Anna in the school yard. The results demonstrated that Susan is better remembered if she is a fellow class mate, whereas we care less about her if she belongs to a different class. Moreover, Susan is also remembered if she constantly does Anna's homework, and this is unusual in our class. Susan offending Anna also makes us angry and motivates our desire to punish her. This is even independent of whether we care for Anna's wellbeing or not. If Susan deliberately pushed in order to hurt Anna, we get very angry, and also a little bit afraid. This is also the case when Anna is not even hurt by Susan's mean-spirited attempt. But are Susan's and Anna's class affiliations important for our desire to punish Susan's wrongdoing? If we are really attached to our class, we care for our class-mate Anna's wellbeing more than for other victims of pushing. In fact, if we perceive our class as the nicest of all classes, we are willing to punish our class-mate Susan very harshly. In spite of those biases, we generally feel angry at and are punitive towards all pushers to some degree.

Cognitive, emotional, and behavioral reactions to deviants may initiate the avoidance of perpetrators, the thorough evaluation of behavioral standards, and the expectations of punishment for own deviance. Whereas reactions to deviants regulate social life, and thereby are useful, for example in global cooperation, they also bear societal risks. People have group-biases that make them pay more attention to and exaggerate punishment of deviance that concerns their group. This might hinder socialization of ingroup deviants, cooperation between groups, and also successful migration between groups, as new group members may not be familiar with the rules. Moreover, people universalize their moral reactions, instigating prejudice and discrimination of those who do not behave in line. To some extent tolerance, even within groups, fosters social life. However, if people would only care for their personal sake, cooperative and harmonious relations would hardly emerge.

References

Abrams, D., Marques, J. M., Bown, N., & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. *Journal of Personality and Social Psychology, 78*, 906-912. doi: 10.1037/0022-3514.78.5.906

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63*, 368-378. doi: 10.1037/0022-3514.63.3.368

Almor, A. & Sloman, S. A. (2000). Reasoning versus text processing in the Wason selection task: A nondeontic perspective on perspective effects. *Memory & Cognition, 28*, 1060-1070. doi: 10.3758/bf03209354

Altemeyer, B. (1981). *Right-wing authoritarianism*: University of Manitoba Press Winnipeg.

Altemeyer, B. (1996). *The authoritarian specter*. Cambridge, MA, US: Harvard University Press.

Alter, A. L., Kernochan, J., & Darley, J. M. (2007). Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. *Law and Human Behavior, 31*, 319-335. doi: 10.1007/s10979-006-9060-x

Ames, D. L. & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science, 24*, 1755-1762. doi: 10.1177/0956797613480507

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*, 1-70. doi: 10.1037/h0093718

Ask, K. & Pina, A. (2011). On being angry and punitive: How anger alters perception of criminal intent. *Social Psychological and Personality Science, 2*, 494-499. doi: 10.1177/1948550611398415

Averill, J. R. (2001). Studies on anger and aggression: Implications for theories of emotion. In W. G. Parrott & W. G. Parrott (Eds.), *Emotions in social psychology: Essential readings* (pp. 337-352). New York, NY, US: Psychology Press.

Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390-1396. doi: 10.1126/science.7466396

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin, 137*, 594-615. doi: 10.1037/a0023489

References

Balliet, D. & Van Lange, P. A. M. (2013). Trust, punishment, and cooperation across 18 societies: A meta-analysis. *Perspectives on Psychological Science, 8*, 363-379. doi: 10.1177/1745691613488533

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140*, 1556-1581. doi: 10.1037/a0037737

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition, 107*, 817-828. doi: 10.1016/j.cognition.2007.11.013

Barclay, P. & Lalumière, M. L. (2006). Do people differentially remember cheaters? *Human Nature, 17*, 98-113. doi: 10.1007/s12110-006-1022-y

Batchelder, W. H. & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548-564. doi: 10.1037/0033-295X.97.4.548

Batchelder, W. H. & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57-86. doi: 10.3758/BF03210812

Batson, C. D. (2008). Moral masquerades: Experimental exploration of the nature of moral motivation. *Phenomenology and the Cognitive Sciences, 7*, 51-66. doi: 10.1007/s11097-007-9058-y

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety, W. Ickes, J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3-15). Cambridge, MA, US: MIT Press.

Batson, C. D. (2011). What’s Wrong with Morality? *Emotion Review, 3*, 230-236. doi: 10.1177/1754073911402380

Batson, C. D., Chao, M. C., & Givens, J. M. (2009). Pursuing moral outrage: Anger at torture. *Journal of Experimental Social Psychology, 45*, 155-160. doi: 10.1016/j.jesp.2008.07.017

Batson, C. D., Kennedy, C. L., Nord, L.-A., Stocks, E. L., Fleming, D. Y. A., Marzette, C. M., . . . Zerger, T. (2007). Anger at unfairness: Is it moral outrage? *European Journal of Social Psychology, 37*, 1272-1285. doi: 10.1002/ejsp.434

Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences, 36*, 59-78. doi: 10.1017/S0140525X12000672

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*, 243-267. doi: 10.1037/0033-2909.115.2.243

Bayen, U. J. & Kuhlmann, B. G. (2011). Influences of source–item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language, 64*, 1-17. doi: 10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 197-215. doi: 10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28*, 480-500. doi: 10.3758/BF03198562

Beauchamp, T. L. (2001). *Philosophical ethics: An introduction to moral philosophy*. New York: McGraw-Hill.

Bell, R. & Buchner, A. (2009). Enhanced source memory for names of cheaters. *Evolutionary Psychology, 7*, 317-330. doi: 10.1177/147470490900700213

Bell, R. & Buchner, A. (2010a). Justice sensitivity and source memory for cheaters. *Journal of Research in Personality, 44*, 677-683. doi: 10.1016/j.jrp.2010.08.011

Bell, R. & Buchner, A. (2010b). Valence modulates source memory for faces. *Memory & Cognition, 38*, 29-41. doi: 10.3758/MC.38.1.29

Bell, R. & Buchner, A. (2011). Source memory for faces is determined by their emotional evaluation. *Emotion, 11*, 249-261. doi: 10.1037/a0022597

Bell, R. & Buchner, A. (2012). How Adaptive Is Memory for Cheaters? *Current Directions in Psychological Science, 21*, 403-408. doi: 10.1177/0963721412458525

Bell, R., Buchner, A., Erdfelder, E., Giang, T., Schain, C., & Riether, N. (2012). How specific is source memory for faces of cheaters? Evidence for categorical emotional tagging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 457-472. doi: 10.1037/a0026017

Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1512-1529. doi: 10.1037/a0028219

Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old–new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition, 117*, 261-275. doi: 10.1016/j.cognition.2010.08.020

Bell, R., Giang, T., & Buchner, A. (2012). Partial and specific source memory for faces associated to other- and self-relevant negative contexts. *Cognition & Emotion, 26*, 1036-1055. doi: 10.1080/02699931.2011.633988

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 456-472. doi: 10.1037/xlm0000034

Bell, R., Schain, C., & Echterhoff, G. (2014). How selfish is memory for cheaters? Evidence for moral and egoistic biases. *Cognition, 132*, 437-442. doi: http://dx.doi.org/10.1016/j.cognition.2014.05.001

Benard, S. (2012). Cohesion from conflict: Does intergroup conflict motivate intragroup norm enforcement and support for centralized leadership? *Social Psychology Quarterly, 75*, 107-130. doi: 10.1177/0190272512442397

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group Affiliation and Altruistic Norm Enforcement. *The American Economic Review, 96*, 217-221. doi: 10.1257/000282806777212594

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442*, 912-915. doi: 10.1038/nature04981

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*, 706-712. doi: 10.1111/j.1467-9280.2007.01964.x

Bettencourt, B. A., Brewer, M. B., Croak, M. R., & Miller, N. (1992). Cooperation and the reduction of intergroup bias: The role of reward structure and social orientation. *Journal of Experimental Social Psychology, 28*, 301-319. doi: 10.1016/0022-1031(92)90048-O

Boldry, J. G., Gaertner, L., & Quinn, J. (2007). Measuring the measures: A meta-analytic investigation of the measures of outgroup homogeneity. *Group Processes & Intergroup Relations, 10*, 157-178. doi: 10.1177/1368430207075153

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: an experimental study. *Games and Economic Behavior, 41*, 1-25. doi: http://dx.doi.org/10.1016/S0899-8256(02)00012-X

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science, 328*, 617-620. doi: 10.1126/science.1183665

Boyd, R. & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology & Sociobiology, 13*, 171-195. doi: 10.1016/0162-3095(92)90032-Y

Boyd, R. & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B, 364*, 3281-3288. doi: 10.1098/rstb.2009.0134 1471-2970

Brambilla, M. & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition, 32*, 397-408. doi: 10.1521/soco.2014.32.4.397

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). In-Group or Out-Group Extemity: Importance of the Threatened Social Identity. *Personality and Social Psychology Bulletin, 19*, 381-388. doi: 10.1177/0146167293194003

Braun, J. & Gollwitzer, M. (2012). Leniency for out-group offenders. *European Journal of Social Psychology, 42*, 883-892. doi: 10.1002/ejsp.1908

Bray, R. M. & Noble, A. M. (1978). Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization. *Journal of Personality and Social Psychology, 36*, 1424-1430. doi: 10.1037/0022-3514.36.12.1424

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*, 307-324. doi: 10.1037/0033-2909.86.2.307

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*, 475-482. doi: 10.1177/0146167291175001

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*, 429-444. doi: 10.1111/0022-4537.00126

Brewer, M. B. (2004). Taking the Social Origins of Human Nature Seriously: Toward a More Imperialist Social Psychology. *Personality and Social Psychology Review, 8*, 107-113. doi: 10.1207/s15327957pspr0802_3

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist, 62*, 728-738. doi: 10.1037/0003-066X.62.8.728

Brewer, M. B. & Campbell, D. T. (1976). *Ethnocentrism and intergroup attitudes: East African evidence*. Oxford, England: Sage.

Brewer, M. B., & Caporael, L. R. (2006). An Evolutionary Perspective on Social Identity: Revisiting Groups. In M. Schaller, J. A. Simpson, D. T. Kenrick, M. Schaller, J. A. Simpson, D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 143-161). Madison, CT, US: Psychosocial Press.

Brewer, M. B. & Harasty, A. S. (1996). Seeing groups as entities: The role of perceiver motivation. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition, Vol 3: The interpersonal context* (pp. 347-370). New York, NY, US: Guilford Press.

Brewer, M. B. & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology, 50*, 543-549. doi: 10.1037/0022-3514.50.3.543

Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology, 69*, 29-40. doi: 10.1037/0022-3514.69.1.29

Bröder, A. & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/Journal of Psychology, 215*, 52-60. doi: 10.1027/0044-3409.215.1.52

Buchner, A., Bell, R., Mehl, B., & Musch, J. (2009). No enhanced recognition memory, but better source memory for faces of cheaters. *Evolution and Human Behavior, 30*, 212-224. doi: 10.1016/j.evolhumbehav.2009.01.004

Buss, D. (2015). *Evolutionary psychology: The new science of the mind*: Psychology Press.

Campbell, D. T. (1965). *Ethnocentric and other altruistic motives.* Paper presented at the Nebraska symposium on motivation.

Caporael, L. R. (2001). Evolutionary psychology: Toward a unifying theory and a hybrid science. *Annual Review of Psychology, 52*, 607-628. doi: 10.1146/annurev.psych.52.1.607

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology, 42*, 437-451. doi: 10.1016/j.jesp.2005.06.007

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish?: Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*, 284-299. doi: 10.1037/0022-3514.83.2.284

Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of Personality and Social Psychology, 95*, 1316-1324. doi: 10.1037/a0012165

Caruso, E. M., Waytz, A., & Epley, N. (2010). The intentional mind and the hot hand: Perceiving intentions makes streaks seem likely to continue. *Cognition, 116*, 149-153. doi: http://dx.doi.org/10.1016/j.cognition.2010.04.006

Čehajić, S., Brown, R., & González, R. (2009). What do I Care? Perceived Ingroup Responsibility and Dehumanization as Predictors of Empathy Felt for the Victim Group. *Group Processes & Intergroup Relations, 12*, 715-729. doi: 10.1177/1368430209347727

Chaurand, N. & Brauer, M. (2008). What determines social control? People's reactions to counternormative behaviors in urban environments. *Journal of Applied Social Psychology, 38*, 1689-1715. doi: 10.1111/j.1559-1816.2008.00365.x

Chekroun, P. & Nugier, A. (2011). 'I'm ashamed because of you, so please, don't do that!': Reactions to deviance as a protection against a threat to social image. *European Journal of Social Psychology, 41*, 479-488. doi: 10.1002/ejsp.809

Chiappe, D., Brown, A., Dow, B., Koontz, J., Rodriguez, M., & McCulloch, K. (2004). Cheaters are looked at longer and remembered better than cooperators in social exchange situations. *Evolutionary Psychology, 2*, 108-120. doi: 10.1177/147470490400200117

Chudek, M. & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences, 15*, 218-226. doi: 10.1016/j.tics.2011.03.003

Cialdini, R. B. & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591-621. doi: 10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*, 1015-1026. doi: 10.1037/0022-3514.58.6.1015

Cialdini, R. B. & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, G. Lindzey, D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology, Vols 1 and 2 (4th ed)* (pp. 151-192). New York, NY, US: McGraw-Hill.

Cohen, E. (2012). The Evolution of Tag-Based Cooperation in Humans: The Case for Accent. *Current Anthropology, 53*, 588-616. doi: 10.1086/667654

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276. doi: 10.1016/0010-0277(89)90023-1

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York, NY US: Oxford University Press.

Cosmides, L., & Tooby, J. (2008). Can a general deontic logic capture the facts of human moral reasoning? How the mind interprets social exchange rules and detects cheaters. In W. Sinnott-Armstrong, W. Sinnott-Armstrong (Eds.) , *Moral psychology, Vol 1: The evolution of morality: Adaptations and innateness* (pp. 53-119). Cambridge, MA, US: MIT Press.

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review, 8*, 144-153. doi: 10.1177/1754073914558466

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7-29. doi: 10.1177/0956797613504966

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353-380. doi: 10.1016/j.cognition.2008.03.006

Darley, J. M. (2002). Just punishments: Research on retributional justice. In M. Ross, D. T. Miller, M. Ross & D. T. Miller (Eds.), *The justice motive in everyday life* (pp. 314-333). New York, NY, US: Cambridge University Press.

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior, 24*, 659-683. doi: 10.1023/A:1005552203727

Darley, J. M. & Huff, C. W. (1990). Heightened damage assessment as a result of the intentionality of the damage-causing act. *British Journal of Social Psychology, 29*, 181-188. doi: 10.1111/j.2044-8309.1990.tb00898.x

Darley, J. M. & Pittman, T. S. (2003). The Psychology of Compensatory and Retributive Justice. *Personality and Social Psychology Review, 7*, 324-336. doi: 10.1207/S15327957PSPR0704_05

Darley, J. M. & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology, 41*, 525-556. doi: 10.1146/annurev.ps.41.020190.002521

Darwin, C. (1871/1901). *The descent of man and selection in relation to sex* (new ed.). London, England: John Murray.

Davis, K. (1949). *Human society*. Oxford, England: Macmillan.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology, 35*, 1-11. doi: 10.1037/0022-3514.35.1.1

De Cremer, D. & van Vugt, M. (1998). Collective identity and cooperation in a public goods dilemma: A matter of trust or self-efficacy? *Current Research in Social Psychology, 3*, 1-11. Retrieved from http://www.uiowa.edu/crisp/prior-issues-covered

De Cremer, D. & Van Vugt, M. (1999). Social identification effects in social dilemmas: A transformation of motives. *European Journal of Social Psychology, 29*, 871-893. doi: 10.1002/(SICI)1099-0992(199911)29:7<871::AID-EJSP962>3.0.CO;2-I

de Rivera, J., Gerstmann, E., & Maisels, L. (2002). Acting righteously: The influence of attitude, moral responsibility, and emotional involvement. In M. Ross, D. T. Miller, M. Ross & D. T. Miller (Eds.), *The justice motive in everyday life* (pp. 271-288). New York, NY, US: Cambridge University Press.

DeScioli, P. & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin, 139*, 477-496. doi: 10.1037/a0029065

Doosje, B., Spears, R., de Redelijkheid, H., & van Onna, J. (2007). Memory for stereotype (in)consistent information: The role of in-group identification. *British Journal of Social Psychology, 46*, 115-128. doi: 10.1348/014466606X103517

Dovidio, J. F., Ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., ... Pearson, A. R. (2004). Perspective and prejudice: Antecedents and mediating mechanisms. *Personality and Social Psychology Bulletin, 30*, 1537-1549. doi: 10.1177/0146167204271177

Duckitt, J. (1989). Authoritarianism and Group Identification: A New View of an Old Construct. *Political Psychology, 10*, 63-84. doi: 10.2307/3791588

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8*, 100-110. doi: 10.1037/1089-2680.8.2.100

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature, 8*, 231-246. doi: 10.1007/BF02912493

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*, 351-362. doi: 10.3758/BRM.42.1.351

Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science, 321*, 1844-1849. doi: 10.1126/science.1155805

Ehrenberg, K. & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41*, 369-387. doi: 10.1016/j.jesp.2004.08.001

Eidelman, S. & Biernat, M. (2003). Derogating black sheep: Individual or group protection? *Journal of Experimental Social Psychology, 39*, 602-609. doi: 10.1016/S0022-1031(03)00042-8

Eidelman, S., Silvia, P. J., & Biernat, M. (2006). Responding to Deviance: Target Exclusion and Differential Devaluation. *Personality and Social Psychology Bulletin, 32*, 1153-1164. doi: 10.1177/0146167206288720

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology, 24*, 160-193. doi: 10.1080/10463283.2013.841490

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior, 62*, 287-303. doi: http://dx.doi.org/10.1016/j.geb.2007.06.001

Falk, A. & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*, 293-315. doi: http://dx.doi.org/10.1016/j.geb.2005.03.001

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi: 10.3758/BF03193146

Fehr, E. & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*, 187-190. doi: 10.1016/j.tics.2004.02.007

Fehr, E. & Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63-87. doi: http://dx.doi.org/10.1016/S1090-5138(04)00005-4

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*, 1-25. doi: 10.1007/s12110-002-1012-7

Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140. doi: 10.1038/415137a

Feinberg, J. (1965). The expressive function of punishment. *The Monist*, 397-423. doi: 10.5840/monist196549326

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and Ostracism Promote Cooperation in Groups. *Psychological Science*. doi: 10.1177/0956797613510184

Feldman, S. (2003). Enforcing Social Conformity: A Theory of Authoritarianism. *Political Psychology, 24*, 41-74. doi: 10.1111/0162-895X.00316

Feldman, S. & Stenner, K. (1997). Perceived Threat and Authoritarianism. *Political Psychology, 18*, 741-770. doi: 10.1111/0162-895X.00077

Fiddick, L. & Erlich, N. (2010). Giving it all away: Altruism and answers to the Wason selection task. *Evolution and Human Behavior, 31*, 131-140. doi: 10.1016/j.evolhumbehav.2009.08.003

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science, 20*, 419-422. doi: 10.1111/j.1467-9280.2009.02312.x

Forgas, J. P. & Fiedler, K. (1996). Us and them: Mood effects on intergroup discrimination. *Journal of Personality and Social Psychology, 70*, 28-40. doi: 10.1037/0022-3514.70.1.28

Frank, R. H. (1988). Passions within reason: The strategic role of the emotions. *(1988) Passions within reason: The strategic role of the emotions,* pp 304. New York, NY. US: W W Norton & Co.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology & Sociobiology, 14*, 247-256. doi: 10.1016/0162-3095(93)90020-I

Frijda, N. H. (1988). The laws of emotion. *American psychologist, 43*, 349. doi: 10.1037/0003-066X.43.5.349

Fritsche, I., Jonas, E., & Kessler, T. (2011). Collective reactions to threat: Implications for intergroup conflict and for solving societal crises. *Social Issues and Policy Review, 5*, 101-136. doi: 10.1111/j.1751-2409.2011.01027.x

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent. *Personality and Social Psychology Bulletin, 40*, 986-997. doi: 10.1177/0146167214533130

Funke, F. (2005). The Dimensionality of Right-Wing Authoritarianism: Lessons from the Dilemma between Theory and Measurement. *Political Psychology, 26*, 195-218. doi: 10.1111/j.1467-9221.2005.00415.x

Gaertner, L., Iuzzini, J., Witt, M. G., & Oriña, M. M. (2006). Us without them: Evidence for an intragroup origin of positive in-group regard. *Journal of Personality and Social Psychology, 90*, 426-439. doi: 10.1037/0022-3514.90.3.426

Gaertner, S. L., Mann, J. A., Dovidio, J. F., Murrell, A. J., & Pomare, M. (1990). How does cooperation reduce intergroup bias? *Journal of Personality and Social Psychology, 59*, 692-704. doi: 10.1037/0022-3514.59.4.692

Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. *Journal of Personality and Social Psychology, 102*, 941-960. doi: 10.1037/a0027233

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology, 39*, 16-30. doi: 10.1016/S0022-1031(02)00517-6

Gigerenzer, G. & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127-171. doi: 10.1016/0010-0277(92)90060-U

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24*, 153-172. doi: 10.1016/S1090-5138(02)00157-5

Glanzer, M. & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8-20. doi: 10.3758/BF03198438

Glanzer, M. & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 5-16. doi: 10.1037/0278-7393.16.1.5

Goette, L., Huffman, D., & Meier, S (2006). The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence using Random Assignment to Real Social Groups. IZA Discussion Paper No. 2020; FRB of Boston Working Paper No. 06-7. Retrieved from http://ssrn.com/abstract=892343

Goette, L., Huffman, D., Meier, S., & Sutter, M. (2012). Competition between organizational groups: Its impact on altruistic and antisocial motivations. *Management Science, 58*, 948-960. doi: 10.1287/mnsc.1110.1466

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology, 29*, 781-795. doi: 0.1002/(SICI)1099-0992(199908/09)29:5/6<781::AID-EJSP960>3.0.CO;2-3

Gollwitzer, M. & Keller, L. (2010). What you did only matters if you are one of us: Offenders' group membership moderates the effect of criminal history on punishment severity. *Social Psychology, 41*, 20-26. doi: 10.1027/1864-9335/a000004

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology, 41*, 364-374. doi: 10.1002/ejsp.782

Gordijn, E. H., Wigboldus, D., & Yzerbyt, V. (2001). Emotional consequences of categorizing victims of negative outgroup behavior as ingroup or outgroup. *Group Processes & Intergroup Relations, 4*, 317-326. doi: 10.1177/1368430201004004002

Gordijn, E. H., Yzerbyt, V., Wigboldus, D., & Dumont, M. (2006). Emotional reactions to harmful intergroup behavior. *European Journal of Social Psychology, 36*, 15-30. doi: 10.1002/ejsp.296

Gould, R. V. (1999). Collective violence and group solidarity: Evidence from a feuding society. *American Sociological Review, 64*, 356-380. doi: 10.2307/2657491

Graham, J. & Haidt, J. (2010). Beyond beliefs: Religions bind individuals into moral communities. *Personality and Social Psychology Review, 14*, 140-150. doi: 10.1177/1088868309353415

Gramzow, R. H. & Gaertner, L. (2005). Self-Esteem and Favoritism Toward Novel In-Groups: The Self as an Evaluative Base. *Journal of Personality and Social Psychology, 88*, 801-815. doi: 10.1037/0022-3514.88.5.801

Gramzow, R. H., Gaertner, L., & Sedikides, C. (2001). Memory for in-group and out-group information in a minimal group context: The self as an informational base. *Journal of Personality and Social Psychology, 80*, 188-205. doi: 10.1037/0022-3514.80.2.188

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry, 23*, 206-215. doi: 10.1080/1047840X.2012.686247

Gray, K. & Wegner, D. M. (2011). Dimensions of Moral Emotions. *Emotion Review, 3*, 258-260. doi: 10.1177/1754073911402388

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*, 101-124. doi: 10.1080/1047840X.2012.651387

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York, NY, US: Penguin Press.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*, 364-371. doi: http://dx.doi.org/10.1016/j.cognition.2009.02.001

Greene, J. D. & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*, 517-523. doi: 10.1016/S1364-6613(02)02011-9

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences, 35*, 1-15. doi: doi:10.1017/S0140525X11000069

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312*, 108-111. doi: 10.1126/science.1123633

Gutierrez, R. & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion, 7*, 853-868. doi: 10.1037/1528-3542.7.4.853

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi: 10.1037/0033-295X.108.4.814

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. Oxford: Oxford University Press.(pp. 852-870).

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998-1002. doi: 10.1126/science.1137651

Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*, 65-72. doi: 10.1111/j.1745-6916.2008.00063.x

Haidt, J. & Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert, G. Lindzey, S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology, Vol 2 (5th ed)* (pp. 797-832). Hoboken, NJ, US: John Wiley & Sons Inc.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613-628. doi: 10.1037/0022-3514.65.4.613

Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating Diversities: Moral Diversity Is Not Like Other Kind. *Journal of Applied Social Psychology, 33*, 1-36. doi: 10.1111/j.1559-1816.2003.tb02071.x

Hardin, G. (1968). The Tragedy of the Commons. *Science, 162*, 1243-1248. doi: 10.1126/science.162.3859.1243

Harms, W. & Skyrms, B. (2008). Evolution of moral norms. In M. Ruse (Ed.), *Oxford handbook of the philosophy of biology*. Oxford, UK: Oxford University Press.

Haslam, N. & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology, 65*, 399-423. doi: 10.1146/annurev-psych-010213-115045

Haslam, S. A., McGarty, C., & Turner, J. C. (1996). Salient group memberships and persuasion: The role of social identity in the validation of beliefs. In J. L. Nye, A. M. Brower, J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition? Research on socially shared cognition in small groups* (pp. 29-56). Thousand Oaks, CA, US: Sage Publications, Inc.

Haslam, S. A., Oakes, P. J., Turner, J. C., & McGarty, C. (1995). Social categorization and group homogeneity: Changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness. *British Journal of Social Psychology, 34*, 139-160. doi: 10.1111/j.2044-8309.1995.tb01054.x

Hastie, R. & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25-38. doi: 10.1037/0022-3514.37.1.25

Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong.* New York, US: Ecco/HarperCollins Publishers.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization, 53*, 3-35. doi: http://dx.doi.org/10.1016/S0167-2681(03)00094-5

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly Punishment Across Human Societies. *Science, 312*, 1767-1770.

Herlitz, A. & Lovén, J. (2013). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition, 21*, 1306-1336. doi: 10.1080/13506285.2013.823140

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology, 53*, 575-604. doi:10.1146/annurev.psych.53.100901.135109

Hicks, J. L. & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language, 48*, 489-501. doi: 10.1016/S0749-596X(02)00537-5

Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature, 13*, 105-128. doi: 10.1007/s12110-002-1016-3

Hoffman, M. L. (1990). Empathy and justice motivation. *Motivation and emotion, 14*, 151-172. doi: 10.1007/BF00991641

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York, NY, US: Cambridge University Press.

Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review, 5*, 184-200. doi: 10.1207/S15327957PSPR0503_1

Hogg, M. A. & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes*. Florence, KY, US: Taylor & Frances/Routledge.

Houston, D. M. & Andreopoulou, A. (2003). Tests of both corollaries of social identity theory's self-esteem hypothesis in real group settings. *British Journal of Social Psychology, 42*, 357-370. doi: 10.1348/014466603322438206

Howard, J. W. & Rothbart, M. (1980). Social categorization and memory for in-group and out-group behavior. *Journal of Personality and Social Psychology, 38*, 301-310. doi: 10.1037/0022-3514.38.2.301

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*, 1168-1187. doi: 10.1037/a0020463

Hutcherson, C. A. & Gross, J. J. (2011). The moral emotions: A social–functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100*, 719-737. doi: 10.1037/a0022408

Hutchison, P. & Abrams, D. (2003). Ingroup identification moderates stereotype change in reaction to ingroup deviance. *European Journal of Social Psychology, 33*, 497-506. doi: 10.1002/ejsp.157

Hutchison, P., Abrams, D., Gutierrez, R., & Viki, G. T. (2008). Getting rid of the bad ones: The relationship between group identification, deviant derogation, and identity maintenance. *Journal of Experimental Social Psychology, 44*, 874-881. doi: 10.1016/j.jesp.2007.09.001

Iyer, A., Schmader, T., & Lickel, B. (2007). Why Individuals Protest the Perceived Transgressions of Their Country: The Role of Anger, Shame, and Guilt. *Personality and Social Psychology Bulletin, 33*, 572-587. doi: 10.1177/0146167206297402

Janoff-Bulman, R. & Carnes, N. C. (2013). Surveying the moral landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review, 17*, 219-236.

Jetten, J. & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology, 65*, 461-485. doi: 10.1146/annurev-psych-010213-115151

Jetten, J., Spears, R., & Manstead, A. S. R. (1997). Strength of identification and intergroup differentiation: The influence of group norms. *European Journal of Social Psychology, 27*, 603-609. doi: 10.1002/(SICI)1099-0992

Johns, M., Schmader, T., & Lickel, B. (2005). Ashamed to be an American? The role of identification in predicting vicarious shame for anti-Arab prejudice after 9–11. *Self and Identity, 4*, 331-348.

Kant, I. (1785 /2007). *Grundlegung zur Metaphysik der Sitten*. Frankfurt am Main: Suhrkamp.

Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review, 20*, 693-719. doi: 10.3758/s13423-013-0407-2

Keller, J. (2008). On the development of regulatory focus: The role of parenting styles. *European Journal of Social Psychology, 38*, 354-364. doi: 10.1002/ejsp.460

Kelley, H. H. & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York: Wiley.

Keltner, D., Ellsworth, P. C., & Edwards, K. (1993). Beyond simple pessimism: Effects of sadness and anger on social perception. *Journal of Personality and Social Psychology, 64*, 740-752. doi: 10.1037/0022-3514.64.5.740

Kensinger, E. A. (2007). Negative Emotion Enhances Memory Accuracy: Behavioral and Neuroimaging Evidence. *Current Directions in Psychological Science, 16*, 213-218. doi: 10.1111/j.1467-8721.2007.00506.x

Kensinger, E. A. & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*, 1169-1180. doi: 10.3758/BF03195800

Kerr, N. L., Garst, J., Lewandowski, D. A., & Harris, S. E. (1997). That still, small voice: Commitment to cooperate as an internalized versus a social norm. *Personality and Social Psychology Bulletin, 23*, 1300-1311. doi: 10.1177/01461672972312007

Kerr, N. L., Hymes, R. W., Anderson, A. B., & Weathers, J. E. (1995). Defendant-juror similarity and mock juror judgments. *Law and Human Behavior, 19*, 545-567. doi: 10.1007/BF01499374

Kerr, N. L. & Kaufman-Gilliland, C. M. (1994). Communication, commitment, and cooperation in social dilemma. *Journal of Personality and Social Psychology, 66*, 513-529. doi: 10.1037/0022-3514.66.3.513

Kerr, N. L., Rumble, A. C., Park, E. S., Ouwerkerk, J. W., Parks, C. D., Gallucci, M., & van Lange, P. A. M. (2009). 'How many bad apples does it take to spoil the whole barrel?': Social exclusion and toleration for bad apples. *Journal of Experimental Social Psychology, 45*, 603-613. doi: 10.1016/j.jesp.2009.02.017

Kessler, T. & Cohrs, J. C. (2008). The evolution of authoritarian processes: Fostering cooperation in large-scale groups. *Group Dynamics: Theory, Research, and Practice, 12*, 73-84. doi: 10.1037/1089-2699.12.1.73

Kessler, T. & Hollbach, S. (2005). Group-based emotions as determinants of ingroup identification. *Journal of Experimental Social Psychology, 41*, 677-685. doi: 10.1016/j.jesp.2005.01.001

Kessler, T. & Mummendey, A. (2002). Sequential or parallel processes? A longitudinal field study concerning determinants of identity-management strategies. *Journal of Personality and Social Psychology, 82*, 75-88. doi: 10.1037/0022-3514.82.1.75

Kessler, T., Neumann, J., Mummendey, A., Berthold, A., Schubert, T., & Waldzus, S. (2010). How do we assign punishment? The impact of minimal and maximal standards on the evaluation of deviants. *Personality and Social Psychology Bulletin, 36*, 1213-1224. doi: 10.1177/0146167210380603

Klauer, K. C. & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin, 26*, 1074-1093. doi: 10.1177/01461672002611005

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*, 309-324. doi: 10.1080/09515080307771

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. *Moral development and behavior: Theory, research, and social issues*, 31-53.

Kramer, R. M. & Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology, 46*, 1044-1057. doi: 10.1037/0022-3514.46.5.1044

Kramer, R. M. & Goldman, L. (1995). Helping the group or helping yourself? Social motives and group identity in resource dilemmas. In D. A. Schroeder (Ed.), *Social dilemmas: Perspectives on individuals and groups* (pp. 49-67): Greenwood Publishing Group.

Kroneisen, M. & Bell, R. (2012). Sex, cheating, and disgust: Enhanced source memory for trait information that violates gender stereotypes. *Memory*, 1-15. doi: 10.1080/09658211.2012.713971

Kuppens, T. & Yzerbyt, V. Y. (2012). Group-based emotions: The impact of social identity on appraisals, emotions, and behaviors. *Basic and Applied Social Psychology, 34*, 20-33. doi: 10.1080/01973533.2011.637474

Küppers, V. & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *The Quarterly Journal of Experimental Psychology, 67*, 2042-2059. doi: 10.1080/17470218.2014.904914

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior, 28*, 75-84. doi: 10.1016/j.evolhumbehav.2006.06.001

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

Leach, C. W., Bilali, R., & Pagliaro, S. (2015). Groups and morality. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, J. A. Simpson, M. Mikulincer, P. R. Shaver, . . . J. A. Simpson (Eds.), *APA handbook of personality and social psychology, Volume 2: Group processes* (pp. 123-149). Washington, DC, US: American Psychological Association.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234-249. doi: 10.1037/0022-3514.93.2.234

Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: Schadenfreude at the suffering of another group. *Journal of Personality and Social Psychology, 84*, 932-943. doi: 10.1037/0022-3514.84.5.932

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., . . . Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology, 95*, 144-165. doi: 10.1037/0022-3514.95.1.144

Lerner, M. J. (1980). The Belief in a Just World *The Belief in a Just World: A Fundamental Delusion* (pp. 9-30). Boston, MA: Springer US.

LeVine, R. A. & Campbell, D. T. (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior*. Oxford, England: John Wiley & Sons.

Li, Y., Li, Q., & Guo, C. (2009). Differences of relevance in implicit and explicit memory tests: An ERP study. *Chinese Science Bulletin, 54*, 2669-2680. doi: 10.1007/s11434-009-0396-8

Lickel, B., Schmader, T., Curtis, M., Scarnier, M., & Ames, D. R. (2005). Vicarious Shame and Guilt. *Group Processes & Intergroup Relations, 8*, 145-157. doi: 10.1177/1368430205051064

Lieberman, D. & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology, 5*, 289-305. doi: 10.1177/147470490700500203

Little, A. C., Jones, B. C., DeBruine, L. M., & Dunbar, R. I. M. (2013). Accuracy in discrimination of self-reported cooperators using static facial information. *Personality and Individual Differences, 54*, 507-512. doi: 10.1016/j.paid.2012.10.018

Livingstone, A. G., Haslam, S. A., Postmes, T., & Jetten, J. (2011). “We are, therefore we should”: Evidence that in-group identification mediates the acquisition of in-group norms. *Journal of Applied Social Psychology, 41*, 1857-1876. doi: 10.1111/j.1559-1816.2011.00794.x

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces - KDEF CD Rom from the Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.*

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology, 79*, 602-616. doi: 10.1037/0022-3514.79.4.602

Mackie, D. M. & Smith, E. R. (2002). Intergroup emotions and the social self: Prejudice reconceptualized as differentiated reactions to outgroups. *The social self: Cognitive, interpersonal, and intergroup perspectives*, 309-326.

Malle, B. F. & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101-121. doi: 10.1006/jesp.1996.1314

Marques, J. M., Abrams, D., Paez, D., & Martinez-Taboada, C. (1998). The role of categorization and in-group norms in judgments of groups and their members. *Journal of Personality and Social Psychology, 75*, 976-988.

Marques, J. M., Abrams, D., & Serôdio, R. G. (2001). Being better by being right: Subjective group dynamics and derogation of in-group deviants when generic norms are undermined. *Journal of Personality and Social Psychology, 81*, 436-447. doi: 10.1037/0022-3514.81.3.436

Marques, J. M. & Paez, D. (1994). The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability. *European Review of Social Psychology, 5*, 37-68. doi: 10.1080/14792779543000011

Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The "Black Sheep Effect": Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology, 18*, 1-16. doi: 10.1002/ejsp.2420180102

McAuliffe, K. & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*. doi: 10.1098/rstb.2015.0073

McCann, S. J. H. (2008). Societal threat, authoritarianism, conservatism, and U.S. state death penalty sentencing (1977-2004). *Journal of Personality and Social Psychology, 94*, 913-923. doi: 10.1037/0022-3514.94.5.913

McGraw, K. M. (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of Personality and Social Psychology, 53*, 247-256. doi: 10.1037/0022-3514.53.2.247

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology & Sociobiology, 17*, 119-128. doi: 10.1016/0162-3095(95)00131-X

Mehl, B. & Buchner, A. (2008). No enhanced memory for faces of cheaters. *Evolution and Human Behavior, 29*, 35-41. doi: 10.1016/j.evolhumbehav.2007.08.001

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, 658-673. Retrieved from http://www.jstor.org/stable/2118074

Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For Members Only: Ingroup Punishment of Fairness Norm Violations in the Ultimatum Game. *Social Psychological and Personality Science, 5*, 662-670. doi: 10.1177/1948550614527115

Messick, D. M. & Mackie, D. M. (1989). Intergroup relations. *Annual Review of Psychology, 40*, 45-81.

Mieth, L., Bell, R., & Buchner, A. (2016). Memory and disgust: Effects of appearance-congruent and appearance-incongruent information on source memory for food. *Memory, 24*, 629-639. doi: 10.1080/09658211.2015.1034139

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*, 143-152. doi: 10.1016/j.tics.2006.12.007

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and Social Psychology Bulletin, 24*, 769-783. doi: 10.1177/0146167298247009

Mikula, G. & Wenzel, M. (2000). Justice and social conflict. *International Journal of Psychology, 35*, 126-135. doi: 10.1080/002075900399420

Minear, M. & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630-633. doi: 10.3758/BF03206543

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In D. Narvaez, D. K. Lapsley, D. Narvaez, D. K. Lapsley (Eds.) , *Personality, identity, and character: Explorations in moral psychology* (pp. 341-354). New York, NY, US: Cambridge University Press. doi:10.1017/CBO9780511627125.016

Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology, 11*, 99-111. doi: 10.1037/1089-2680.11.2.99

Montada, L. (1998). Justice: Just a rational choice? *Social Justice Research, 11*, 81-101. doi: 10.1023/A:1023299119352

Montada, L. & Schneider, A. (1989). Justice and emotional reactions to the disadvantaged. *Social Justice Research, 3*, 313-344. doi: 10.1007/BF01048081

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review, 5*, 119-124. doi: 10.1177/1754073912468165

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42-54. doi: 10.3758/BRM.42.1.42

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology, 22*, 103-122. doi: 10.1002/ejsp.2420220202

Mummendey, A., Kessler, T., Klink, A., & Mielke, R. (1999). Strategies to cope with negative social identity: Predictions by social identity theory and relative deprivation theory. *Journal of Personality and Social Psychology, 76*, 229-245. doi: 10.1037/0022-3514.76.2.229

Nairne, J. S. & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science, 17*, 239-243. doi: 10.1111/j.1467-8721.2008.00582.x

Nelissen, R. M. A. & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision Making, 4*, 543-553.

Noë, R. & Hammerstein, P. (1994). Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology, 35*, 1-11. doi: 10.1007/BF00167053

Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science, 314*, 1560-1563. doi: 10.1126/science.1133755

Nowak, M. A. & Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of theoretical biology, 194*, 561-574.

Nowak, M. A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291-1298. doi: 10.1038/nature04131

Nugier, A., Chekroun, P., Pierre, K., & Niedenthal, P. M. (2009). Group membership influences social control of perpetrators of uncivil behaviors. *European Journal of Social Psychology, 39*, 1126-1134. doi: 10.1002/ejsp.602

O'Mara, E. M., Jackson, L. E., Batson, C. D., & Gaertner, L. (2011). Will moral outrage stand up? Distinguishing among emotional reactions to a moral violation. *European Journal of Social Psychology, 41*, 173-179. doi: 10.1002/ejsp.754

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology, 30*, 125-144. doi: 10.1111/j.2044-8309.1991.tb00930.x

Oda, R. (1997). Biased face recognition in the prisoner's dilemma game. *Evolution and Human Behavior, 18*, 309-315. doi: 10.1016/S1090-5138(97)00014-7

Oda, R. & Nakajima, S. (2010). Biased face recognition in the Faith Game. *Evolution and Human Behavior, 31*, 118-122. doi: 10.1016/j.evolhumbehav.2009.08.005

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature, 441*, 502-505. doi: http://www.nature.com/nature/journal/v441/n7092/suppinfo/nature04605_S1.html

Okimoto, T. G. & Wenzel, M. (2010). The symbolic identity implications of inter and intra-group transgressions. *European Journal of Social Psychology, 40*, 552-562.

Okimoto, T. G. & Wenzel, M. (2011). Third-party punishment and symbolic intragroup status. *Journal of Experimental Social Psychology, 47*, 709-718. doi: http://dx.doi.org/10.1016/j.jesp.2011.02.001

Ostrom, T. M., Carpenter, S. L., Sedikides, C., & Li, F. (1993). Differential processing of in-group and out-group information. *Journal of Personality and Social Psychology, 64*, 21-34. doi: 10.1037/0022-3514.64.1.21

Ostrom, T. M. & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin, 112*, 536-552. doi: 10.1037/0033-2909.112.3.536

Otten, S. & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: evidence from effective priming tasks. *European Journal of Social Psychology, 29*, 1049-1071. doi: 10.1002/(SICI)1099-0992(199912)29:8<1049::AID-EJSP985>3.0.CO;2-Q

Pagliaro, S., Ellemers, N., & Barreto, M. (2011). Sharing moral values: Anticipated ingroup respect as a determinant of adherence to morality-based (but not competence-based) group norms. *Personality and Social Psychology Bulletin, 37*, 1117-1129. doi: 10.1177/0146167211406906

Park, B. & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology, 42*, 1051. doi: 1051-1068. doi:10.1037/0022-3514.42.6.1051

Parks, C. D. & Stone, A. B. (2010). The desire to expel unselfish members from the group. *Journal of Personality and Social Psychology, 99*, 303-310. doi: 10.1037/a0018403

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. Retrived from http://bildungswissenschaften.uni-saarland.de/personal/paulus/homepage/empathie.html

Peck, J. R. (1993). Friendship and the evolution of co-operation. *Journal of Theoretical Biology, 162*, 195-228. doi: 10.1006/jtbi.1993.1083

Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology, 59*, 475-486. doi: 10.1037/0022-3514.59.3.475

Piaget, J. (1932). *The moral judgment of the child*. Oxford, England: Harcourt, Brace.

PICS. (2008). *Psychological Image Collection at Stirling*. Retrieved from: http://pics.psych.stir.ac.uk/2D_face_sets.htm

Pinto, I. R., Marques, J. M., Levine, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who triggers the black sheep effect? *Journal of Personality and Social Psychology, 99*, 107-119. doi: 10.1037/a0018187

Platow, M. J., Grace, D. M., & Smithson, M. J. (2012). Examining the preconditions for psychological group membership: Perceived social interdependence as the outcome of self-categorization. *Social Psychological and Personality Science, 3*, 5-13. doi: 10.1177/1948550611407081

Preacher, K. J. & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers, 36*, 717-731. doi: 10.3758/BF03206553

Preacher, K. J. & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891. doi: 10.3758/BRM.40.3.879

Quigley, B. M. & Tedeschi, J. T. (1996). Mediating Effects of Blame Attributions on Feelings of Anger. *Personality and Social Psychology Bulletin, 22*, 1280-1288. doi: 10.1177/01461672962212008

Rabbie, J. M. & Horwitz, M. (1988). Categories versus groups as explanatory concepts in intergroup relations. *European Journal of Social Psychology, 18*, 117-123. doi: 10.1002/ejsp.2420180204

Rai, T. S. & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118*, 57-75. doi: 10.1037/a0021867

Rockenbach, B. & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature, 444*, 718-723. doi: 10.1038/nature05229

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133-148. doi: 10.1037/0033-2909.121.1.133

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition, 108*, 771-780. doi: http://dx.doi.org/10.1016/j.cognition.2008.07.001

Rozin, P., Haidt, J., McCauley, C., Dunlop, L., & Ashmore, M. (1999). Individual differences in disgust sensitivity: Comparisons and evaluations of paper-and-pencil versus behavioral measures. *Journal of Research in Personality, 33*, 330-351. doi: 10.1006/jrpe.1999.2251

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*, 574-586. doi: 10.1037/0022-3514.76.4.574

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125*, 207-218. doi: 10.1016/j.cognition.2012.06.017

Runciman, W. G. (1966). Relative deprivation and social justice: Study attitudes social inequality in 20th century England. University of California Press

Russell, P. S. & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion, 11*, 233-240. doi: 10.1037/a0022598

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science, 300*, 1755-1758. doi: 10.1126/science.1082976

Schaller, M. & Maass, A. (1989). Illusory correlation and social categorization: Toward an integration of motivational and cognitive factors in stereotype formation. *Journal of Personality and Social Psychology, 56*, 709-721. doi: 10.1037/0022-3514.56.5.709

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior, 35*, 169-175. doi: 10.1016/j.evolhumbehav.2013.12.006

Seewald, D., Hechler, S., & Kessler, T. (2016). Divorcing the puzzles: When group identities foster in-group cooperation. *The Behavioral and brain sciences, 39*, e23. doi: 10.1017/S0140525X15000539

Seip, E. C., Dijk, W. W., & Rotteveel, M. (2014). Anger motivates costly punishment of unfair behavior. *Motivation and Emotion*. doi: 10.1007/s11031-014-9395-4

Shepherd, L., Spears, R., & Manstead, A. S. R. (2013). 'This will bring shame on our nation': The role of anticipated group-based emotions on collective action. *Journal of Experimental Social Psychology, 49*, 42-57. doi: 10.1016/j.jesp.2012.07.011

Sherif, M. (1936). *The psychology of social norms*. Oxford, England: Harper.

Sherif, M. & Sherif, C. W. (1953). *Groups in harmony and tension; an integration of studies of intergroup relations*. Oxford, England: Harper & Brothers.

Sherman, J. W., Klein, S. B., Laskey, A., & Wyer, N. A. (1998). Intergroup bias in group judgment processes: The role of behavioral memories. *Journal of Experimental Social Psychology, 34*, 51-65. doi: 10.1006/jesp.1997.1342

Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies:"Altruistic" punishment of in-group members. *Evolution and Human Behavior, 25*, 379-393. doi: 10.1016/j.evolhumbehav.2004.08.001

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83). Chicago, IL: University of Chicago Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359-1366. doi: 10.1177/0956797611417632

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass, 4*, 267-281. doi: 10.1111/j.1751-9004.2010.00254.x

Smith, E. R. (1993). Social identity and social emotions: toward new concepitualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition and stereotyping: Interactive processes in group perception* (pp. 297).

Smith, E. R. & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality and Social Psychology Bulletin, 22*, 635-642. doi: 10.1177/0146167296226008

Smith, J. R., Hogg, M. A., Martin, R., & Terry, D. J. (2007). Uncertainty and the influence of group norms in the attitude-behaviour relationship. *British Journal of Social Psychology, 46*, 769-792. doi: 10.1348/014466606X164439

Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50. doi: 10.1037//0096-3445.117.1.34

Sober, E. & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA, US: Harvard University Press.

Spears, R., Jetten, J., Scheepers, D., & Cihangir, S. (2009). Creative Distinctiveness: Explaining ingroup bias in minimal groups. In S. Otten, K. Sassenberg & T. Kessler (Eds.), *Intergroup Relations: The Role of Motivation and Emotion*. East Sussex: Psychology Press.

Sperber, D. & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition, 85*, 277-290. doi: http://dx.doi.org/10.1016/S0010-0277(02)00125-7

Stangor, C. & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111*, 42-61. doi: 10.1037/0033-2909.111.1.42

Staub, E. (1990). Moral exclusion, personal goal theory, and extreme destructiveness. *Journal of Social Issues, 46*, 47-64. doi: 10.1111/j.1540-4560.1990.tb00271.x

Steffens, N. K., Haslam, S. A., Ryan, M. K., & Kessler, T. (2013). Leader performance and prototypicality: Their inter-relationship and impact on leaders' identity entrepreneurship. *European Journal of Social Psychology, 43*, 606-613. doi: 10.1002/ejsp.1985

Stellmacher, J. & Petzel, T. (2005). Authoritarianism as a Group Phenomenon. *Political Psychology, 26*, 245-274. doi: 10.1111/j.1467-9221.2005.00417.x

Stürmer, S., Snyder, M., & Omoto, A. M. (2005). Prosocial emotions and helping: The moderating role of group membership. *Journal of Personality and Social Psychology, 88*, 532-546. doi: 10.1037/0022-3514.88.3.532

Sumner, W. (1906). *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores and Morals*. Oxford, England: Ginn.

Suzuki, A. & Suga, S. (2010). Enhanced memory for the wolf in sheep’s clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition, 117*, 224-229. doi: 10.1016/j.cognition.2010.08.004

Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology, 53*, 1038-1051. doi: 10.1037/0022-3514.53.6.1038

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

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations, 33*(47), 74. Retrieved from https://www.researchgate.net/publication/226768898_An_Integrative_Theory_of_Intergroup_Conflict

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual review of psychology, 58*, 345-372. doi: 10.1146/annurev.psych.56.091103.070145

Tangney, J. P., Wagner, P., Fletcher, C., & Gramzow, R. (1992). Shamed into anger? The relation of shame and guilt to anger and self-reported aggression. *Journal of Personality and Social Psychology, 62*, 669-675. doi: 10.1037/0022-3514.62.4.669

Tarrant, M., Branscombe, N. R., Warner, R. H., & Weston, D. (2012). Social identity and perceptions of torture: It's moral when we do it. *Journal of Experimental Social Psychology, 48*, 513-518. doi: 10.1016/j.jesp.2011.10.017

Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for outgroup members. *British Journal of Social Psychology, 48*, 427-446. doi: 10.1348/014466608X373589

Taylor, T. S. & Hosch, H. M. (2004). An examination of jury verdicts for evidence of a similarity-leniency effect, an out-group punitiveness effect or a black sheep effect. *Law and Human Behavior, 28*, 587-598. doi: 10.1023/B:LAHU.0000046436.36228.71

Terry, D. J. & Hogg, M. A. (1996). Group norms and the attitude–behavior relationship: A role for group identification. *Personality and Social Psychology Bulletin, 22*, 776-793. doi: 10.1177/0146167296228002

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*, 451-471. doi: 10.1037/0033-295X.109.3.451

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of personality and social psychology, 78*, 853. doi: 10.1037//0022-3514.78.5.853

Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., . . . Rescober, P. (2007). People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology, 43*, 195-209. doi: 10.1016/j.jesp.2006.02.009

Tomasello, M. & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64*, 231-255. doi: 10.1146/annurev-psych-113011-143812

Tooby, J. & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human morality and sociality: Evolutionary and comparative perspectives* (pp. 91-234). New York: Palgrave MacMillan.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly review of biology*, 35-57. doi: 10.1086/406755

Turiel, E. (1983). *The development of social knowledge: Morality and convention*: Cambridge University Press.

Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions, and vagaries. In J. Kagan, S. Lamb, J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155-243). Chicago, IL, US: University of Chicago Press.

Turner, J. C. (1982). Towards a cognitive redefinition of the social group. In: H. Tajfel (Ed.), *Social identity and intergroup relations* (pp. 15-40). New Yrok, NY: Cambridge University Press.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In E.J. Lawler (Ed.), *Advances in group processes* (Vol. 2, pp. 77-122). Greenwich, CT: JAI.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Cambridge, MA, US: Basil Blackwell.

Tyler, T. R. & Blader, S. L. (2000). *Cooperation in groups: Procedural justice, social identity, and behavioral engagement*. New York, NY US: Psychology Press.

Tyler, T. R. & Boeckmann, R. J. (1997). Three strikes and you are out, but why? The psychology of public support for punishing rule breakers. *Law & Society Review, 31*, 237-265. doi: 10.2307/3053926

Van Bavel, J. J. & Cunningham, W. A. (2012). A social identity approach to person memory: Group membership, collective identification, and social role shape attention and memory. *Personality and Social Psychology Bulletin, 38*, 1566-1578. doi: 10.1177/0146167212455829

Van de Wetering, S. (1996). Authoritarianism as a group-level adaptation in humans. *Behavioral and Brain Sciences, 19*, 780-781. doi: 10.1017/S0140525X00044034 van Prooijen, J.-W. (2006). Retributive Reactions to Suspected Offenders: The Importance of Social Categorizations and Guilt Probability. *Personality and Social Psychology Bulletin, 32*, 715-726. doi: 10.1177/0146167205284964

van Zomeren, M., Spears, R., Fischer, A. H., & Leach, C. W. (2004). Put Your Money Where Your Mouth Is! Explaining Collective Action Tendencies Through Group-Based Anger and Group Efficacy. *Journal of Personality and Social Psychology, 87*, 649-664. doi: 10.1037/0022-3514.87.5.649

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior, 28*, 260-271. doi: 10.1016/j.evolhumbehav.2007.04.006

Vidmar, N. (2001). Retribution and revenge. In J. Sanders, & V. L. Hamilton (Eds.), *Handbook of justice research in law* (pp. 31-63). Dordrecht, Netherlands: Kluwer Academic Publishers.

Vitaglione, G. D. & Barnett, M. A. (2003). Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. *Motivation and Emotion, 27*, 301-324. doi: 10.1023/A:1026231622102

Volstorf, J., Rieskamp, J., & Stevens, J. R. (2011). The good, the bad, and the rare: Memory for partners in social interactions. *PLoS ONE, 6*. doi: 10.1371/journal.pone.0018945

Walker, I. & Pettigrew, T. F. (1984). Relative deprivation theory: An overview and conceptual critique. *British Journal of Social Psychology, 23*, 301-310. doi: 10.1111/j.2044-8309.1984.tb00645.x

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences, 14*, 383-388. doi: 10.1016/j.tics.2010.05.006

Wenzel, M. (2004). Social Identification As a Determinant of Concerns About Individual-, Group-, and Inclusive-Level Justice. *Social Psychology Quarterly, 67*, 70-87. doi: 10.1177/019027250406700107

Wenzel, M. (2009). Social identity and justice: Implications for intergroup relations. In S. Otten, K. Sassenberg & T. Kessler (Eds.), *Intergroup relations: The role of motivation and emotion* (pp. 61-79). New York, NY US: Psychology Press.

Wenzel, M., Okimoto, T. G., Feather, N. T., & Platow, M. J. (2008). Retributive and restorative justice. *Law and Human Behavior, 32*, 375-389. doi: 10.1007/s10979-007-9116-6

Wiese, H., Komes, J., & Schweinberger, S. R. (2013). Ageing faces in ageing minds: A review on the own-age bias in face recognition. *Visual Cognition, 21*, 1337-1363. doi: 10.1080/13506285.2013.823139

Wilkowski, B. M. & Chai, C. A. (2012). Explicit person memories constrain the indirect reciprocation of prosocial acts. *Journal of Experimental Social Psychology, 48*, 1037-1046. doi: 10.1016/j.jesp.2012.04.005

Wilson, D. S., Ostrom, E., & Cox, M. E. (2013). Generalizing the core design principles for the efficacy of groups. *Journal of Economic Behavior & Organization, 90*, S21-S32. doi: 10.1016/j.jebo.2012.12.010

Wilson, D. S. & Wilson, E. O. (2010). Evolution 'for the good of the group'. In P. W. Sherman, J. Alcock, P. W. Sherman & J. Alcock (Eds.), *Exploring animal behavior: Readings from the American Scientist (5th ed)* (pp. 344-353). Sunderland, MA, US: Sinauer Associates.

Wiltermuth, S. S. & Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*, 1-5. doi: 10.1111/j.1467-9280.2008.02253.x

Wit, A. P. & Kerr, N. L. (2002). 'Me versus just us versus us all' categorization and cooperation in nested social dilemmas. *Journal of Personality and Social Psychology, 83*, 616-637. doi: 10.1037/0022-3514.83.3.616

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition, 100*, 283-301. doi: 10.1016/j.cognition.2005.05.002

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*, 110-116. doi: 10.1037/0022-3514.51.1.110

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 106*, 11520-11523. doi: 10.1073/pnas.0900636106

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. *Advances in group processes, 16*, 161-197.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior, 24*, 290-301. doi: 10.1016/S1090-5138(03)00035-7

Young, L. & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120*, 202-214. doi: http://dx.doi.org/10.1016/j.cognition.2011.04.005

Yzerbyt, V., Dumont, M., Gordijn, E., & Wigboldus, D. H. J. (2002). Intergroup Emotions and Self-Categorization. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 67 - 88). New York: Psychology Press.

Yzerbyt, V., Dumont, M., Wigboldus, D., & Gordijn, E. (2003). I feel for us: The impact of categorization and identification on emotions and action tendencies. *British Journal of Social Psychology, 42*, 533-549. doi: 10.1348/014466603322595266

Appendix

Appendix A.

Target ratings in terms of fairness (Study 1, Section 2), and likability in encoding and test phase (Study 2 and Study 3, Section 2) as a function of target type

	Study 1		Study 2		Study 3	
	Ingroup	Outgroup	Ingroup	Outgroup	Ingroup	Outgroup
	M (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*	*M* (*SD)*
Uncooperative/ Cheater	1.87 (.86)	1.85 (.85)	2.41 (.72)	2.41 (.66)	1.90 (.80)	1.88 (.80)
Neutral	-	-	3.77 (.58)	3.70 (.55)	3.93 (.72)	3.82 (.66)
Cooperative/ Trustworthy	5.16 (.76)	5.16 (.74)	4.44 (.77)	4.40 (.71)	4.86 (.67)	4.89 (.71)
Uncooperative/ Cheater	-	-	3.37 (.59)	3.39 (.55)	3.33 (.64)	3.26 (.70)
Neutral	-	-	3.59 (.56)	3.56 (.53)	3.43 (.56)	3.47 (.61)
Cooperative/ Trustworthy	-	-	3.64 (.64)	3.54 (.51)	3.53 (.52)	3.57 (.63)

Note: M=*mean,* SD= *standard deviation; all variables varied from* 1= *not at all to* 7= *very much*

Appendix B.

Means and standard deviations of pre-test evaluations (*N*= 76, 50 female) of target behavior (cheating. irrelevant and trustworthiness descriptions) as used in the Study 3 (Section 2)

	Plausibility	Valence	Category
	M (*SD)*	*M* (*SD)*	*M* (*SD)*
Cheating behavior	4.86 (.74)	1.41 (.43)	1.13 (.14)
Neutral behavior	4.47 (.74)	3.98 (.39)	2.05 (.12)
Trustworthy behavior	4.81 (.66)	5.67 (.36)	2.74 (.27)

Rating scales for "plausibility" ranged from 1= *not at all to* 6= *very much; rating scale for "valence" ranged from* 1= *negative to* 6= *positive; categories:* 1= *cheating,* 2= *neutral,* 3= *trustworthy*

Appendix C.

Scenario description as included in study material of Study 1 (Section 3)

Introduction:
Lies Dir bitte den folgenden Text aufmerksam durch:

Stell Dir vor du trainierst gerade zusammen mit deiner Mannschaft. Ihr habt einen guten Teamgeist und achtet aufeinander. Mit eurem Trainer habt ihr beim letzten Training eine neue Taktik besprochen, die ihr jetzt zusammen einübt. Spieler A, B und C spielen auf der gleichen Position und üben jetzt die gleiche Technik ein. Spieler B ist der beste Spieler auf dieser Position und wird deshalb meistens bei Spielen eingesetzt. Zuerst setzten Spieler A und Spieler B die Technik wie besprochen um. Nun sind Spieler B und Spieler C an der Reihe.

Intentions:
Beim Durchführen der Technik stößt Spieler C absichtlich Spieler B, um sich dadurch einen Vorteil in der Teamaufstellung zu verschaffen.

No intentions:
Beim Durchführen der Technik stößt Spieler C aus Versehen mit Spieler B zusammen.

Harm:
Spieler B verletzt sich dabei. Damit wird er, als ein wertvoller Spieler für das Team in nächster Zeit beim Training und auch bei Punktspielen ausfallen.

No harm:
Spieler B verletzt sich dabei nicht und kann das Training fortsetzen.

Appendix D.

Scenario description as included in study material of Study 2&3 (Section 3)

Introduction:
*In folgendem Fragebogen geht es um die Beurteilung von Situationen. Unsere Hauptfragestellung bezieht sich auf Gefühle, die durch eine bestimmte Schilderung ausgelöst werden. Deshalb haben wir eine Geschichte auf unterschiedliche Weise dargestellt. Alle Proband*innen bekommen dieselbe Geschichte in unterschiedlichen Textformen zur Beurteilung.*

Gib bitte deine spontane Reaktion an, nachdem du den Text sorgfältig *gelesen hast.*

Ulrich H. wird von seinen Mitmenschen als unauffällig und angenehm beschrieben. Er wohnt in einer Zweizimmerwohnung in einem Mietshaus. Er ist Schulbusfahrer für die örtliche Grundschule. Diesen Beruf führt er schon lange aus, obwohl er täglich eine schwierige und gefährliche Strecke zurücklegen muss. Ulrich H. fährt sicher und mit Bedacht.

Intention:
Jeden Tag stellt er sich vor, wie die Kinder schreien und weinen, während der Bus eine tiefe Klippe herunterstürzt. Ulrich H. steigert sich lange in seine gewalttätigen Vorstellungen hinein, und denkt sich aus, wie der Bus mit Kindern an den Felsen zerschellt. Irgendwann beschließt Ulrich H. seine Fantasie auszuleben. Er wird immer entschlossener in seinem Vorhaben. An einem Tag im Oktober, an dem die Kinder gerade mit dem neuen Schuljahr begonnen haben, ist der Bus besonders voll besetzt. Ulrich H. trifft die letzten Vorbereitungen. Damit sein Vorhaben nicht mehr verhindert werden kann, lockert Ulrich H. die Bremsen des Busses, so dass sie bei hoher Geschwindigkeit den Bus nicht mehr zum Stehen bringen.

No intention:
Ulrich H. hat große Angst, dass eines Tages ein Unfall passiert und die Kinder zu Schaden kommen. An einem Tag im Oktober, an dem die Kinder gerade mit dem neuen Schuljahr begonnen haben, ist der Bus besonders voll besetzt. Es ist unmöglich für Ulrich H. zu bemerken, dass sich die Bremsen des Busses gelockert haben, so dass sie bei hoher Geschwindigkeit den Bus nicht mehr zum Stehen bringen.

Harm:
Am nächsten Morgen steht Ulrich H. auf, nimmt sein Frühstück zu sich und verlässt das Haus. Eine Stunde später sterben die 48 Kinder, die den Bus zur Schule nehmen. Der Schulbus stürzt in einer Kurve von einer Klippe ca. 40 Meter in ein Flussbett. Ulrich H.s Leiche wird zerquetscht hinter dem Lenkrad gefunden, als der Bus geborgen wird.

No harm:
Am nächsten Morgen steht Ulrich H. auf, nimmt sein Frühstück zu sich und erleidet noch am Tisch einen Herzinfarkt. Kurz darauf stirbt Ulrich H. Die 48 Kinder, die den Bus zur Schule nehmen, erfahren nie von den lockeren Bremsen und Ulrich H.s Absichten. Der Schulbus wird nach seinem Tod in einer Werkstatt generalüberholt und die Bremsen ausgewechselt.

Appendix E.

Correlations between emotional adjectives and punishment Study 2 (Section 3)

Measure		2	3	4	5	6
1.	Empathy with victims	.55*	.35*	.45*	-.15	.28*
2.	Anger		.67*	.56*	-.12	.50*
3.	Fear			.37*	.09	.28*
4.	Sadness				-.36*	.04
5.	Content					-.06
6.	Willingness to punish					

*p< .01, ^{+}p< .05

Appendix F.

Correlations between emotional adjectives and punishment Study 3 (Section 3)

Measure		2	3	4	5	6
1.	Empathy with victims	.33*	.29*	.36*	-.36*	.28*
2.	Anger		.46*	.33*	-.22*	.64*
3.	Fear			.64*	-.17^{+}	.20*
4.	Sadness				-.16^{+}	-.04
5.	Content					-.11
6.	Willingness to punish					

*p< .01, ^{+}p< .05

Appendix G.

Results of original ANOVA of all emotional reactions in Study 3 (Section 3), not controlled for plausibility

Effect	Dependent variable	*F*(1,205)	*p*	η^2
Intention	anger	99.95	.00	.32
	fear	6.02	.02	.03
	sadness	1.30	.26	.01
	content	2.48	.12	.01
	willingness to punish	300.73	.00	.59
Harm	anger	1.64	.20	.01
	fear	3.80	.05	.02
	sadness	11.46	.00	.05
	content	19.18	.00	.08
	willingness to punish	2.15	.144	.004
Intention * Harm	anger	1.13	.29	.003
	fear	.003	.96	<.001
	sadness	.30	.58	.001
	content	5.37	.02	.02
	willingness to punish	.34	.56	<.001

Appendix H.

PCA factor loadings of emotional adjectives after oblimin rotation (converged in 17 iterations) displayed by the pattern matrix in Study 3 (Section 4)

component	1	2	3	4	5
	(anger)	(irritation)	(disinterest)	(content)	
Eigenvalue	6.31	2.01	1.70	1.15	1.02
verärgert	**.748**				
belustigt					-.886
überrascht		**.600**			-.507
empört	**.795**				
zerstreut		**.851**			
entrüstet	**.823**				
verwirrt		**.768**			
aufgebracht	**.836**				
zufrieden				**.855**	
wütend	**.852**				
kühl			**.765**		
beruhigt				**.792**	
gelangweilt			**.738**		
schockiert	.540	.373			
unbeteiligt			**.659**		
angewidert	**.822**				
verdrossen	**.701**				

Appendix I.

Scenario description as included in study material of Study 4 (Section 4)

Introduction:

*In folgendem Fragebogen geht es um die Betrachtung von Zeitungsartikeln. Unsere Hauptfragestellung bezieht sich auf Gefühle. die durch die Art der Meldung ausgelöst werden: Deshalb haben wir eine Meldung auf unterschiedliche Weise dargestell.t Deswegen bekommen alle Proband*innen dieselbe Meldung in unterschiedlichen Textformen zur Beurteilung. Wir möchten von dir wissen: Welche Emotion löst der Zeitungsartikel bei dir aus?*

Gib bitte deine spontane Reaktion an nachdem du den Text sorgfältig gelesen hast.

Torture:

Heute verstarb ein Armeeangehöriger der westlichen Bodentruppen/ des Islamischen Staats (IS) an den Folgen brutaler Verhörmethoden in einem Hochsicherheitsgefängnis in Europa/ in dem vom IS besetzten Gebiet. Der Verhaftete stand unter Verdacht die innereuropäische Sicherheit/ Sicherheit des IS zu gefährden. Um an Informationen zu gelangen wurde er durch Geheimdienstmitarbeiter über mehrere Tage körperlicher Misshandlung unterzogen. Dabei erlitt er schwerwiegende gesundheitliche Schäden und ist noch im Gefängnis verstorben.

Court Trial:

Heute begann das Gerichtsverfahren gegen einen Armeeangehörigen der westlichen Bodentruppen/ des Islamischen Staats (IS). der in einem Hochsicherheitsgefängnis in Europa/ in dem vom IS besetzten Gebiet untergebracht ist. Der Verhaftete steht unter Verdacht die innereuropäische Sicherheit/ Sicherheit des IS zu gefährden. Um seine Aussage aufzunehmen wurde er durch Geheimdienstmitarbeiter über mehrere Tage befragt. Das Verfahren wird von Menschenrechtsorganisationen beobachtet und begleitet. Die Beweislage wird von einem europäischen/ islamischen Gericht begutachtet und die Verhandlungen werden noch einige Wochen andauern.

Appendix J.

Correlations between emotional adjectives and punishment Study 4 (Section 4)

Measure		2	3	4	5	6	7
1.	Plausibility	.07	-.20^{+}	.13	.05	.03	.05
2.	Anger		.62*	.67*	-.24^{+}	.62*	.53*
3.	Shame			.37*	-.08	.48	.36*
4.	Fear				.02	.47*	.40*
5.	Content					-.25*	-.30*
6.	Willingness to punish						.56*
7.	Empathy with victims						

**p*< .01. $^{+}$*p*< .05

Appendix K.

Results of original ANOVA of all emotional reactions in Study 4 (Section 4). not controlled for plausibility

Effect	Dependent variable	*F*(1.94)	*p*	η^2
Treatment	anger	68.66	<.001	0.40
	fear	13.51	<.001	0.12
	content	11.92	.001	0.11
	shame	11.53	.001	0.10
	Willingness to punish	58.16	<.001	.12
Perpetrator Group Membership	outrage	.48	.49	0.00
	fear	.18	.67	0.00
	content	1.30	.26	0.01
	shame	.10	.75	0.00
	Willingness to punish	.86	.36	.002
Victim Group Membership	outrage	.72	.40	0.00
	fear	.001	.97	0.00
	content	1.34	.25	0.01
	shame	.007	.93	0.00
	Willingness to punish	3.08	.08	.02
Treatment * Perpetrator Group Membership	outrage	6.54	.01	0.04
	fear	.10	.75	0.00
	content	1.86	.18	0.02
	shame	9.61	.003	0.08
	Willingness to punish	7.08	.006	<.001
Treatment * Victim Group Membership	outrage	.95	.33	.01
	fear	.007	.93	<.001
	content	.14	.71	<.001
	shame	.54	.47	<.001
	Willingness to punish	.006	.94	<.001
Perpetrator * Victim Group Membership	outrage	.35	.55	<.001
	fear	.22	.64	<.001
	content	1.14	.29	.01
	shame	.54	.47	<.001
	Willingness to punish	.16	.69	<.001
Treatment * Perpetrator * Victim Group Membership	outrage	1.62	.21	.01
	fear	1.44	.23	.01
	content	.36	.55	.00
	shame	.31	.58	<.001
	Willingness to punish	<.001	1.00	<.001

Acknowledgements

My special thanks go to those people who behaved favorably towards me in the last years and thereby contributed in maintaining our positive and cooperative relationships. First of all I want to thank my supervisors, Thomas Kessler and Franz J. Neyer, for the good cooperation and their support. I want to thank Raoul Bell for counseling me on a frequent basis. I am grateful to my colleagues from the Research Unit Person Perception and the Departments of Social and Personality Psychology for their comments and advices. Especially, I want to thank my colleagues and friends Mareike Grotheer, Jutta Proch, Pamela Pensini, Daniel Seewald, and Elisabeth Schubach for providing me with academic and social support.

I would additionally like to thank the Deutsche Research Foundation (DFG) for their financial support. Thanks go to my hard-working student assistants for their help with data collection: Eva Junge, Maria Besselmann, Vera Seyffert, Selina Schmid, Blume Logar, Sarah Matthias, und Rowi Bender.

Last, but not least, of course I want to thank my family, Lydia, Georg and Hanna Hechler for their encouragement and warm support; and Leonie, Jessica, Timmi, and Daniel for patience, love, and beer.

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena bekannt ist. Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe sowie nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, sind als solche kenntlich gemacht.

Ich habe die vorliegende Dissertation nicht, auch nicht in Teilen, für eine Staatliche Prüfung bzw. als Dissertationsschrift bei einer anderen Hochschule bzw. Fakultät eingereicht.

Ich versichere, dass ich nach bestem Wissen die Wahrheit gesagt und nichts verschwiegen habe.

____

Stefanie Hechler

Curriculum Vitae

Name	Stefanie Hechler
Address	Sophienstraße 17 07743 Jena
E-Mail	stefanie.hechler@uni-jena.de
Date of birth	13.05.1985
Place of birth	Nürnberg

Education

03/2016 - now	PhD student and research assistant in the Person Perception Research Unit at the University of Jena, Germany (DFG-Grant to Prof. Dr. Thomas Kessler and Prof. Dr. Franz J. Neyer)
10/2015 – 03/2016	Research assistant at the Department of Personality Psychology and Psychological Assessment at the University of Jena, Germany
10/2012 – 09/2015	PhD student and research assistant in the Person Perception Research Unit at the University of Jena, Germany (DFG-Grant to Prof. Dr. Thomas Kessler and Prof. Dr. Franz J. Neyer
09/2005 - 02/2012	Psychology student at the University of Bremen, Germany
08/2007 – 03/2008	Academic semester abroad at the University of Valencia, Spain
09/1995 - 06/ 2005	High school student at the Helene-Lange-Gymnasium Fürth, Germany
08/2001 – 07/2002	Student exchange at the Herman-Norcross High-School

Publications

Hechler, S., Neyer, F. J., & Kessler, T. (accepted). The infamos among us: Enhanced reputational memory for uncooperative ingroup members. *Cognition*.

Seewald, D., Hechler, S., & Kessler, T. (2016). Divorcing the puzzles: When group identities foster in-group cooperation. The Behavioral and Brain Sciences, 39, e23. doi: 10.1017/S0140525X15000539

Kessler, T., Proch, J., Hechler, S., & Nägler, L. A. (2015). Political diversity versus stimuli diversity: Alternative ways to improve social psychological science. Behavioral and Brain Sciences, 38, e148. doi:10.1017/S0140525X14001241

Hechler, S. (2012). Ungerechtigkeit und Protestverhalten. *ZeS Report* (17). Bremen: Zentrum für Sozialpolitik, Universität Bremen.

Stefanie Hechler