Punishing 'Them' Harder?
An Investigation of Group Differences in Reactions to Norm Violation

Dissertation
zur Erlangung des akademischen Grades
doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften
der Friedrich-Schiller-Universität Jena
von Dipl.-Psych. Johann Jacoby
geboren am 23. August 1975 in Düsseldorf

Gutachter

1. PD Dr. Kai Sassenberg

2. PD Dr. Sabine Otten

Prof. Dr. Reinhard Blickhan

Tag des Kolloquiums: 15. September 2006

At the beginning [...], briefly quote or summarize the thoughts of an unknown psychologist or philosopher. This works best if the figure is from the 19th century. If this is not possible, a few phrases (not more than three) from Aristotle will do. (Weiner, 1984, p. 926)

Table Of Contents

1 INTRODUCTION

Punishment for bad deeds is a universal feature of human societies (Vidmar, 2000). While it is legitimate and common across societies that behavior which violates a societal standard is met with the infliction of harm, it is still debated how the appropriate amount of such harm could be both morally justified as well as how lay people (i.e., non-lawyers) determine an amount appropriate for a given transgression. While the latter issue may seem trivial, it is quite important, as law abiding of the constituents of a legal system will be considerably influenced by the extent to which their own sense of an appropriate response to transgressions overlaps with the empirical responses of legal institutions, and thus by their trust that these institutions will uphold justice (Robinson & Darley, 1995).

Empirical research on how lay people determine the appropriate amount of punishment has compiled a host of results consistent with the notion that we recommend punishment proportionate to the severity of a crime (e.g., Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002). According to this research, utilitarian aspects such as rehabilitation or deterrence play a minor role (Carlsmith et al., 2002; Darley, Carlsmith, & Robinson, 2000). Also, it seems that lay people's reactions to criminal offenses (or at least severe deeds deemed wrong) are mediated by a rather diffuse negative affective experience (Carlsmith et al., 2002) which is taken as an indicator of the offenses' severity. This affectively colored process linking information about an offense to psychologically appropriate punishment intensity may thus be vulnerable to negative affect stemming from other sources than information about a deed of which we have just learned. One such source could be an offender's group membership, that is whether he or she belongs to a group the perceiver belongs to him or herself (*ingroup*) or a group of which the perceiver is not a member (*outgroup*). Assuming for now that this group membership has a biasing influence, such a phenomenon would have an

important consequence for intergroup research, research which is – among others – generally concerned with ingroup vs. outgroup distinctions.

Throughout history, distinctions between groups or categories of people have been and are still pervasive (Sidanius & Pratto, 2001). To a large part, intergroup research has been motivated by the desire to understand and explain these differentiations *per se* as well as maltreatment of human beings based on such categorization. Interpersonal violence among individuals is one thing, but the most horrible and terrifying acts against humanity (such as genocide and most prominently the Shoah) have regularly been committed by individuals acting on behalf of a group against other individuals as members of another group (see for example Newman & Erber, 2002). However, despite a long history of research in the intergroup domain, which often focused on the role of mere categorization, it is still somewhat of a riddle to intergroup research why and how the infliction of harm on individual others belonging to 'the others' (outgroup derogation) emerges (Brewer, 1999). Simply put, derogation of relatively content-free outgroups (i.e., minimal categories) is not well understood when it comes to the infliction of harm, or more generally, the allocation of negative resources to members of certain groups (see also Mummendey & Otten, 1998).

The currently reported work aims to contribute to this research in that it identifies a class of contexts where the infliction of harm on individuals by individuals is already relatively common: reactions to norm violation, including punishment tendencies. The term 'norm violation', as used throughout this dissertation, simply and globally denotes behavior commonly believed to be wrong. As mentioned above many of such norm violations are intricately associated with a punishing response. The punishment need not be permanent, it may come in different degrees of severity and it can consist in a variety of procedures applied to those who have done wrong. Invariably however, and even independent of whether such

punishment is intended to hurt the norm violator or to, for example, teach him or her a lesson, it is aversive to the violator: No one likes to be punished, except for a few to which the present research and theorizing is not to be generalized.

These circumstances, under which it is legitimate to treat somebody negatively, provide a promising domain in which to study intergroup behavior. Domains in which intergroup behavior is commonly examined, especially in experimental research, are usually neutral or even positive. A typical experiment randomly assigns participants to a minimal group (e.g., A or B) and then has participants allocate positive points to members of their own and those of another group (*ingroup* and *outgroup*, respectively; such a paradigm is called a *Minimal Group Paradigm*, MGP, Tajfel, Flament, Billig, & Bundy, 1971). In such a situation, there is no reason, let alone legitimacy, to hurt targets belonging to a different group than oneself. Preferential treatment for the own group (ingroup favoritism) on the other hand is less illegitimate and may even be considered normative (Blanz, Mummendey, & Otten, 1997). It seems that outgroup derogation typically does not occur because there is no reason for derogation of any individual, let alone group members. But in situations where negative behavior is legitimate against anyone found guilty and responsible of doing 'something bad', we may find systematic differences in treatment of ingroup versus outgroup members again, providing a step in the direction of an understanding of how outgroup derogation occurs.

How would such a phenomenon come about? Traditional explanations of ingroup favoritism may not be promising, as they predict differential treatment of ingroup versus outgroup targets from that treatment providing a means to increase self-esteem as a member of the ingroup (Tajfel & Turner, 1979; Crocker & Luhtanen, 1990). However, in the case of targets who have violated a strong norm, such a gain in self-esteem is rather unlikely. Therefore, an alternative explanation is proposed.

Ingroups are automatically associated with positive affect (e.g., Otten & Wentura, 1999), whereas outgroups are not specifically associated with any valence. Ingroup members, who have committed a transgression therefore could be spared of the full intensity of a negative affective reaction. They would thus profit from a buffer which outgroup targets lack. In neutral or positive contexts, the differential association with positive affect leads to merely preferential treatment of ingroup members compared to outgroup members. But no harm as such is inflicted on members of an outgroup (they are simply *not treated as positively* as ingroup members). In the context of norm-violation, and thus events associated with negative affect, this buffer for the ingroup could however turn into *more negative treatment* of an outgroup and its members.

This hypothesis runs counter to common phenomena of a *Black Sheep Effect* (henceforth referred to as BSE, Marques & Paez, 1994). This effect consists in relative derogation of an ingroup member behaving in a socially undesirable fashion compared to an outgroup member behaving in the same way. Thus, the above prediction for an intergroup bias in favor of the ingroup will have to face the seemingly contradicting theory and empirical evidence for the BSE and it will need to advance an argument as to why the opposite pattern (i.e., outgroup derogation) will occur. The argument advanced here, and to be discussed in more detail in Chapter 5, is that the occurrences of the BSE in the published literature have mostly been observed in the context of *mildly* norm-violating behaviors, for which there exists a descriptive or prescriptive, but group-specific norm. With more severe offenses and transgressions against overarching, global norms, the processes leading to the BSE (see below) could lose in importance. In this case, the process hypothesized above, protecting ingroup versus outgroup members from strong negative reactions, may lead to less negative treatment of an ingroup target.

The present dissertation develops the foregoing ideas and reports results from empirical tests of the resulting predictions. First, punishment reactions to norm violations in general and crimes in particular are discussed and a model of how reactions to such norm violations are formed will be advanced (Chapters 2 and 3). This discussion will be followed by a sketch of the problem of intergroup research regarding the understanding of harmful treatment of outgroups and demonstrate why the norm violation context provides a promising ground for progress in this area (Chapter 4). The discussion of the ideas outlined until then in relation with research on the BSE will take up most of Chapter 5. Chapter 6 will summarize three hypotheses concerning the reactions to norm violation in group contexts. The empirical chapters, following a general description of the methodological approach chosen here (Chapter 7 and 8), then describe studies and meta-analyses regarding each of the three hypotheses (Chapters 9 through 15). A general discussion (Chapter 16) with a review of the findings and their integration will sum up the content of this dissertation.

2 Determinants of Punishment

Punishment for violations of prescriptive norms may be considered a human invariant (Vidmar, 2000; Carlsmith, 2006). All societies and cultures have developed some sort of institutionalized procedure inflicting harm on perpetrators of acts deemed unacceptable, that is, crimes. Some scholars even argue that punishment, particularly punishment which raises considerable costs to the punisher – so called *altruistic punishment –,* is a phylogenetically evolved mechanism which has emerged because it provided an advantage to groups, communities, and societies who practiced it (de Quervain, Fischbacher, Treyer, Schellhammer, Schnyder, Buck, & Fehr, 2004; Fehr & Fischbacher, 2003; Fehr, Fischbacher, & Gächter, 2002). Punishment impulses and procedures can safely be considered ubiquitous, widespread and one of the most fundamental features across cultures on various levels of civilization.

However, in many disciplines such as philosophy, law, sociology, political science, there has been considerable debate as to *how* punishment is assigned or its infliction may be justified. Two distinct broad classes of approaches have emerged.

2.1 Just Deserts as a Determinant of Punishment

The first approach, commonly referred to as the *just deserts* or *deservingness* perspective, has a long history and can be traced back to authoritative philosophers such as Kant (1797/1990, see, e.g., Carlsmith, 2006; Carlsmith et al., 2002). This perspective essentially holds that punishment should be proportionate to the wrongdoing as its main purpose is 'righting of the scales of justice' (Sargent, 2004, p. 1485) that the transgression has tipped to imbalance (see also Carlsmith et al., 2002). Aspects that play a role in the assignment of appropriate intensity of punishment in the just deserts perspective are

- The *degree of harm* done (offenders causing more harm should receive more severe punishment)
- The perpetrator's *intentions* (offenders causing harm inadvertently should be punished less than those committing their transgression whole-heartedly)
- *Mitigating circumstances* (e.g., offenders acting under high pressure and threat to their life, physical or psychological integrity should be punished less severely than those acting freely)

An important consequence of this approach is that punishment should be proportionate to the harm done and the internal wickedness ('innere Bösartigkeit', Kant, 1797/1990, p. 195) of the perpetrator and his deed. Punitive responses essentially secure retribution. Considerations of future recidivation, chances of rehabilitation, or specific and general deterrence do not play a role in the pure form of this approach sketched here. These latter issues figure prominently in the other perspective on punishment to be discussed next.

2.2 Utilitarian Punishment Motives

The opposing approach, the *utilitarian* or *consequentialist* philosophy of punishment (e.g., Bentham, 1843/1962) fundamentally insists that punishment exclusively serve the end to minimize future incidences of transgressions and therefore should be meted out according to the principle of maximum utility. Exemplary aspects determining severity (and also the nature) of punishment according to this perspective are

- *The likelihood of recidivation*. The more likely the offender is deemed to commit the same or a similar crime again, the more severe punishment should be. This follows from the general idea that punishment should deter the offender from repeating norm-violating behavior as opposed to third parties (*special deterrence*). According to this

view, higher punishment increases the cost of future transgressions, and the more likely the offender is to re-offend in the first place, the higher the cost should be.

- *The base rate of the crime as well as rate of detection and conviction in society*. Third parties witnessing the conviction and punishment of an offender should also perceive higher cost for a transgression so that they are also deterred from committing it (*general deterrence)*. The higher the initial likelihood of a crime being committed in society as a whole, and the higher the possibility to 'get away with it', the more severe punishment should be.
- *The likelihood of rehabilitation*. Offenders who are more likely to desist from a crime already after a mild punishment should be punished less than those for whom prospects of rehabilitation are lesser. The nature of punishment could even consist in educational measures if the likelihood of rehabilitation and penitence is considered high. A special case is the situation where chances of rehabilitation are considered minimal, such that permanent protection of society from the offender by permanently holding him or her in a penitentiary or health care institution is considered appropriate (*incapacitation*).

2.3 Two Philosophies - Two Psychological Processes

These two perspectives rarely figure in a pure form in modern penal codes or even in individual attitudes. However, they point to two fundamentally different psychological processes of arriving at an adequate measure of punishment for a given transgression. While the utilitarian perspective prescribes a thorough integration of facts and projections into the future not directly related to the actual transgression to be avenged (i.e., probabilities), the just deserts approach considers the deed itself and its severity. Thus the former approach essentially requires a relatively cold processing of facts and may be considered 'rule-based'

(Nichols & Mallon, 2006). The question arises how severity is determined so that the appropriate amount of punishment in terms of proportionality to the wrongdoing may be inflicted (as in punishment from a just deserts perspective).

It is proposed here that such an assessment is made on the basis of the intensity of an initial negative affective reaction we experience upon learning about a crime. Of course, the assessment of mitigating circumstances or the perpetrator's intentions do require processing of information over and above the mere severity (i.e., intentionality, mitigating circumstances establishing lessened responsibility, see p. 13). Lay persons, whose punitive reactions are at focus here, however rarely have this kind of information. I argue that when the offender's responsibility is established (i.e., no doubts about his or her intention to commit the crime and the freedom to do so are raised), the first reactions to a crime will be very basic and affective in nature. This affective reaction is the main determinant of punishment severity found appropriate by lay persons. Presently, it is assumed and argued that utilitarian concerns figure relatively unimportant in punishment recommendation based on empirical evidence to be discussed next.

The importance of affective reactions of moral judgments has been highlighted by Haidt (2001). In his *social intuitionist model* (SIM) he argues that moral judgments generally start with a 'gut feeling' and verbalized rational arguments are essentially post-hoc justifications of these feelings. According to the model, these gut feelings basically reflect an inevitable and self-evident truth that a certain action is good or bad and thereby acceptable or unacceptable. As transgressions that are in principle punishable most often constitute immoral behaviors whose condemnation has found their way into penal codes or accepted social practice (e.g., the educational domain), it seems plausible, following Haidt (2001), that judgment about them is mediated by negative affect experienced upon learning of them. In line with Haidt (2001) it

is presently believed that the judgment of most transgressions we ordinarily encounter are based on learned evaluative associations. This is all the more plausible considering the strength and relative ease with which we make moral judgments in everyday life.

Trafimow, Bromsgard, Finlay and Ketelaar (2005) have argued that the violation of a perfect duty arouses negative affect which in turn gives greater weight to such violations in person evaluation than violation of imperfect duties, which do not arouse such affect. In line with a distinction made by Kant (1797/1990), perfect duties ('vollkommene Pflichten', or 'Rechtspflichten', Kant, 1797/1990, pp. 77ff.) are taken by Trafimow et al. (2005) as obligations which may not be violated without a condemnation of those committing the violation by the social environment. Imperfect duties ('unvollkommene Pflichten' or 'Tugendpflichten', Kant, 1797/1990, pp. 77ff.) on the other hand comprise desired actions whose omission however is not considered as a transgression but more as a failure to perform maximally desirable behavior. Imperfect duties are not in principle punishable. The violation of perfect duties (as which the adherence to positive law must be taken, at least in modern, constitutional states) is hypothesized by Trafimow et al. (2005)to carry negative affect which leads to its higher weighting in attribution processes and therefore in evaluation of the transgressor compared to the violation of an imperfect duty. Empirical results reported by Trafimow et al. (2005) clearly show that violations of perfect duties are distinctively associated with negative affect.

Finally, research by Carlsmith and colleagues (Carlsmith et al., 2002; Carlsmith, 2006, Darley et al., 2000; see also Darley & Pittman, 2003) has consistently shown that the intensity of recommended punishment is much less influenced by information relating to the aspects pertinent in the utilitarian punishment perspective (incapacitation, Darley et al., 2000; deterrence, Carlsmith et al., 2002) than by the degree of harm done. They conclude

convincingly that lay people's punishment recommendations are mainly made from a just deserts perspective (see also McFatter, 1978; McFatter, 1982; Rucker, Polifroni, Tetlock, & Scott, 2004). Moreover, Carlsmith and colleagues (Carlsmith et al., 2002) found evidence that the relationship between the severity of a deed and punishment intensity is mediated by moral outrage, an emotional reaction related to anger. Weiner (1995; see also Weiner, Graham, & Reyna, 1997) has similarly proposed that reactions to norm violations (specifically punishment impulses) are mediated by the emotion of anger (which comprises negative affect, Frijda, Kuipers, & ter Schure, 1989).

Thus, in sum, there is good reason to assume that reactions to norm violations which are in principle punishable, specifically recommended punishment intensity, are mainly driven by negative affect aroused by the behavior. Aspects relevant to a just deserts or deservingness perspective seem to be most pertinent to lay people's punishment recommendations. They arouse moral outrage, a negative affective experience whose intensity is taken as a subjective indicator of appropriate punishment intensity.

3 Affect - A Possible Inadmissible Influence on Punishment

If punishment recommendations by lay people are indeed mainly driven by a negative affective reaction to it, as argued in the preceding chapter, then affective experiences which are not directly connected to a crime could play an important role in punishment recommendation. Such 'contamination' of judgment processes, especially under uncertainty, have been repeatedly theoretically formulated and empirically found (see Forgas, 1995; Schwarz & Clore, 1983). For example, van den Bos (2003) found that participants' judgments about the fairness of their treatment in the laboratory was influenced by a mood induction unrelated to that procedure proper. Lerner, Goldberg, and Tetlock (1998) have found similar carry-over effects for anger aroused in a different domain which increased punitiveness.

However, such influences of affect on punishment recommendations could also stem from the person of the norm violator him or herself[1]. If a target is evaluated rather positively in a neutral situation, then the positive affect associated with that target influences judgments in valence-laden situations. Specifically, diffuse positive affect associated with a target could buffer the negative affective experience leading to recommended punishment intensity. Such a target would then receive a milder punishment recommendation[2] than another, to which an initial affective reaction is not as positive or neutral. Note that such an influence is inconsistent with the central principle of modern law according to which similar crimes should be punished by similar punishment. Features which are unrelated to the crime and its

1 The distinction between affect elicited by the target person of a judgment itself on one hand, and that evoked by external factors, but possibly influencing judgment of a target on the other is similar to that between integral versus incidental affect, respectively, made by Bodenhausen (1993, Bodenhausen, Mussweiler, Gabriel, & Moreno, 2001). But this latter distinction focuses on chronic emotional reactions, accompanied by specific representations of the target, as well as affect arising within concrete personal interactions with the target. Presently however, the focus is on diffuse general positive versus negative affect independent of representations and interactions (see Chapter 20 below).

2 Assuming that punishment (i.e., actual harmful treatment of an offender) is a function of the amount found appropriate by perceivers, then such a difference would also extend to behavior. Presently, the focus is however on the amount considered appropriate by a perceiver who does not necessarily carry out the punishment of the recommended intensity.

circumstances (such as personal liking for the offender) should not have any weight in the assignment of punishment.

4 Inadmissible Influence in Intergroup Settings

If indeed, initial positive affect associated with a target turns out to determine recommended punishment severity, then this could have important implications for intergroup research and theorizing. Specifically, it offers a possibility to pinpoint situations in which a phenomenon generally referred to as the *positive-negative asymmetry in social discrimination* (Mummendey & Otten, 1998) may *not* occur. This asymmetry consists in the considerable decrease or even elimination of bias in favor of an ingroup if resources to be allocated to an ingroup versus an outgroup are negative rather than positive.

4.1 Intergroup Bias and the Positive-Negative Asymmetry

Intergroup research can rely on a quite stable phenomenon: ingroup favoritism (Hewstone, Rubin, & Willis, 2002). It consists in the general preference of groups we belong to (i.e., ingroups) as opposed to those that we do not belong to (outgroups, for the classical empirical finding using the MGP, see Tajfel et al., 1971). This phenomenon most importantly occurs even for arbitrary and ad-hoc group categorizations without any content of group membership except the random assignment to one of two groups. However, ingroup preference is mainly expressed on positive dimensions (e.g., allocation of goods or points, evaluations on positive adjectives). Thus, in studies regarding intergroup differences discrimination – in the sense of making a difference between members of two groups – is almost uniquely a better treatment of the ingroup, its members, or products. The outgroup and its members are not easily treated negatively in terms of infliction of harm on them. Therefore a distinction has been made by Brewer (1999) between 'ingroup love' and 'outgroup hate' (p. 429). While the former denotes the relatively mild form of ingroup favoritism just described, the latter refers to more negative treatment of an outgroup or its members at large or the more

negative treatment of an outgroup and its members compared to an uncategorized control group.

The case of infliction of more harm on outgroup than ingroup members is at focus here. Empirical studies by Mummendey and colleagues (for a review see Mummendey & Otten, 1998) have repeatedly shown that arbitrary categorization alone is not enough to elicit intergroup differences favoring the ingroup when it comes to the allocation of negative resources (e.g., infliction of harm, withdrawal of positive resources). They concluded and found empirical evidence that so-called *aggravating conditions* (Mummendey & Otten, 1998) are necessary in order for an intergroup bias to emerge in the allocation of negative resources, such as heightened category salience (Mummendey, Otten, Berger, & Kessler, 2000).

Thus, it seems that mere categorization into arbitrary groups is not sufficient for actual (more) negative treatment of an outgroup and its members to occur. However, most empirical studies investigating intergroup differences in the allocation of negative resources have been conducted in contexts in which the groups as a whole or their members behaved at least neutral or even positive. To behave negatively towards any individual or group in such a situation is not legitimate and therefore not a common mode of evaluation or treatment. Blanz et al. (1997) have indeed obtained evidence that ingroup-favoring treatment regarding positive resources is considered more normative and less illegitimate than that regarding negative resources. Less direct evidence comes from Otten, Mummendey and Buhl (1998). They found longer reaction times in allocations of negative resources and both positive *and* negative resources compared to the allocation of positive resources only. This suggests that the allocation of negative resources (even if allocated simultaneously with positive resources) triggers legitimacy concerns which in turn disrupt information processing, a finding in line with the argument that this allocation of negative resources is a rather unusual behavior in

neutral or positive intergroup contexts. Furthermore, in a study reported by Wenzel and Mummendey (1996), an ingroup bias in evaluations of ingroup versus outgroup products occurred if adjectives in a list provided for these evaluations were exclusively positive, but not if the list contained only negative or even both positive as well as negative adjectives. It seems that the mere presence of negative adjectives in the material alone eliminated discrimination by making salient the possibility of discrimination on explicitly negative dimensions. On the other hand, the allocation of positive resources is in principle desirable in those contexts and therefore permits for 'mindless' ingroup favoritism (Mummendey & Otten, 2001). After all, giving anybody more positive things than somebody else will be less undesirable than actually inflicting more harm on somebody than somebody else. The allocation of negative resources in a neutral or positive context increases the salience of potentially unfair or discriminatory behavior and thus triggers deeper and more mindful behavior which then also eliminates intergroup differences.

But what if the behavior by members of an ingroup versus an outgroup to be evaluated is already negative and negative treatment of such a target is legitimately responded to with the infliction of harm? The case examined here is legitimate punishment for norm violations. Reacting to wrongdoing with negative evaluations, offensive emotions and negative, punishing treatment is quite legitimate and 'natural' as it was discussed in the previous chapter. Therefore legitimacy concerns regarding a negative treatment of any offender (i.e., independent of possible information about category membership) should be diminished. In this case, an intergroup bias could re-emerge such that ingroup members are treated more favorably than outgroup members. Harsher negative treatment for an outgroup target could be more appropriate than for an ingroup target. To the author's knowledge, this idea has not been

explicitly empirically explored until now, even though it has been formulated, for example in the very similar domain of retaliatory aggression by Marilynn Brewer:

> When aggression is not clearly justified by the circumstances, most individuals do not display discriminatory aggression against outgroup members (and may show more aggression toward another ingrouper than toward an outgroup target). But when aggression can be justified by provocation or other circumstances, so that it does not appear to be motivated by prejudice, aggression against outgroup members in particular is greatly increased. (Brewer, 2003, p. 82).

Of course, independently of group membership, punishment should be found appropriate. It is not expected that an outgroup target will be treated negatively while the ingroup counterpart would not. For both targets, punishment is appropriate but by a mechanism to be developed shortly, it should be less harsh for an ingroup member.

In sum, I argue that categorization relatively poor in content (i.e., which does not bear any diagnostic information regarding the norm-violating behavior in question) is sufficient to elicit relative lenience towards an ingroup target compared to an outgroup target. This relative lenience should be apparent in punishment recommendations. Such recommendations are of course far from actual behavior. Considering the fact that modern criminal justice has been highly institutionalized, very few lay persons recommending punishment for a crime will actually be willing or able to execute it. But the recommended severity of punishment will presently serve as a relatively proximal correlate of the legitimacy of negative treatment and – as it encompasses emotional reactions (see above) which could trigger action tendencies (Frijda et al., 1989) potentially flowing into actual behavior –, as a more distal predictor of a perceiver's own behavior.

4.2 Why a Difference in Punishment Recommendations?

So far, the hypothesis simply claims that an effect – an ingroup bias – should arise under certain conditions, but no rationale was given why there would be such a bias. The first and

most obvious one is based on one very popular explanation for the ingroup favoritism effect regarding the allocation of positive resources: Social Identity Theory (SIT, Tajfel & Turner, 1979). SIT posits that individuals do not only have a personal, individual identity distinguishing them from other individuals, but also a social identity, which defines them in terms of their membership in a group or category. Just as individuals in general strive to have a positive view of themselves personally (self-esteem; Baumeister, 1998), they also strive to view themselves positively as members of a category. One consequence of this striving (among others which are not discussed here) is the creation and maintenance of differences between the group they belong to (their ingroup) and a corresponding outgroup. In the case of minimal content of the categorization constituting group membership, they do so by 'inventing' differences which make their ingroup positively distinct from outgroups (*creative distinctiveness*, Spears, Jetten, & Scheepers, 2002). Ways to create such positive distinctiveness are evaluating the ingroup as a whole as more positive on a certain trait, evaluating the ingroup's products or behaviors as superior or even endowing the ingroup with more positive resources (as in the classic MGP; Tajfel et al., 1971).

For the present case, in which norm-violating behaviors are at focus, this explanation however runs into trouble. It is not plausible that individuals can draw positive self-esteem from viewing their ingroup less negatively than the outgroup when they evaluate and treat an outgroup target more negatively than an ingroup member. The possibly 'invented' fact that the transgression by a member of the ingroup is to be condemned but that the same behavior of an outgroup member is to be condemned even more strongly will not allow for positive self-esteem flowing from this difference. For a perceiver who considers herself to be law abiding and in general not a norm-violating person, the self-esteem that might be drawn from belonging to a group whose member has done a (in her mind) slightly less bad thing than

another, is still negative. Therefore, it is considered unlikely that a potential ingroup bias on reactions to norm-violations would be driven by similar processes than those invoked by SIT to explain ingroup favoritism on positive dimensions.

A different approach to explain ingroup favoritism in minimal group settings claims that an ingroup (and its members) profit from an association with automatic positive affect. According to this perspective, perceivers do not have any information about minimal groups (and their members) and their attributes by definition. They therefore rely on information about themselves and project this information onto unknown targets (social projection, see Krueger, 1998). This projection is assumed to occur automatically, that is, without the awareness of those projecting, spontaneously, quickly, and with low effort (Bargh, 1996). One such piece of information is a basic evaluation and since individuals tend to have positive self-esteem and view themselves as positive rather than negative (Baumeister, 1998), they project that positive valence onto other people.

It has been shown that projection onto ingroups is stronger than onto outgroups (Cadinu & Rothbart, 1996; Otten, 2002; Clement & Krueger, 2002, Robbins & Krueger, 2005). It is therefore possible that a better evaluation of ingroups than outgroups stems from a relatively simple mechanism by which an ingroup profits from social projection of positive valence whereas outgroups do not or at least to a lesser degree. This difference in automatic positive connotation of the ingroup, but not the outgroup has been found in various empirical works.

Otten and Wentura (1999) found in evaluative priming tasks that the label of an ingroup in a minimal group context acts similar to *a priori* positive primes (i.e., times and errors for reactions to positive targets were decreased if primed with an ingroup label, just like with *a priori* positive primes not related to the ingroup or outgroup), whereas the outgroup label neither sped up reactions to positive nor negative target words. Perdue, Dovidio, Gurtman,

and Tyler (1990) found similar effects using ingroup (e.g., we) and outgroup pronouns (e.g., them) as subliminal primes.

Results from a paradigm of spontaneous trait inferences (Uleman, Hon, Roman, & Moskowitz, 1996) reported by Otten and Moskowitz (2000) suggest that for ingroup targets trait inferences from specific positive behaviors to positive traits were made, whereas this was not the case for outgroup targets or for negative behaviors and traits.

In a similar vein, Gramzow and Gaertner (2005) formulated the *Self-As-Evaluative-Base* hypothesis (SEB) and found confirming evidence that indeed personal self-esteem measured by the Rosenberg (1965) self-esteem scale predicted (biased) positive evaluation of a novel ingroup after controlling for expectancy-based processing (Gramzow, Gaertner, & Sedikides, 2001), *Collective Self-Esteem* (Crocker & Luhtanen, 1990), *Right-Wing Authorianism* (Altemeyer, 1981) and *Narcissism* (Raskin & Terry, 1988). Self-esteem may be thought of as a global positive evaluation of the self. Therefore, the predictive value of this construct for ingroup bias is consistent with the assumption that positive affect associated with the self is projected onto an ingroup.

If ingroups but not outgroups are associated with general positive affect as the foregoing argument suggests, then affective reactions towards an ingroup versus outgroup or their respective members should be subject to influence of that positive affect. Specifically, a reaction characterized by negative affect to a specific event involving the ingroup or one of its members could be inhibited by that positive affect associated *a priori* with the ingroup. For an outgroup member no such *a priori* affective advantage exists. Therefore, the negative reaction to the specific event in question is not buffered in the same way as for an ingroup member. In the case of a punitive reaction towards a norm violation as discussed above then, the punishment recommendation for an ingroup member would be milder than for an outgroup

member as the indicator for the severity of the deed, the negative affect experienced, is – in sum – less intense. As mentioned above, of course, it is not argued that for an ingroup member no punishment at all would be recommended, but that this recommendation should be attenuated by the positive affect elicited by ingroup membership.

To be clear, a number of research efforts have been dedicated to group differences in punishment with naturally occurring groups and found such differences. However, this research has mostly targeted group memberships such as black versus white offenders in the US or the UK (e.g., Hodson, Hooper, Dovidio, & Gaertner, 2005; Johnson, Whitestone, Jackson, & Gatto, 1995; Sommers & Ellsworth, 2000a; 2000b[3]) and thus used naturally occurring groups with a long history of often conflict-laden group relations. Moreover, the attribute 'criminal' is part of the stereotype about Blacks (Hodson et al., 2005) such that attributional and stereotype consistency issues probably weight heavily in judgments about black as well as white offenders (Gordon, 1990; Gordon, Bindrim, & McNicholas, 1988). In the present research however, the main interest is to see whether group membership *per se*, that is independent of specific stereotypes of perceived criminal behavior propensity, will be sufficient in creating differentially negative reactions.

3 See also for example Donnerstein & Donnerstein (1973, 1975, 1978), Donnerstein, Donnerstein, Simon, & Ditrichs (1972) for research with white and black participants on aggressive reactions to provocation, which are conceptually quite similar to punishment reactions as conceived of here. However, as 'aggressive' is also a stereotypical trait for African Americans (Devine, 1989), the interpretation of this research with respect to mere categorization effects on aggression is stricken by similar problems as research on punishment. Similar concerns complicate the discussion of research by Moreno and Bodenhausen (2001) regarding the stereotyped group of gay men and lesbians, for which negative evaluation was found if a seemingly legitimate reason for such evaluation was present.

5 Objection? The Black Sheep Effect

The theoretical reasoning in the last chapter leads to the prediction of a simple difference in punishment reactions to the advantage of an ingroup target compared to an outgroup target. While this hypothesis may seem straightforward (or even trivial to some) at first glance, it is not. There is a line of research showing the exact opposite effect, namely an advantage of outgroup targets compared to ingroup targets in the context of evaluations of targets displaying socially undesirable behavior. This effect has been labeled the *Black Sheep Effect* (henceforth referred to as BSE, Marques & Paez, 1994). In the following sections, research on the BSE and explanations for it will be discussed first. Then an argument why it may not apply to negative reactions like punishment recommendations shall be advanced. This argument will flow into the prediction that under certain conditions – actually the more common conditions when it comes to punishment reactions – a reversal of the BSE will occur.

5.1 The Black Sheep Effect

The first publications of the BSE under that name were by Marques, Yzerbyt and Leyens (1988) and Marques and Yzerbyt (1988). They found that evaluations of targets from an ingroup behaving in an undesirable manner (i.e., violating a norm) are more negative than those for comparable outgroup members while for desirable behaviors, a typical ingroup favoritism effect emerged (i.e., more positive evaluations of ingroup than of outgroup targets). The targets presented in Marques et al. (1988) were qualified as 'unlikeable' (Study 1, intergroup context: Belgians = ingroup versus North Africans), 'putting amusement behind studying' (Study 2, intergroup context: Belgians = ingroup versus North Africans), or 'triggering the events at Heysel Stadium in Brussels, Belgium in May 1985'[4] (Study 3,

4 In this incident, football rioters had started fights in the course of which 39 people died.

intergroup context: Belgians fans = ingroup versus German fans). Marques and Yzerbyt (1988) had participants listen to short good or poor quality speeches and rate the speakers (intergroup context: law = ingroup versus philosophy students).

These original publications were followed by a number of replications of the basic effect (Abrams, Marques, Bown, & Henson, 2000; Abrams, Rutland, & Cameron, 2003; Abrams, Rutland, Cameron, & Marques, 2003; Marques, Abrams, Paez, & Martinez-Taboada, 1998; Marques, Abrams, & Serôdio, 2001; Marques, Robalo, & Rocha, 1992), which also provided evidence that it is a genuine intergroup effect (e. g. Branscombe, Wann, Noel, & Coleman, 1993). These papers also provided evidence that the BSE is especially prone to emerge if the violated norm is relevant as opposed to not relevant to the perceiver (Marques, 1990) and specific to the ingroup rather than general, or in other words, applying to both ingroup and outgroup equally (Marques et al., 1988). A further condition enhancing the effect is norm insecurity (i.e., general heterogeneity in ingroup behavior is perceived along with individual violation, Marques et al., 2001) and participants expecting to be accountable to ingroup members rather than outgroup members for their responses in the questionnaires used (Marques et al., 1998).

Thus, the ingroup favoritism predicted in Chapter 4 seems to be at odds with research on the BSE. While the argument proposed here predicts less negative reactions to norm violation by an ingroup target, there is much evidence for a derogation of an ingroup target compared to an outgroup target (i.e., a BSE). However, there are features across the studies on the BSE and theoretical developments of its explanation, which are different from the general setting proposed here (i.e., punishment reactions). Therefore, they at least do not logically contradict the processes hypothesized here and the predictions of milder punishment recommended for an ingroup target because of a differential positive affect buffer.

5.2 Why Would the Black Sheep Effect not Occur for Punishment Reactions?

First, the violated norms or transgressions used in BSE studies are almost all descriptive rather than prescriptive. While the former are in essence distribution information about actual behavior of a certain group of people, the latter are standards to which one should to abide regardless of whether most individuals of a given population or group actually do abide by it (see Cialdini, Kallgren and Reno, 1991). These latter norms are also distinguishable from descriptive norms in that their violation is subject to social sanctions, such as punishment or reprehension. Descriptive norms are not *a priori* associated with such consequences for those violating them.[5]

Studies in the BSE literature typically operationalize norm violation as a deviation from allegedly modal and relatively harmless behaviors for which there is a differential propensity in the ingroup versus outgroup (e. g., Marques et al., 2001; Marques et al., 1998). There is only one exception: In experiment 3 of Marques et al. (1988), the norm-violating behavior was rioting at a football game, which resulted in many people killed. But in all other cases, the behaviors described as norm-violating are of rather mild undesirability and do not constitute standards whose violation is commonly subject to punishment (being unlikeable, putting amusement behind study, Marques et al., 1988; giving a poor speech, Marques & Yzerbyt, 1988; attitudes towards initiation practices, opinion on homosexual people having the right to choose their own sexual life, Marques et al., 2001; ranking characters in a criminal story according to their responsibility for the death of a woman, Marques et al., 1998; attitudes towards preferential treatment of overseas students in housing allocation, Abrams et al., 2002). The typical norm that is described as being violated in BSE experiments is thus a modal attitude or behavior, differing in distribution between ingroup and outgroup, rather than

5 In this last respect this distinction is similar to the one of perfect versus imperfect duties discussed on p. 16.

a clearly punishable behavior. Therefore a look at whether a BSE emerges with violations of overarching norms, that is norms which are not distinctive for ingroup and outgroup, and which are in principle subject to punishment is clearly instructive in its own right.[6]

Second, as has been mentioned above, the BSE is most pronounced if the norm violated is specific to the ingroup (Marques et al., 1988, Study 2), that is if the expectation for an ingroup member adhering to the norm is higher in the first place. This finding has received particular attention in the most comprehensive and developed theoretical explanation of the BSE, the *Subjective Group Dynamics* model (SGD, Abrams, Marques, Randsley de Moura, Hutchinson, & Bown, 2004; Abrams, Randsley de Moura, Hutchinson, & Viki, 2005). This model, in line with SIT (Tajfel & Turner, 1979) rests on the assumption that positive ingroup distinctiveness is essential for a positive social identity. This means that people strive for a realistic or merely perceived difference in the behaviors or attributes of the ingroup on the one and the outgroup on the other hand. This difference is preferentially positive, such that an overall description of the ingroup is more positive than that of an outgroup. A member of the ingroup deviating from the typical behavior of ingroup members, that is who is more similar to the outgroup members than prototypical ingroup members, therefore elicits considerable negative reactions. This target threatens positive ingroup distinctiveness and, additionally, shakes the validity of the ingroup norm that constitutes its superiority. A comparable outgroup member who shows the same behavior as the ingroup deviant also threatens the positive distinctiveness of the ingroup (or even distinctiveness in general), but at the same time validates the ingroup norms. Their behavior is, according to SGD, considered as beneficial to ingroup superiority as it provides support for the ingroup specific behavior that distinguishes

6 The doubts as to whether a BSE invariably emerges with heavier norm violations and truly bad deeds, will be of interest again when a possible moderator in the *Intergroup Punishment Difference Hypothesis* (see below) is discussed. I will come back to this later.

it positively from the outgroup. As an example, consider a religious group X drawing much of its subjective ethical superiority over another denomination Y from categorically rejecting abortion. Group Y is very liberal regarding abortion. If a member of X stated that abortion were admissible in some rigidly described instances, this statement would undermine the validity of the superiority claim of group X (i.e., the statement would be too liberal) and attract considerable negative reaction. A member of Y however, making the exact same statement would have to be regarded as deviating from the group norm of Y (i.e., that member would be considered too conservative by her own ingroup), but gives support to the notion that being stricter on the subject than group Y on average (as is group X) is a more viable stance on the subject (see also Bègue, 2001, for such an example). Therefore, the same statement would be regarded as a sign of apostasy from the superior attitudes if made by an ingroup member and therefore condemned, but as an indication of the falsity of the common attitude of an outgroup if made by an outgroup member – and thus relatively positively received.

While distinctive norms (i.e., norms positively distinguishing the ingroup from the outgroup) may be prescriptive in that complying with them is mandatory for ingroup members over and above a description of how ingroup members actually do behave, the norm deviations examined in past research on the BSE are still quite different from those considered here, namely, crimes. Norms prohibiting crimes are prescriptive in the sense that they are to be adhered to by everyone without exception and also non-distinctive in that ingroups and outgroups (living at least under the same penal code) do not descriptively differ with regard to the behavior in question. Indeed, Study 2 from Marques et al. (1988) shows that there is no BSE if the norm violated is one not distinguishing ingroup and outgroup. They had participants rate individuals (ingroup = Belgians, outgroup = North Africans) who

violated the norm of lending course notes to fellow students which had been determined by a pretest to be applicable to both ingroup and outgroup. In a different condition, the violated norm was preferring to party instead of studying for school, which had been pretested to be a norm distinguishing the ingroup positively from the outgroup. While in the latter condition a BSE was found, a null difference emerged in the former condition. Moreover, in Experiment 3 of Marques et al. (1998), no BSE occurred if norm distinctiveness was not salient. Participants rated four normative targets and one deviant target from either the ingroup or the outgroup. The salience of the distinctiveness of the norm was manipulated by explicitly informing participants of the ingroup and outgroup norms being diametrically opposed to each other (high salience) or not mentioning this alleged fact (low salience). Thus, in the latter, low salience condition, participants in fact did not know anything about the modal behaviors of the ingroup nor the outgroup but only their group membership. In this condition, no BSE, but only ingroup favoritism occurred.

In sum, it remains an open question whether a BSE would also occur when a behavior violates an overarching prescriptive norm for which descriptive behavior distributions are not different for ingroup and outgroup and which is normally and legitimately punished. It is advocated here that it will not. In contrast, in the case of a violation of a non-distinctive and prescriptive norm, an outgroup target will be reacted to more negatively than an ingroup target because of the process outlined in Chapter 3. Specifically, the positive affect associated with the ingroup and its members will buffer negative reactions to the undesirable, norm-violating behavior, while no such attenuation is available to the outgroup and its members.

6 Summary and Hypotheses

From the arguments developed above, three complexes of hypotheses were derived and tested. These hypotheses will be presented in the following.

6.1 The *Intergroup Punishment Difference Hypothesis*

The *Intergroup Punishment Difference Hypothesis* states that punishment recommended for an ingroup target who has broken the law (i.e., has clear violated a societal norm) will be less intense than recommended punishment for the same behavior carried out by an outgroup target. This follows from the arguments above suggesting that:

- Punishment is assigned proportionate to the negative affective experience upon learning about a norm deviation,
- The intensity of this negative affect is influenced by affective experience flowing from sources extraneous to the norm violation in question,
- Ingroup members enjoy an automatic association with positive affect whereas outgroup members do not, and
- This association acts as an attenuating buffer in punishment reactions.

Hypothesis 1:	*Evaluations will be more negative and punishment recommendations will be higher if a criminal offender is an outgroup member rather than an ingroup member.*

A Possible Moderator Regarding the Intergroup Punishment Difference Hypothesis

A possible moderator of this effect is considered here: It is possible that reactions to lighter offenses (e.g., burglary with theft of a smaller amount of money) may not elicit strong affective reactions, but that reactions are affectively relatively mild in the first place. In every day life, for example, learning of lesser crimes may be relatively common, such that there are

scripts for judgments about these lighter norm deviations. If the punishment recommendation flowing from the affective reaction process is thus less intense, then the advantage in the *a priori* association with positive affect for an ingroup target may not have as much influence as in the case where the judgment is more strongly (or maybe even exclusively) determined by the affective reaction. Therefore, the severity of the offense may moderate the intergroup difference in punishment, such that the difference is stronger for heavier compared to lighter offenses.

Hypothesis 1a: The difference predicted in Hypothesis 1 will be more pronounced in the case of a heavier rather than a lighter offense.

Note that in the discussion of the BSE in Chapter 5, a similar idea was implicit: While ingroup members may be evaluated less positively than outgroup targets (i.e., a BSE) with regard to relatively mild transgressions, this difference could be reversed for more severe transgressions (i.e., a difference consistent with the one predicted here). While this argument will be featured in more detail and lead to the next hypothesis, it also supports the plausibility of the expectation that the ingroup favoritism expected presently should be more pronounced with heavier offenses.

6.2 The *Reversed Black Sheep Effect Hypothesis*

The second hypothesis examined is named the *Reversed Black Sheep Effect Hypothesis*. It predicts that a BSE may emerge in the evaluation of a norm-violating target when the propensity of that violation in the ingroup as a whole is lower than in the outgroup (independent of a target's individual behavior). If no such difference between the groups is perceived, the difference between an ingroup target and an outgroup target will be in the opposite direction. In Chapter 5, it was pointed out that studies on the BSE commonly use undesirable, but not strictly punishable behaviors for which there is an *a priori* descriptive difference consisting in a higher incidence rate of the violation in the outgroup compared to the ingroup (i.e., a distinctive norm). In connection with the reasoning leading to the Intergroup Punishment Difference Hypothesis, it is expected that once these constraints regarding the BSE are relaxed (i.e.,the transgression is a categorically punishable behavior and no distinctive norms for the groups exist), a bias in favor of ingroup targets will re-emerge. Thus, the hypothesis is a hypothesis of moderation of the difference in evaluation between an ingroup and an outgroup target by norm distinctiveness.

Hypothesis 2:	*For a punishable behavior, an ingroup member will be evaluated more negatively than a comparable outgroup member if the propensity of that behavior is lower in the ingroup than in the outgroup. If the propensities are equal, an ingroup member will be evaluated less negatively than an outgroup member.*

6.3 The *Prior Positive Affect Hypothesis*

Derived from the general theoretical considerations on punishment stated in Chapter 2, this last hypothesis theoretically expands the Intergroup Punishment Difference Hypothesis and follows up on the studies testing this first hypothesis.

Theoretically, this hypothesis proposes the generalization of Hypothesis 1, which predicts a difference between an ingroup and an outgroup target in punitive reactions. Specifically, an affective buffer such as the one assumed for ingroups and their members should also influence punishment tendencies towards targets who are similarly associated with positive affect and of which the perceiver is not a member him or herself. Positive valence elicited by members of a rather positively evaluated third party group[7] should attenuate punishment reactions in the same way as the affect stemming from ingroup membership (see Chapter 3 for the argument that extraneous should affect punishment recommendation).

Hypothesis 3:	*Punitive reactions towards an offender from an* a priori *positively evaluated social category will be milder than those towards an offender from a less positively evaluated category.*

A Possible Moderator Regarding the Prior Positive Affect Hypothesis

A qualification of this hypothesis might be in place, as evidence to a boundary condition arose in the first study testing the Prior Positive Affect Hypothesis (A-1, p. 125) and was explicitly tested as a moderator in three subsequent studies (see Chapters 13 and 14).

Specifically, the use of irrelevant information (such as *a priori* positive affect) in a judgment process should be especially likely in a spontaneous and rather less deliberative state. In a very carefully scrutinizing mind set, in contrast, a judging person may become aware of the biasing effect of the prior positive affect association and try to correct for it (e.g.,

7 Categories to which the perceiver does not belong (i.e., the self is not directly involved) will henceforth be referred to as third party groups. If you will, these groups are two outgroups differing on a relevant dimension (here: general positivity associated with them and determined by a pretest in a neutral setting).

Schwarz & Clore, 1983; Strack & Hannover, 1996; Strack, Schwarz, Bless, Kübler, & Wänke, 1993). Such a correction attempt may be unsuccessful however, in that it overestimates the potential bias and thus the correction results not only in an elimination of a difference as a function of that bias, but even in a reversal of that difference (Fein, Hoshino-Browne, Davies, & Spencer, 2003; Wegener & Petty, 1997). Therefore, the hypothesized effect should be particularly pronounced in a state of heuristic and spontaneous processing and less pronounced, possibly reversed if participants make judgments in an analytic and careful state.

Hypothesis 3a:	*The effect hypothesized in Hypothesis 3 will be most pronounced if processing is spontaneous, while it will be attenuated, eliminated or even reversed if processing is careful.*

7 General Method

For each of the three hypothesis complexes, results from several empirical studies will be reported. All studies are essentially scenario studies, presenting participants information about a norm violation and then asking them for various judgments about this violation and the violator. Within each of the three complexes, studies used similar materials and scenarios.

The individual studies within complexes are often very similar in design, but heterogeneous in results. They therefore do not proceed in a logical or strategic order enlarging the body of results step by step as it is generally known from, say, publications in social psychological journals. Rather, because the study sets within each complex are quite homogeneous as far as materials and basic design are concerned, individual studies could be integrated using meta-analytical procedures. This approach has at least three advantages.

The first, very obvious advantage is that collected data, which – due to small sample sizes or unfortunate laboratory or sampling conditions – did not yield clear and interpretable significant results, do not have to be omitted. Note that this argument is not equivalent to 'squeezing out of the data as much as you can'. If indeed non-significant effects from individual studies constitute sampling error around a population mean of zero (as it follows inherently from the logic of significance testing), and, thus, speak against the (alternative) hypothesis, this will become obvious in a meta-analytic integration much clearer then it would in individual studies. The same line of reasoning applies of course to results (significant or not) that even point in the opposite direction of the hypothesis, but, finally, also to results consistent with the predictions. Thus, the meta-analytic strategy employed here does not capitalize on 'failed' studies, but responsibly uses all available evidence (Rosenthal & DiMatteo, 2001).

Secondly, the integration of several studies of the same kind protects the researcher from prematurely abandoning a paradigm or even a theory on the grounds of one 'failed' study. One may always charge several aspects of a study for the lack of an effect, some of which are irrelevant for the theory and general concept behind the research (e.g., finding a null effect by chance; occasionally occurring failed randomization in assigning participants to conditions by chance, not due to lack of diligence on the researcher's part or insufficient sample size). Other null effects actually speak to the viability of the theory or lack thereof. Studies of these two types are not easily distinguishable. Several studies of similar design however should attenuate the effect of the mentioned irrelevant and randomly occurring aspects while preserving the influence of the relevant aspects. The caution advocated here against relying on single studies of course also applies to the opposite case where clear and significant findings of a single study are taken as clear and final support for a given hypothesis that may not even be theoretically sound. Thus, the integration (by way of meta-analysis or otherwise) of several individual studies is presently considered epistemologically superior to individual studies.

Finally, the interpretation of a set of studies becomes clearer. When not all studies that investigate a particular research question provide consistent and clear results, researchers are faced with the task of integrating them and drawing conclusions based on a heterogeneous picture of results. This can be done narratively, in a qualitative way. This approach is especially suited for a rather large set of studies with very diverse independent and dependent variables, different experimental settings and a host of potential moderators. When the set of studies however is small and homogeneous, employ very similar designs and variables, yet comprises heterogeneous effects, temptations run high of over-interpreting unexpected deviations or lack of finding a significant effect in one attempt of replication. Under such circumstances, meta-analytic procedures provide a good way of summarizing these effects in

order to make a judgment about the viability of the hypothesis based on all of the studies. At the same time one does not risk being carried away by post-hoc hypotheses that cannot be directly tested on the data at hand. Meta-analytic integration treats all observed cases from a set of studies as one large sample. The result then indicates, as one finding, whether there is an effect across all the studies consistent with the hypothesis or not.

Given these advantages, individual studies from the three complexes will be synthesized in a meta-analysis. This allows for a relatively clear and unambiguous picture concerning the viability of the hypotheses without giving disproportionate interpretative weight to single significant or non-significant findings.

8 The Meta-Analytic Approach Employed in the Present Dissertation

8.1 The Basic Hypothesis and its Numerical Equivalent

Throughout the studies conducted, the essential hypothesis is that reactions to a norm violation differ as a function of the group which the person described as violating the norm (i.e., the target) is a member. This theoretical hypothesis can be stated as an empirical prediction of the form

> If the target person is from group A, negative reactions will be less intense than if the target is a member of group B.

Given that negative reactions are operationalized as answers to items on answer scales of which it is assumed that they can be treated as continuous variables, the empirical prediction translates into a statistical hypothesis of the form

$$\mu_A < \mu_B \tag{1}$$

where μ is the population mean of answers on a given measure and A and B denote group membership of a presented target. This relationship is equivalent with the hypothesis that the difference ($\mu_A - \mu_B$) is different from zero. Usually such a test is achieved by means of a t-test on a sample of individual observations. In order to integrate results from several studies, the difference is expressed as the effect size (Cohen, 1992), which takes into account the standard deviation of the difference:

$$d = \frac{M_A - M_B}{SD} \tag{2}$$

where *M* are empirical means of a particular condition and *SD* represents the standard deviation . The present dissertation uses the within-cell standard deviation from the complete design from which the difference of means stems. For each individual study, common

analyses such as ANOVAs are reported along with effect sizes *d* which are later used in the meta-analysis. At the end of each complex, these effect sizes are presented in an overview table, rounded to two decimal figures. All subsequent analyses however use the exact results of outputs from SPSS12 and Microsoft Excel sheets which were frequently used to calculate effect sizes, with all available decimal figures. Results of these analyses are then again reported rounded for presentation to the second decimal figure.

8.2 Integration of Effect Sizes

Effect sizes are combined according to the Weighted Integration Method proposed by Hedges and Olkin (1985). In this procedure, individual effect sizes are first corrected for a sampling bias Hedges and Olkin (1985, p.80) identified. These individual corrected effect sizes *d*' are calculated according to the formula

$$d' = d\left(1 - \frac{3}{4\mathrm{N} - 9}\right) \tag{3}$$

The effect sizes from individual studies reported throughout this dissertation are however the uncorrected effect sizes. The mean effect sizes d_m resulting from the meta-analytic processing on the other hand are means of the correct effect sizes *d*'.

Individual effect sizes differentially influence the mean effect size as they are weighted by their estimated variance, which in turn is a function of sample size. As estimates from larger samples are generally more precise than those from small samples, the latter should be disproportionately accounted for (Hedges & Olkin, 1985). The exact formula for the final integration of the effect sizes reflects this weighting is:

$$d_m = \frac{\sum_{i=1}^{p} \frac{d_i'}{\frac{n_1+n_2}{n_1 \times n_2} + \frac{d_i'^2}{2(n_1+n_2)}}}{\sum_{i=1}^{p} \frac{1}{\frac{n_1+n_2}{n_1 \times n_2} + \frac{d_i'^2}{2(n_1+n_2)}}} \quad (4)$$

where d_i' are effect sizes from individual studies and n_1 and n_2 are the sizes of the compared cells (Hedges & Olkin, 1985, p. 86, Formula 14 and p. 112, Formula 8).

8.3 Calculations

This procedure is presently carried out by way of the software package *Computer Programs for Meta-Analysis* by Schwarzer (1989). The program provides the mean effect size over a set of effect sizes from individual studies, their variances, standard errors, and confidence intervals (*CI*) and *p* values indicating the probability of a particular mean effect size occurring if the population parameter is actually zero.

The program also provides results from the analysis with correction for imperfect reliability of the measures. Effect sizes are usually attenuated by reliabilities below perfection (i.e., below 1.0). It is thus possible to estimate the mean effect size if the measure had shown perfect reliability. Results from these corrected analyses are also reported, but actually never really differ from those for uncorrected effect sizes. The reliabilities used for this correction are Cronbach's α from the entire study from which a particular effect size has been obtained.

Generally, the studies reported here comprise several dependent variables. Meta-analyses are conducted for each of these measures separately with special caution exercised in the interpretation of differential or similar effects on different variables, as these measures are occasionally highly correlated.

8.4 Tests for Moderation

Also, some potential moderators of possible overall effects in the group differences are incorporated in the designs of the studies as manipulated independent variables. These variables are used to partition the sets of effect sizes for each complex in order to test for their moderating role. To this end, meta-analyses over the complete set of effect sizes are complemented by meta-analyses over the subsets formed through grouping by levels of the potential moderator. In order to test the moderation, a procedure proposed by DeCoster (2004) will be applied. He proposes to calculate a contrast from means and variances of p subsets according to the following formula:

$$Z = \frac{\sum_{i=1}^{p} c_i d_{m_i}}{\sqrt{\sum_{i=1}^{p} c_i^2 s_i^2}} \tag{5}$$

where c_i are contrast coefficients for subset i, d_{m_i} represents the mean effect size in subset i and s_i^2 is the variance of effect sizes within subset i. In the particular case of two subsets, this reduces to

$$Z = \frac{(1)d_{m_1} + (-1)d_{m_2}}{\sqrt{(1)(1)s_1^2 + (-1)(-1)s_2^2}} = \frac{d_{m_1} - d_{m_2}}{\sqrt{s_1^2 + s_2^2}} \tag{6}$$

which essentially describes a simple standardized difference of mean effect sizes.

Given large sample sizes (which is generally assumed if several individual studies are pooled), the resulting statistic Z is standard normally distributed and will be tested using the p-value corresponding to Z. The exact nature of the moderation will be described using the obtained mean effect sizes of the subsets.

8.5 Additional Remarks

Finally, a few remarks should be kept in mind while reading the following sections. Degrees of freedom occasionally vary because of individual missing data. Unless otherwise stated, confidence intervals reported are 95% confidence intervals. All post hoc tests are least significant difference tests (no correction for error accumulation). The problem of error accumulation over 5% arises in very few cases and is also attenuated by the integration of individual studies in meta-analyses. Means and standard deviations of age information in sample descriptions are rounded to the nearest integer as raw data usually is only accurate to that extent.

9 Studies Investigating the Intergroup Punishment Difference Hypothesis

As the reader will recall, this hypothesis predicted that in evaluation and recommendation of punishment, a bias in favor of the ingroup will be apparent: Ingroup targets should be reacted to less negatively than outgroup targets if they are reported to have committed a crime. Also, a tentative moderation hypothesis was advanced: as the influence of a negative affective reaction to a crime should be clearer the more severe the crime, the basic effect should be more apparent for heavier offenses than for lighter ones.

The Intergroup Punishment Difference Hypothesis was tested using simple scenario studies which followed the same broad scheme: Participants read about offenses and were asked about their views on the offender and his behavior as well about as punishment for the offender. This offender's ingroup or outgroup membership was varied by indicating place of residence, occupation, or age information in a casual way. Also, the offense cases used will be coded as 'lighter' versus 'heavier' in order to test the moderation hypothesis (see Hypothesis 1a on page 35). More details on the specific materials and procedures will be given in the reports of individual studies.

9.1 Study P-1

Participants, Design and Procedure

Participants were 95 students of Friedrich-Schiller-Universität Jena (FSU). They volunteered to participate in several studies conducted in a quiet corner of a hallway in the main building of the university. Participants were on average 21 years old (*SD* = 3 years), 76% were female.

The whole session comprised a short identification questionnaire, an unrelated experiment manipulations of which were fully counterbalanced with those of the present

study and then the present questionnaire. Participants were randomly assigned to one of four conditions of a 2 (offender's *Group membership*: ingroup vs. outgroup) × 2 (*Offense*: lighter vs. heavier) between participants design. Cell sizes were $n = 23$ or $n = 24$. After handing back the questionnaire they received a chocolate bar for compensation, were thanked and a few days later debriefed by email.

The questionnaire presented a short description of a student either from Jena (ingroup condition) or Erfurt (outgroup condition) who had either broken into a store at night and stolen goods worth about €2800 (burglary, lighter offense) or had started a fight with another person in a restaurant and beaten him so that the victim had to be treated at a hospital (assault, heavier offense). Participants first read a short paragraph describing the offense and then, on the next page, indicated how much the offender should be sentenced to on an anchored rating scale (1 = no time in prison, 2 = one day, 3 = two weeks, 4 = two months, 5 = six months, 6 = one year, 7 = three years, 8 = seven years, 9 = 15 years, 10 = 30 years, 11 = life time, *Time in prison*). The steps of this scale are not equidistant in terms of chronological time and some of the higher values are certainly out of range, but the scale was modeled after a scale that has often been used in punitive justice research (Darley et al., 2000; Robinson & Darley, 1995; Weiner, Graham, & Reyna, 1997). Participants also answered in a free format how many *Hours of community service* they found appropriate as punishment for the offender.

These two questions were followed by several additional items answered on 9 point rating scales. They pertained to general harshness of recommended punishment ('The offender should be punished ...', ranging from 1 = 'very mildly' to 9 = 'very harshly', *Harsh punishment*; 'The offender cannot expect mercy', ranging from 1 = 'do not agree at all' to 9 = 'completely agree', *No mercy*), the perceived *Severity* of the offense (1 = 'very light' to 9 = 'very high') and *Emotional reactions* to the offender and the offense ('The offense triggered moral outrage in

me', 1 = 'none at all' to 9 = 'very intense', 'Please indicate how much the offender and the offense caused the following feelings in you': *Anger, Pity, Furor, Liking, Indifference, Contempt,* from 1 = 'I do not feel this at all' to 9 = 'I feel this very intensely'). Also, participants indicated how likely they thought the offender would commit the same or a similar offense in the future (*Likelihood of recidivation*, 1 = 'very unlikely' to 9 = 'very likely'), how similar they found themselves to the offender in norms, values and convictions (S*imilarity to self*, 1 = 'very dissimilar' to 9 = 'very similar') and attributed the offender's behavior on a bipolar continuum from 1 = 'personally' to 9 = 'situationally/his environment' (*Attribution*). Finally, using a graphical scale similar to the one proposed by Schubert and Otten (2002) participants indicated to what degree they perceived the offender to be integrated in general society (*Integration*). On this scale, participants mark one of seven diagrams representing the offender as a small circle and society as a larger circle. There are seven such diagrams sorted according to how distant the offender is from society. The diagram coded as one (if chosen) shows the offender to be far apart from society, in the following pictures this distance decreases, the smaller circle increasingly overlaps with the larger one and finally, on the diagram coded as seven, the offender is represented to be fully included in society.

As these items were differently scaled, scores of *Time in prison, Hours of Community Service, Harsh punishment* and *No mercy* were standardized into *z*-scores; they formed a reliable scale (Cronbach's $\alpha = .74$) and were averaged for an index of *Recommended punishment severity*.

Results

Means and standard deviations for all scales and single item measures are reported in Table 1.

Table 1

*Means and standard deviations of dependent variables in Study P-1 (*N = *95)*

	offense							
	lighter (burglary)				heavier (assault)			
Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Recommended punishment severity	-0.17	0.83	-0.42	0.65	0.28	0.60	0.33	0.66
Severity	4.71	1.68	4.18	1.14	6.83	1.47	6.61	1.70
Likelihood of recidivation	6.21	1.69	5.95	1.84	6.71	1.49	7.22	1.59
Similarity to Self	2.42	1.56	2.23	1.95	1.96	1.37	1.74	1.10
Attribution	5.92	2.75	5.36	2.24	3.75	1.96	3.70	2.51
Integration	4.71	1.65	3.86	2.01	4.33	1.63	4.04	1.67
Pity	1.62	1.38	1.55	1.63	1.83	1.61	1.30	0.70
Liking	3.13	2.23	2.77	2.54	3.21	2.34	3.48	2.37
Outrage	4.30	1.96	4.39	2.04	6.32	1.59	6.80	1.38

Note. Higher values on *Attribution* indicate situational attribution, lower values personal attribution. 'Ingroup' indicates a target from Jena, 'outgroup' indicates a target from Erfurt. *Recommended punishment severity* values are z-scores.

A 2 (*Offense*: burglary vs. assault) × 2 (*Group membership* of the offender: ingroup vs. outgroup) ANOVA revealed only a main effect of *Offense*, $F(1,90) = 18.15$, $p < .001$, $\eta_p^2 = .17$, with recommended punishment being higher for assault ($M = 0.31$, $SD = 0.10$) than for burglary ($M = -0.30$, $SD = 0.10$). Neither the main effect of *Group membership* nor the interaction term were significant, both Fs(1,90) < 1.10, both ps > .29, both η_p^2s < .02.

Effect sizes capturing ingroup versus outgroup differences in *Recommended punishment severity* for the two offenses separately were $d = 0.33$ and $d = -0.07$, for the heavier and lighter offenses, respectively.

Severity, *Likelihood of recidivation*, S*imilarity to self*, *Attribution*, and *Integration* were kept as single item measures. *Emotional reactions* items were subjected to a factor analysis with varimax rotation, which yielded two factors accounting for 42.70% and 7.27% of the variance (eigenvalues = 2.99 and 0.59, respectively). *Anger*, *Moral Outrage*, *Furor* and *Contempt* loaded positively on the first factor and *Indifference* negatively (all loadings $|\lambda| > .61$). *Pity* and *Liking* loaded positively on the second factor (λs = .55 and .38, respectively). Absolute values of all cross-factor loadings were smaller than .18. Therefore, the items of the first factor were averaged to a scale of *Outrage* (Cronbach's $\alpha = .87$) and *Pity* and *Liking* were treated as single items.

All these additional variables were subjected to 2 (*Offense*: burglary vs. assault) × 2 (*Group membership* of the offender: ingroup vs. outgroup) ANOVAs. *Group membership* had no main effect on any of these variables, all Fs(1,89) < 2.65, all ps > .12, all η_p^2s < .03. *Offense* had main effects on *Severity, Likelihood of recidivation, Attribution,* and *Outrage,* all Fs(1,89) > 6.58, all ps < .02, all η_p^2s > .07, but not on *Similarity to Self*, *Integration*, *Pity*, and *Liking,* all Fs(1,89) < 2.26, all ps > .13, all η_p^2s < .03. The main effects reflected more negative reactions and more personal attribution in the face of the more severe offense (assault, see Table 1 for means). No interactions emerged, all Fs(1,89) < 1.24, all ps > .27, all η_p^2s < .02.

Discussion

In sum, contrary to predictions, *Group membership* did not have an effect on *Recommended punishment severity* nor on the other variables measured. It also was not involved in any interactions. *Offense* did have a main effect on *Recommended punishment severity*, *Severity*, *Likelihood of recidivation*, *Attribution*, and *Outrage*, consistent with the

assumption that the more severe the offense, the more negative the reactions toward that offense and the perpetrator.

In the next study, another test of the hypothesis was undertaken using a different ingroup versus outgroup distinction and different crimes. Also, the ranges of the *Recommended punishment severity* items was changed to fit the different crimes used. Particularly the *Time in prison* scale used in the last study was not used in the next, as participants seemed to not use the whole range. Also two additional items asking for assignment of specific punishment (probation and fine, see below) were added for this scale.

9.2 Study P-2

Participants, Design and Procedure

Sixty students of FSU participated in this study (mean age = 22 years, *SD* = 3 years; 62% female). They volunteered, were compensated and debriefed as the participants in Study P-1.

Participants were randomly assigned to one of four conditions. These conditions resulted from orthogonal crossing of the factor offender's *Group membership* (ingroup: German born in Germany versus outgroup: German born in Kazachstan) and *Offense* (lighter offense: burglary with theft of goods worth about €9000 versus heavier offense: rampage and severe assault on an innocent bystander in a bar). There were material errors in the questionnaire filled out by six participants. Data from these participants were excluded. Two participants indicated to be not born in Germany or that their native language was not German. As the intergroup context used here was that between German born as ingroup and non-German born as outgroup, data from these participants were also omitted, leaving a final total sample of N = 52. Cell sizes were n = 14 or n = 12.

Participants again read the description of an offense first. On the next page of the questionnaire, they rated the offender on five bipolar adjective scales (*likable* vs. *unlikable*, *good* vs. *bad*, *reliable* vs. *unreliable*, *friendly* vs. *unfriendly*, *moral* vs. *immoral, negative* vs. *positive*). All rating scales but *negative* vs. *positive* ranged from 1, representing the positive, to 7, indicating the negative counterpart in the ingroup condition while, due to a material error for the outgroup offender version, *good* vs. *bad* and *moral* vs. *immoral* were reversed. The analyses pertaining to the measure of general *Evaluation* resulting from these adjective ratings were therefore conducted using two different versions of the scale: one for which only the four adjectives that were identically coded in all questionnaires were averaged (*Short evaluation scale*, α = .64 for the total sample, .69 for the ingroup conditions and .61 for the outgroup conditions) and one comprising all six adjectives for which the two reversed items in the outgroup version were first recoded (*Long evaluation scale*, α = .76 for the total sample, .78 for the ingroup conditions and .74 for the outgroup conditions).

On seven point rating scales ranging from 1 = 'do not agree at all' to 7 = 'completely agree' participants then indicated agreement to statements about the legitimacy of temporary *Exclusion* of the offender ('The offender deserves to be eschewed by other people because of his behavior at least temporarily'), his general *Deservingness* of punishment ('The offender deserves punishment for what he has done'), the *Wrongness* of the offender's behavior ('I find the offender's deed wrong'), and the degree to which they thought that the offender had excluded himself from of society by his behavior (*Self-exclusion,* 'By his behavior, the offender has put himself outside of the community of decent people at least temporarily').

Finally, participants indicated the amount of *Time in prison* they found appropriate for the offender on an eight point scale in steps of one half year, accordingly ranging from one half year to four years, whether the offender should be let off with *Probation* (dichotomous

Table 2

*Means and standard deviations of dependent variables in Study P-2 (*N = *52)*

	offense							
	lighter (burglary)				heavier (assault)			
Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Evaluation (Long scale)	4.68	0.76	4.54	0.75	5.61	0.79	5.35	0.86
Recommended Punishment Severity	-0.47	0.72	-0.02	0.50	0.34	0.40	0.22	0.78
Exclusion	4.50	1.79	4.36	1.91	5.25	1.55	4.83	1.95
Wrongness	6.36	1.34	6.83	0.39	6.86	0.36	6.67	0.65
Self-Exclusion	4.71	1.07	5.14	1.46	0.58	1.62	6.25	0.97

Note. 'Ingroup' denotes a target born in Germany and living in Rudolstadt near Jena, 'outgroup' denotes a target born in Kazachstan with German citizenship and living in Rudolstadt. Higher values on Evaluation indicate more negative evaluation. *Recommended punishment severity* values are z-scores.

item, yes versus no), how many hours of *Community service* they found appropriate as punishment (free answer format, from 40 to 520 hours), and which amount of a monetary *Fine* they thought legitimate (free answer format, from €200 to €3500). The items pertaining to recommended punishment (*Deservingness, Time in prison, Probation* [reverse coded], *Community service,* and *Fine*) were *z*-standardized and averaged to a scale of *Recommended punishment severity* ($\alpha = .71$).

Results

Means and standard deviation for all scales and single items are reported in Table 2.

Regardless of the type of *Evaluation scale* used (*Short* or *Long*), only the offense had a main effect on evaluation, both Fs(1,48) > 9.20, both ps < .004, both η_p^2s > .16. The main effect of *Group membership* and the interaction term were not significant, all Fs < 1. Effect sizes capturing ingroup versus outgroup differences on the *Long evaluation scale* for each

offense separately were $d = 0.18$ and $d = 0.33$ for the lighter and heavier offenses, respectively.

Offense had a main effect on *Recommended punishment severity*, $F(1,48) = 9.10$, $p = .004$, $\eta_p^2 = .16$. Responses were higher for the assault case ($M = 0.28$, $SD = 0.13$) than for the burglary case ($M = -0.24$, $SD = 0.13$). There was no main effect of *Group membership,* $F < 1$, and a non-significant interaction, $F(1,48) = 2.67$, $p = .109$, $\eta_p^2 = .05$. Descriptively, however, means pointed towards a clearer *Group membership* difference within the burglary case, $F(1,48) = 3.46$, $p = .069$, $\eta_p^2 = .07$, than in the assault case, $F < 1$. *Recommended punishment severity* in the burglary case (lighter offense) tended to be higher for an outgroup offender ($M = -0.02$, $SD = 0.17$) than for an ingroup offender ($M = -0.46$, $SD = 0.17$), while for assault, if anything, the means tended to be in the opposite direction ($M = 0.34$, $SD = 0.40$ and $M = 0.22$, $SD = 0.78$), even though this difference was far from statistical significance.

Effect sizes capturing ingroup versus outgroup differences on *Recommended punishment severity* for each offense separately were $d = -0.70$ and $d = 0.20$ for the lighter and heavier offenses, respectively.

In analyses of additional variables there was only a main effect of *Offense* on *Self-Exclusion*, $F(1,48) = 7.41$, $p = .009$, $\eta_p^2 = .13$. Participants agreed more that the offender had excluded himself if he had committed assault ($M = 5.92$, $SD = 1.35$) rather than burglary ($M = 4.93$, $SD = 1.27$). All other main and interaction effects were not significant, all Fs$(1,48) < 2.28$, all ps $> .13$, η_p^2s $< .05$.

Discussion

In sum, a main effect of offense emerged on *Evaluation* and *Punishment*, attesting to the validity of the severity distinction between the offenses. Contrary to predictions, there were no main effects of the offender's *Group membership,* but a tendency for an interaction:

punishment recommendations were a marginally milder for an ingroup than an outgroup member in the lighter offense, but did not differ for the heavier offense. This is contrary to the moderation hypothesis stating that the intergroup difference in punishment tendencies should be more pronounced the heavier the offense and thus, the more affective reaction the offense triggers.

The results were not clear-cut, therefore another test was attempted. As the offenses used in Study P-2 could also differ in more than severity, the next study used the qualitatively same offense for the lighter and heavier conditions, and varied only the severity by manipulating the harm done in the offense. Also, the *Recommended punishment severity* items were somewhat changed (see below) as answers on most of these items were heavily skewed.

9.3 Study P-3

Participants and Design

Fifty-nine students of FSU participated in this study. They were approached around campus, if they agreed to participate in the study, they were given the questionnaire and filled it out by themselves. They were assigned randomly to one of four cells of a 2 (*Group membership:* ingroup vs. outgroup) × 2 (*Offense:* lighter vs. heavier) between participants design.

Three students were observed to copy answers from another participant and in one case, the questionnaire pages were stapled in a wrong order. Also, in two manipulation check questions in the end of the questionnaire, three participants did not report the offenders' group membership correctly and three indicated that they had to go back to the description of the offense to be able to answer the manipulation checks. These cases were omitted from

analyses. The final sample thus comprised data from 49 participants (mean age = 22 years, *SD* = 2 years; 63% female) resulting in cell sizes of *n* = 12 or *n* = 13.

Participants were later debriefed by email.

Procedure

As in Studies P-1 and P-2, participants obtained and filled out all materials in one questionnaire package. First they answered a few questions about their satisfaction with studying at FSU and in Thüringen (the state which FSU is located in) to make salient the comparison to Sachsen, the neighboring state. They then again read a short description of a burglary in the course of which the offender either stole goods worth €9000 from a store at night and no mention was made of damage he caused while breaking in (lighter offense) or stole goods and caused considerable damage while breaking in, resulting in a total damage of €20,000 (heavier offense). The offender was either portrayed as a student from Jena (ingroup condition) or from Dresden (in Sachsen, outgroup condition).

Participants then answered questions regarding the offense and the offender. They again rated the offender on the same 11 bipolar adjective scales as before for an index of *Evaluation* (α = .88, higher values indicate more negative evaluation). They then indicated on seven-point rating scales (1 = do not agree at all to 7 = completely agree) their agreement with statements regarding *Deservingness* ('The offender deserves punishment for what he has done'), *Wrongness* of the behavior ('I find the offender's deed wrong'), the legitimacy of temporary *Exclusion* of the offender ('The offender deserves to be eschewed by other people because of his behavior at least temporarily') and *Self-Exclusion* ('By his behavior, the offender has put himself outside of the community of decent people at least temporarily').

They then indicated how much *Time in prison* they thought appropriate for the offense in free answer format (up to 60 months), whether the offender should be let off with *Probation*

(five-point rating scale, anchors: 1 = 'definitely', 2 = 'rather yes 3 = 'undecided/don't know' 4 = 'rather not', 5 = 'definitely not'), recommended number of hours of *Community Service* (free answer restricted to a number between 40 and 520) and recommended amount of a monetary *Fine* (free answer format, restricted to an amount between €200 and €25,000). These recommended punishment items and the item *Deservingness* were *z*-standardized and averaged for a scale of *Recommended punishment severity* ($\alpha = .58$).

On the last page, participants finally indicated their agreement on seven point rating scales (1 = 'do not agree at all', 7 = 'completely agree') with statements regarding *Damage to the Reputation* of Jena ('The offender has damaged the reputation of Jena by his behavior' for both ingroup and outgroup offender), *Unpleasantness* of being connected to the offender ('I would find it unpleasant to have contact with the offender'; 'It would be embarrassing to be friends with the offender', and 'I would rather not be connected to the offender', $\alpha = .89$), and *Threat to social order* implied by the offender's behavior ('Behavior like the offender's threatens the social order' and 'People like the offender are a danger to society's functioning, $\alpha = .79$).

Results

Means and standard deviations of all dependent measures are shown in Table 3.

A 2 (*Offense*: lighter vs. heavier) × 2 (*Group membership*: ingroup vs. outgroup) ANOVA on *Evaluation* ratings revealed no main effects, both *F*s < 1, but a marginal *Offense* × *Group membership* interaction, $F(1,45) = 2.90$, $p = .096$, $\eta_p^2 = .06$. Simple main effects within offenses were not significant by themselves, but in opposite directions. While descriptively, the ingroup offender tended to be rated less negatively than the outgroup offender in the lighter offense scenario, $F(1,46) = 0.92$, $p = .342$, $\eta_p^2 = .02$, this pattern was reversed – and stronger – in the heavier offense condition, $F(1,46) = 2.09$, $p = .155$, $\eta_p^2 = .05$.

Effect sizes capturing ingroup versus outgroup differences on *Evaluation* for each offense separately were d = -0.39 and d = 0.59 for the lighter and heavier offenses, respectively.

Recommended punishment severity was not affected by *Offense*, *Group membership,* or their interaction, all Fs(1,45) < 1.84, all ps > .18, all η_p^2s < .04. Descriptively however, there was a difference to the advantage of the ingroup offender in the heavier offense scenario while in the lighter offense scenario, this difference was very small (see Table 3 for details). Effect sizes capturing ingroup versus outgroup differences on *Recommended punishment severity* for each offense separately were d = -0.10 and d = -0.68 for the lighter and heavier offenses, respectively.

Regarding additional measures there was a significant main effect of *Offense* on *Threat to social order*, $F(1,45) = 4.51$, $p = .039$, $\eta_p^2 = .09$. *Threat to social order* was higher for the heavier offense scenario (M = 4.65, SD = 1.28) than for the lighter offense scenario (M = 3.82, SD = 1.46). There was also a marginal effect of *Offense* on *Self-Exclusion*, $F(1,45) = 3.03$, p = .089, $\eta_p^2 = .06$, such that the offender was seen to have put himself out of society more if he had committed the heavier offense (M = 5.13, SD = 1.60) rather than the lighter one (M = 4.36, SD = 1.58). This main effect was marginally moderated by *Group membership*, $F(1,45) = 3.23$, p = .079, $\eta_p^2 = .07$. While for the ingroup offender, agreement to the item differed in line with the just reported main effect of *Offense*, $F(1,45) = 6.13$, p = .017, $\eta_p^2 = .12$, there was no such effect of *Offense* for the outgroup offender, F < 1.

There were no other main and interaction effects on additional measures, all Fs(1,45) < 2.80, all ps > .10, all η_p^2s < .06.

Table 3

*Means and standard deviations of dependent variables in Study P-3 (*N = *59)*

	offense							
	lighter (burglary €9000)				heavier (burglary €20000)			
Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Evaluation	5.00	0.62	5.30	0.63	5.57	1.01	5.10	0.79
Recommended Punishment Severity	-0.13	0.48	-0.06	0.79	-0.11	0.74	0.35	0.62
Wrongness	6.08	1.51	6.54	.88	6.50	1.00	6.33	1.61
Exclusion	2.83	1.47	3.23	1.92	3.83	1.99	4.00	1.95
Self-Exclusion	4.00	1.86	4.69	1.25	5.58	1.62	4.67	1.50
Damage to Reputation	1.83	1.34	2.54	1.90	2.75	1.91	2.00	1.13
Unpleasantness	2.72	1.54	3.23	1.41	3.36	1.75	3.72	1.23
Threat to social order	3.46	1.44	4.15	1.46	4.63	1.33	4.67	1.29

Note. Higher values on Evaluation indicate more negative evaluation. 'Ingroup' denotes a target living in Jena, 'outgroup' denotes a target living in Dresden. *Recommended punishment severity* values are z-scores.

Discussion

In sum, on *Evaluation*, a descriptive tendency for an ingroup favoritism effect was found for the lighter offense, but a descriptive reversal for the heavier offense. No significant effects were found for the punishment measure. However, there was a rather large effect consistent with predictions for punishment in the heavier offense condition. Sample size was rather small, therefore it seemed worthwhile to trying to replicate the general pattern of means for the heavier offense found here and better protect it against chance fluctuations. Also, instead of a lighter offense, another heavy offense was used in the next study to increase generalizability in the event of replication.

9.4 Study P-4

Participants and Design

Eighty students of FSU participated in this study. They were approached around campus, if they agreed to participate in the study, they were given the questionnaire and filled it out by themselves. They were randomly given one of four questionnaires resulting from crossing the offender's *Group membership* (ingroup versus outgroup) and the *Offense* (lighter versus heavier). After returning the questionnaire they were thanked, given a candy bar for compensation and later debriefed by email.

One participant gave answers to the punishment recommendation items (see below) that were considerably out of the specified bounds and very unreasonable, so that his data was omitted from analyses. The final sample thus comprised $N = 79$ with a mean age of 22 years, $SD = 2$ years; 68% female. Cell sizes were $n = 20$ or $n = 19$.

Procedure

The questionnaire was a shortened version of the one used in Study P-3. It only contained the *Evaluation* on 11 bipolar adjective scales ($\alpha = .82$), the *Deservingness* item and the remaining items for the *Recommended punishment severity* scale (*Time in prison*, *Probation*, *Community service*, and *Fine*). The *Deservingness* item was changed in wording to 'Which intensity of a punishment do you think the offender deserves' and answered on a rating scale from 1('rather mild punishment') to 7 ('very harsh punishment'). The upper bound for the answer to the *Community service* item was increased to 800 hours. The *Recommended punishment severity* scale from averaged z-scores had acceptable reliability ($\alpha = .69$).

Regarding independent variables, the lighter offense in Study P-3 (burglary) was replaced by a very short description of a heavy assault to see whether the general tendency of the result

for the heavier offense from P-3 would generalize to a qualitatively different crime. The offender was described as either a student from Jena (ingroup) or from Dresden (outgroup).

This time, unlike in Study P-3, participants did not answer questions concerning their satisfaction at FSU first to keep materials as short as possible.

Results

Means and standard deviations of dependent measures are shown in Table 4.

None of the independent variables alone nor their interaction had an effect on *Evaluation*, all Fs(1,75) < 1.72, all ps > .19, all η_p^2s < .03. Effect sizes capturing ingroup versus outgroup differences on *Evaluation* for each offense separately were $d = 0.53$ and $d = 0.06$ for the burglary and assault, respectively.

Only *Offense* had a main effect on *Recommended punishment severity*, $F(1,75) = 4.60$, $p = .035$, $\eta_p^2 = .06$. Participants recommended higher punishment in the burglary case ($M = 0.16$, $SD = 0.63$) than in the assault case ($M = -0.17$, $SD = 0.68$). Neither the main effect of *Group membership* nor the interaction were significant, both Fs < 1. Effect sizes capturing ingroup versus outgroup differences on *Recommended punishment severity* for each offense separately were $d = 0.11$ and $d = -0.13$ for the burglary and assault cases, respectively.

Table 4

*Means and standard deviations of dependent variables in Study P-4 (*N = *79)*

	offense							
	burglary				assault			
Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Evaluation	5.40	0.75	5.02	0.69	5.33	0.67	5.28	.81
Recommended Punishment Severity	0.19	0.61	0.12	0.66	-0.21	0.57	-0.12	.80

Note. Higher values on Evaluation indicate more negative evaluation. *Recommended punishment severity* values are z-scores. 'Ingroup' denotes a target living in Jena, 'outgroup' denotes a target living in Dresden.

Discussion

The replication of the ingroup favoritism effect on punishment recommendation in the heavy offense condition from Study P-3 failed. Also, no similar pattern was found for evaluation. The next study was a second attempt to replicate the ingroup favoritism effect on punishment for a heavier offense. The sample were participants of a different study. In this other study, a control condition comprising a number of measures relating to stereotyping of East Germans by West Germans to measure a baseline, was considerably shorter than others. Participants in that condition therefore also partook in the present study.

9.5 Study P-5

Participants and Design

Participants were 29 students of FSU who filled out the questionnaire after completing materials for a control condition of an unrelated study. Participants were randomly assigned to one of two *Group membership* (ingroup versus outgroup) offender conditions, only one heavier offense was used here. Future participants were approached around campus and asked whether they would like to come to a seminar room for a study taking about 30 minutes. After

completing and returning the questionnaire they were thanked, paid €2 for compensation and later debriefed by email.

Three participants indicated on control questions at the end of the questionnaire that they had participated in this study before, their data was therefore omitted from analyses. Also, data from four participants were excluded because they did not answer correctly to manipulation check questions regarding the independent variable in the end of the questionnaire. The final sample comprised 22 participants (mean age = 22 years, *SD* = 2 years; 50% female).

Procedure

This study presented the same offense to all participants (burglary with €20000 damage from previous studies) and manipulated simply the offender's *Group membership (*student from Jena, ingroup, *n* = 12 versus Dresden, outgroup, *n* = 10).

Participants again answered questions about their study satisfaction as in previous studies and then completed the same questionnaire as in Study P-3.

Results

Means and standard deviations of *Evaluation* (α = .89) and *Recommended punishment severity* (α = .74) indices as well as of other dependent measures are given in Table 5.

Both *Evaluation* scores as well as *Recommended punishment severity* were unaffected by the offender's *Group membership*, both *F*s < 1. Effect sizes capturing ingroup versus outgroup differences on *Evaluation* and *Recommended punishment severity* were *d* = -0.01 and *d* = -0.14, respectively.

None of the differences for additional variables were significant, all *F*s < 1.

Table 5

*Means and standard deviations of dependent variables in Study P-5 (*N = *29)*

Group membership	Group membership			
	ingroup		outgroup	
	M	*SD*	*M*	*SD*
Evaluation	4.95	0.70	4.95	1.16
Recommended Punishment Severity	-0.04	0.62	0.05	0.75
Wrongness	6.17	1.12	6.30	1.06
Exclusion	3.00	1.71	3.50	2.01
Self-Exclusion	3.67	1.97	3.80	1.55
Damage to Reputation	2.25	1.55	1.90	1.45
Unpleasantness	3.17	1.51	3.33	1.64
Threat to social order	3.75	1.37	4.20	0.63

Note. Higher values on Evaluation indicate more negative evaluation. 'Ingroup' denotes a target living in Jena, 'outgroup' denotes a target living in Dresden. *Recommended punishment severity* values are z-scores.

Discussion

There were no indications of differential recommended punishment nor evaluation. The next study undertook a last try to find intergroup differences with the questionnaire used in Study P-3 and a large sample size.

9.6 Study P-6

Participants and Design

Study P-6 was identical to Study P-3 except that participants were seated in a separated area of a hallway at FSU and were compensated with a chocolate bar. One hundred and sixty students participated. Data from two participants was discarded because they indicated having completed the same questionnaire before, and an additional 19 data sets were omitted from

analyses because participants did not answer correctly to the manipulation check question for the offender's group membership at the end of the questionnaire. Excluded cases were evenly distributed across conditions, $\chi^2(df = 3, N = 60) = 2.30$, $p = .51$, $\varphi = .12$. The final sample comprised 139 participants (mean age = 22 years, *SD* = 3 years; 58% female) who, before exclusion of the discussed cases, had been randomly assigned to the four cells of a 2 (*Group membership*: ingroup vs. outgroup) × 2 (*Offense*: lighter vs. heavier) between participants design, with cell sized ranging from $n = 32$ to $n = 39$.

Procedure

The questionnaire was identical to the one from Study P-3. The procedure was identical to that of Study P-1.

Results

Means and standard deviations for dependent measures are shown in Table 6.

Evaluation scores ($\alpha = .88$) were subject to no main effects, both Fs < 1, but there was a significant *Group membership* × *Offense* interaction, $F(1,135) = 5.22$, $p = .024$, $\eta_p^2 = .04$. While *Evaluation* of the outgroup offender was more negative than that of the ingroup offender in the heavier case, $F(1,135) = 4.61$, $p = .034$, $\eta_p^2 = .03$, there was no such difference for the lighter offense, $F(1,135) = 1.20$, $p = .275$, $\eta_p^2 < .01$. Effect sizes capturing ingroup versus outgroup differences on *Evaluation* for each offense separately were $d = 0.27$ and $d = -0.51$ for the lighter and heavier offenses, respectively.

Recommended punishment intensity ($\alpha = .69$) was subject to a main effect of *Offense*, $F(1,135) = 5.67$, $p = .019$, $\eta_p^2 = .04$: Scores were higher for the heavier offense ($M = 0.13$, $SD = 0.68$) than for the lighter one ($M = -0.12$, $SD = 0.63$). There was no main effect of *Group membership*, $F < 1$, but a marginal interaction, $F(1,135) = 3.00$, $p = .090$, $\eta_p^2 = .02$. While

there was a tendency for milder punishment recommendation for the ingroup offender for the heavier offense, $F(1,135) = 2.44$, $p = .121$, $\eta_p^2 = .02$, there was no difference in the lighter offense scenario, $F < 1$. Effect sizes capturing ingroup versus outgroup differences on *Recommended punishment intensity* for each offense separately were $d = 0.21$ and $d = -0.37$ for the lighter and heavier offenses, respectively.

Additional measures. The additional index *Threat to Social Order* was computed as in previous studies ($\alpha = .80$). The item 'It would be embarrassing to be friends with the offender' from *Unpleasantness* of being connected with the offender considerably lowered the internal consistency of the scale (for all three items: $\alpha = .48$). Therefore, it was excluded, leaving a two item scale of *Unpleasantness* with good internal reliability ($\alpha = .88$). Results using the complete and the shortened scale do not differ in decisions based on statistical significance.

There was a significant main effect of *Group membership* on *Damage to reputation*, $F(1,135) = 5.69$, $p = .018$, η_p^2s = .04 and a marginal one on *Unpleasantness*, $F(1,135) = 3.38$, $p = .068$, $\eta_p^2 = .02$. Participants rated *Damage to reputation* higher in the case of an ingroup offender ($M = 2.92$, $SD = 1.59$) than of an outgroup offender ($M = 2.29$, $SD = 1.43$). For *Unpleasantness of being* connected with the offender, the difference was in the opposite direction: they expressed more unpleasantness about being connected to the outgroup ($M = 4.62$, $SD = 1.61$) than the ingroup offender ($M = 4.11$, $SD = 1.69$).

There was also a main effect of *Offense* on *Self-Exclusion,* $F(1,135) = 4.21$, $p = .042$, $\eta_p^2 = .03$. Participants believed that the offender had excluded himself to a higher degree if he had committed the heavier offense ($M = 5.46$, $SD = 1.57$) rather than the lighter one ($M = 4.90$, $SD = 1.64$).

There were no other main or interaction effects for additional variables, all Fs < 1.59, all ps > .21, all η_p^2 s < .02.

Table 6

*Means and standard deviations of dependent variables in Study P-6 (*N = *139)*

	offense							
	lighter (burglary €9000)				heavier (burglary €20000)			
Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Evaluation	5.43	0.67	5.21	0.89	5.21	0.93	5.63	0.74
Recommended Punishment Severity	-0.06	0.65	-0.20	0.60	0.02	0.61	0.26	0.75
Wrongness	6.62	0.78	6.52	.94	6.54	0.97	6.78	0.49
Exclusion	3.97	1.71	3.82	1.93	4.38	2.16	4.37	1.62
Self-Exclusion	4.71	1.53	5.09	1.76	5.49	1.52	5.44	1.65
Damage to Reputation	2.76	1.58	2.12	1.52	3.05	1.62	2.47	1.34
Unpleasantness	3.84	1.50	4.25	1.58	4.13	1.86	4.73	1.57
Threat to social order	4.87	1.44	4.50	1.62	4.68	1.38	4.91	1.14

Note. Higher values on *Evaluation* indicate more negative evaluation. 'Ingroup' denotes a target living in Jena, 'outgroup' denotes a target living in Dresden. *Recommended punishment severity* values are z-scores.

9.7 Study P-7

In Study P-7 another test of the hypothesis was undertaken with a different ingroup versus outgroup categorization (young versus old). It also used a large sample and only one heavy offense case different from those used so far: an old woman as a victim and €13000 robbed (see below).

Participants and Design

Participants were 100 students of FSU who participated in an experimental lab session and received €5 for their 30 minute participation (mean age = 22 years, *SD* = 2 years; 51% female). They all completed the same affective priming task involving no between participants manipulations before filling out the questionnaire presently reported. Thus, this

preceding task will not be discussed further here. Participants were randomly assigned to one of two *Group membership* (ingroup versus outgroup) conditions. After the session, participants were debriefed via email.

Procedure

Participants read the story of two men conning an old woman out of €13000 by one of the men pretending to be her son in law on the telephone. The story said that one of the two men was later apprehended and now to be tried. A few pieces of information were given about the offender: criminal record, employment status, age and familiar status. In both versions of the questionnaire, all information except age was identical. In order to manipulate *Group membership*, the offender was described to be either 23 years old (ingroup: younger) or 57 (outgroup: older). The oldest participant indicated to be 30 years old, participants can thus safely be said to be young rather than old.

Dependent measures asked for a *Severity judgment* of the offense (from 1 = 'not severe at all' to 9='very severe'), general *Harshness* of recommended punishment ('How harsh should the offender be punished ...', ranging from 1 = 'very mildly' to 9 = 'very harshly', *Harsh punishment)* and whether the punishment should *Hurt* him (from 1='no, not at all' to 9='yes, very much'). They also indicated how long a *Prison* term should be (1='0 months', 2='1-2 months', 3 = '3-4 months' etc. through 10='17-18 months'), whether the offender should receive probation (from 1='yes, definitely' to 9='no, not at all'), how many *Hour of community service* he should be sentenced to serve (from 0 to 100 hours in ten steps of 10 hours) and the amount of a monetary *Fine* they found appropriate (€0-€3000 in ten steps of €300). All but the first of these items were *z*-standardized and averaged to an index of *Recommended punishment intensity* (α = .65).

Results

Recommended punishment intensity did not differ significantly for the young (M = -0.06, SD = 0.64) versus old (M = 0.07, SD = 0.56) offenders, $F(1,98) = 1.26$, p = .265, $\eta_p^2 = .01$. The effect size for the difference between the young and the old offender was d = -0.22.

9.8 Study P-8

This study also aimed to test the Intergroup Punishment Difference Hypothesis but introduced the additional factor of positive affect associated with the target by way of association of participants name letters and birth dates with the target (see Jones, Pelham, Mirenberg, & Hetts, 2002; Kitayama & Karasawa, 1997; Nuttin, 1985). This was done in order to get a first idea of whether reactions to severe norm violations (i.e., crimes) are at all sensitive to initial differences in associated positive affect.

Also, the study was conducted in the context of an experimental session in a laboratory and after an affective priming task containing no between participants manipulation and not discussed in further detail here as well as a further unrelated experiment. This unrelated affective priming task and the experiment investigated reactions to exclusion in Cyberball, a computer based ball tossing game introduced by Williams, Cheung and Choi (2000) and contained a different manipulation of group membership (Eastern versus Western German interaction partner) than the one employed here (college student versus non college student, see below). Conditions of that unrelated (k = 2) and present (k = 4, see below) experiments were orthogonally crossed. Assignment to the eight resulting cells was random and resulted in evenly distributed numbers of participants in each cell of a design comprising conditions from both studies, $\chi^2(df = 3, N = 79) = 2.30$, $p = .509$, $\varphi = .17$.

All materials and questions were presented on a computer screen and participants gave their answers via mouse clicks.

Participants were presented with information about the offender and answered questions regarding their general liking for him before they learned that he had committed a crime. Therefore, this *Evaluation* measure closely modeled after that in the preceding studies, will not be included in analyses as in the other studies of this complex it was always placed after description of the transgression and therefore cannot be compared.

In the prior information they saw a fabricated profile of the offender from an on line dating community which, apart from the presently interesting manipulation of group membership (see below), also indicated his screen name as either containing individual participants' day of birth and the three first letters of his/her first name or not containing these strings. This factor, *Own characters*, will be kept as an independent variable in analyses, but is of lesser theoretical concern here.

Participants and Design

Seventy-nine students of FSU participated in this study (mean age = 23 years, *SD* = 3 years, 63 % male). They took part in an experimental session in a lab lasting 30 minutes and were paid €5 for compensation. They were debriefed by email shortly after they had come to the lab.

They were randomly assigned to one of four conditions of a 2 (*Own characters*: yes vs. no) × 2 (offenders *Group membership*: ingroup vs. outgroup) between participants design.

Procedure

Participants first saw a profile of a male person (the target) from an on line dating community which had been fabricated to closely resemble original profiles. In this profile, which otherwise contained rather inconspicuous information (e.g., personal food preferences, astrological sign etc.) they learned that the person presented was either also a college student

(ingroup condition) or had completed school in ninth grade and now worked full time as a car salesman (outgroup condition). Also, participants had earlier in the session indicated their first name and their day of birth. From this information, a screen name, that was printed in bold letters on the top of the profile view, was constructed in the form of DOMINIK_XXX_YY. For participants in the *Own characters* condition, XXX was replaced by the first three letters of the participants' first name and YY by their day of birth. For participants in the *No Own characters* condition, XXX and YY were chosen to be definitely different from their own three first name letters and day of birth, but randomly.

After reading the profile and rating the target regarding 12 adjectives (six positive and six negative ones[8]) on rating scales ranging from 1 ('not at all __') to 9 ('very __'), they read an article from an on line newspaper, that was also fabricated, but closely modeled in design after an original one. The article reported that the person from the on line community presented earlier had gotten acquainted with a woman in the on line community and after a few dates he anesthetized her during a diner in her apartment and robbed her apartment. As this offense involved robbing and also physical assault on the victim, it was coded as heavier for the meta-analysis below.

After reading the article, participants indicated to which degree they felt nine emotions upon reading of the crime (*anger*, *fear*, *outrage*, *fury*, *anxiety*, *indignation*, *rage*, *worry*, *irritated*[9]) on rating scales anchored by 1 ('I do not feel this emotion at all') and 9 ('I feel this emotion strongly'). Then participants indicated how harsh a punishment the offender should get (from 1 = 'very mild' to 9 = 'very harsh'), how long of a *Time in prison* (1 = 'very short' to 9 = 'very long') and how much *Community work* (1 = 'very little' to 9 = 'very much') he should

8 The positive adjectives were *angenehm (pleasant), gut (good), verlässlich (reliable), freundlich (friendly), friedlich (peaceful)*, and *höflich (polite)*; the negative adjectives were *naiv (naïve)*, *unehrlich (dishonest)*, *primitiv (primitive)*, *kalt (cold)*, *dumm (stupid)*, *langweilig (boring)*.

9 The German emotion terms used were *Ärger*, *Angst*, *Empörung*, *Wut*, *Furcht*, *Entrüstung*, *Zorn*, *Besorgtheit*, *Aufgebrachtheit,* respectively.

Table 7

*Means and standard deviations of dependent variables in Study P-8 (*N = *79)*

	Own characters							
	yes				no			
Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Offensive emotions	6.00	2.00	5.09	2.08	6.60	1.53	6.14	1.89
Defensive emotions	3.68	1.61	3.68	1.80	3.72	1.71	4.02	2.14
Recommended punishment intensity	6.95	0.95	7.03	1.09	6.72	1.36	6.61	1.37

Note. Means and standard deviations for *Offensive* and *Defensive emotions* are unadjusted, while results in the text are from analyses introducing the respective other as a covariate.

be sentenced to, what they thought would be appropriate as a *Fine* (1 = 'small amount' to 9 = 'high amount') and whether the offender should be let off on *Probation* (1 = 'yes, definitely' to 9 = 'no, definitely not'). These items pertaining to punishment formed a satisfactory scale (α = .71) and were averaged for the index of *Recommended punishment intensity*.

Results

Results from the evaluative rating of the offender on the adjective scales will not be discussed here, as this measurement was taken before participants knew about the offense. Means and standard deviations for the remaining dependent measures are shown in Table 7.

The emotion items were selected to cover the facets anger (*anger*, *fury*, *rage*), fear (*fear*, *anxiety*, *worry*) and moral outrage (*outrage*, *indignation*, *irritation*). A factor analysis with varimax rotation however revealed only two factors. The first factor explained 45.52% of the variance and all anger and outrage items loaded on it higher than .75. The second factor (25.13% of the variance) comprised the fear items *fear* and *anxiety* (loadings = .95 and .91, respectively), and *worry* loaded weakly on both factors (.35 on Factor 1 and .47 on Factor 2).

Therefore, all anger and outrage items were averaged to an index of *Offensive emotion* (α = .93) and *fear* and *anxiety* for an index of *Defensive emotion* (α = .95). The two scales were still considerably correlated, $r(79) = .37$, $p < .001$, therefore the analyses for both emotion scales use the respective other as a covariate.

Both *Group membership* as well as *Own Characters* had marginal main effects on *Offensive emotions,* $F(1,74) = 3.31$, $p = .073$, $\eta_p^2 = .04$ and $F(1,74) = 2.80$, $p = .099$, $\eta_p^2 = .04$, respectively. *Offensive emotions* were higher for the ingroup ($M = 6.31$, $SE = 0.28$) than for the outgroup offender ($M = 5.57$, $SE = 0.29$) and higher if the offenders screen name did not contain *Own characters* ($M = 6.28$, $SE = 0.31$) than if it did ($M = 5.60$, $SE = 0.27$). There was no interaction, $F < 1$.

As could be expected from the intercorrelation of the two emotion scales, the covariate *Defensive emotions* was significant, $F(1,74) = 11.67$, $p = .001$, $\eta_p^2 = .14$. Apart from the significant effect of the covariate *Offensive emotions*, there were no main or interaction effects on *Defensive emotion*, all Fs(1,74) < 1.46, all ps > .23, all η_p^2s < .02. No main or interaction effects on *Recommended punishment severity* emerged, all Fs(1,75) < 1.45, all ps > .23, all η_p^2s < .02. The effect size capturing the ingroup versus outgroup difference in *Recommended punishment severity* irrespective of *Own characters* was $d = -0.02$.

Discussion

This last reported study on simple punishment differences followed a different setting, but was conceptually very similar to the preceding studies. Therefore, it was concluded in the meta-analysis (see below). No effect of *Group membership* on punishment reactions, which were measured on traditional rating scales rather than with open format answer options as in preceding studies, was found. Emotional tendencies, measured here for the first time, even pointed to an effect of *Group membership* opposite to that predicted.

10 Meta-Analysis on Studies Investigating the Intergroup Punishment Difference Hypothesis

10.1 Results

Effect sizes for *Evaluation* and *Recommended punishment intensity* of studies P-1 through P-8 were separately summarized with Computer Programs for Meta-Analysis 5.3 (Schwarzer, 1989) using the Weighted Integration Method proposed by Hedges and Olkin (1985).

Consistent with the general remarks on the meta-analyses made earlier, effect sizes were always computed using the difference $M_{ig} - M_{og}$, such that negative values indicate a bias in favor of the ingroup and positive values a bias in favor of the outgroup, as on both variables higher values indicate more negative evaluation and harsher punishment, respectively. The effect sizes analyzed are shown in Table 8.

The mean effect size for *Evaluation* was $d_m = 0.07$, $SE = 0.11$, $p = .261$. The same result was obtained after correction for attenuation due to imperfect reliability, $d_m = 0.09$, $SE = 0.11$, $p = .210$. Thus, across all individual studies and levels of offense severity, there was no intergroup effect on *Evaluation* ratings. If anything, the descriptive difference pointed in the opposite direction, an evaluative bias in favor of the outgroup (see the positive sign of the mean effect).

Separate analyses of effect sizes from lighter on the one hand and heavier offenses on the other yielded $d_m = 0.11$, $SE = 0.18$, $p = .263$ and $d_m = 0.04$, $SE = 0.14$, $p = .374$, respectively, without correction for imperfect reliability, and $d_m = 0.12$, $SE = 0.18$, $p = .250$ and $d_m = 0.07$, $SE = 0.14$, $p = .308$, respectively,with that correction. The *Z*-test for moderation proposed by

DeCoster (2004; see p. 45) yielded $Z = 0.31$, $p = .756$ and thus no evidence of *Evaluations* being a function of offense severity ($Z = 0.23$, $p = .815$ for corrected effect sizes).

The hypothesis that the severity of the transgression would moderate a difference in evaluative reaction between ingroup and outgroup targets therefore must be rejected.

Effect sizes for *Recommended punishment intensity* averaged to $d_m = -0.09$, $SE = 0.08$, p = .131, without correction, and $d_m = -0.11$, $SE = 0.08$, $p = .084$, with correction. Effect sizes for milder offenses did not differ significantly from zero, $d_m = 0.06$, $SE = 0.15$, $p = .354$, but those for heavier offenses did so marginally, $d_m = -0.15$, $SE = 0.10$, $p = .060$ ($d_m = 0.07$, $SE = 0.15$, $p = .327$ and $d_m = -0.18$, $SE = 0.10$, $p = .029$, respectively, for effect sizes corrected for imperfect reliability). The *Z*-test for moderation however showed no moderation of the effect by severity of the offense, $Z = 1.14$, $p = .253$ for uncorrected and $Z = 1.39$, $p = .164$ for corrected effect sizes.

Thus, while overall, the simple punishment hypothesis receives weak support in the meta-analysis integrating studies P-1 through P-8 (see the marginal difference if effect sizes are corrected for imperfect reliability), it seems that for heavier offenses there is a small difference in favor of ingroup targets. They on average received less intense punishment recommendations than outgroup targets. To be sure, the moderation test is not significant.

It may seem odd that the mean effect size for heavier offenses is significantly different from zero, but not from the slightly positive one for lighter offenses, but looking at the variability of the effect sizes in the meta-analysis, it is clear that effect sizes for lighter offenses are more heterogeneous than those for heavier offenses. This may be due to the smaller number of available effect sizes from lighter offense scenarios than from heavier offense scenarios discussed earlier. Alternatively, inspection of individual effect sizes in Table 8 shows, that the effect sizes from Study P-2 are the largest inconsistent ones with the

mean effect size of the respective subset, more so for the lighter offenses than the heavier offenses. Effect sizes from Study P-2 may thus be responsible for the rather large variability in the subsets. The striking inconsistency of effect sizes from Study P-2 with the remaining ones may be due to the intergroup distinction between a German born in Germany versus one who was born in Kazachstan, a so called 'Deutschstämmiger', used in this study. Grounded in a rather old-fashioned law of nationality in Germany, persons being able to prove that one of their ancestors was German ('Deutschstämmige') but not knowing the language and never having been to Germany may receive German citizenship relatively easy (*ius sanguis,* Rabkov, 2006). Since the disintegration of the Soviet empire and communist regimes in Central Asia as well as Eastern Europe, many of such Deutschstämmige came to Germany to live a better life than they could in the countries where they had been born. The integration process was and is sometimes difficult and a stereotype of them as criminal and aggressive evolved (Rabkov, 2006). Thus, the intergroup distinction of this study included a stereotype that may have flooded the mere categorization effect proposed and examined here, possibly by way of expectancy violation effects (Kernahan, Bartholow, & Bettencourt, 2000). According to such an interpretation, the commission of an aggressive crime such as the heavier one in Study P-2 by a German born violates expectations more than that by a non German born. This expectancy violation would elicit negative affect larger than the one presumably associated to him as an ingroup member and beyond the negative affect from the transgression. The net affect would then be more negative than for the non German born offender, resulting in higher punishment recommendation. For the lighter offense in Study P-2, the expectation could be in the opposite direction (i.e., more expectancy violation for the non German born and thus more additional negative affect) and therefore, the difference is in favor of the German born offender.

Table 8

Overview of effect sizes from ingroup versus outgroup comparisons in Evaluation *and* Recommended punishment intensity *from punishment studies P-1 to P-8.*

Measure	Study	Offense	d (ig – og) *	n_{ig}	n_{og}	Reliability
Recommended punishment intensity	P-1	lighter	0.33	24	24	.74
	P-2	lighter	-0.70	14	14	.69
	P-3	lighter	-0.10	13	12	.58
	P-6	lighter	0.21	39	32	.61
		Mean:	d_m = 0.06, *SE* = 0.15, *CI:* [-0.24, 0.36]			
	P-1	heavier	-0.07	24	22	.74
	P-2	heavier	0.20	12	12	.69
	P-3	heavier	-0.68	12	12	.58
	P-4	heavier	0.11	20	20	.69
	P-4	heavier	-0.13	20	19	.69
	P-5	heavier	-0.14	12	10	.74
	P-6	heavier	-0.37	35	33	.61
	P-7	heavier	-0.22	51	49	.65
	P-8	heavier	-0.02	40	39	.73
		Mean:	d_m =-0.15, *SE* = 0.10, *CI:* [-0.34, 0.04]			
		Overall mean:	d_m = -0.09, *SE* = 0.08, *CI:* [-0.25, 0.07]			
Evaluation	P-2	lighter	0.18	14	14	.76
	P-3	lighter	-0.39	13	12	.88
	P-6	lighter	0.27	39	32	.88
		Mean:	d_m = 0.11, *SE* = 0.18, *CI:* [-0.24, 0.47]			
	P-2	heavier	0.33	12	12	.76
	P-3	heavier	0.59	12	12	.88
	P-4	heavier	0.53	20	20	.82
	P-4	heavier	0.06	20	19	.82
	P-5	heavier	-0.01	12	10	.89
	P-6	heavier	-0.51	35	33	.88
		Mean:	d_m = 0.04, *SE* = 0.14, *CI:* [-0.22, 0.31]			
		Overall mean:	d_m = 0.07, *SE* = 0.11, *CI:* [-0.14, 0.28]			

Note. Reliability is Cronbach's α. Means are calculated according to the Weighted Integration Method (Hedges & Olkin, 1985) without correction for imperfect reliability.
* Positive values indicate more negative reaction to the ingroup than the outgroup, negative values more negative reaction to the outgroup than the ingroup

10.2 Discussion

Overall, the hypothesis that group membership (ingroup versus outgroup) would make a difference for punishment recommendation is mildly supported (see the marginal overall effect size). Presumably, if a certain amount of negative affect is elicited by the offense (heavier offenses), then an ingroup bias emerges. This ingroup bias however is not on positive dimensions, such as allocation of positive points or positive evaluation of group products, but on a negative dimension, namely the recommendation of harmful behavior in the form of punishment for criminal behavior. Thus, at least in a context, for which – according to the theoretical ideas concerning punishment advanced earlier – judgment is clearly affectively laden, the predicted bias in favor of an ingroup target emerges. In the context of a lighter offense however, the difference is not significant and descriptively even in the opposite direction.

Evaluations are, according to presently reported results, not affected by group membership. If anything, effect sizes descriptively are positive, and thus suggest a tendency for a BSE, but this tendency is not even close to conventional levels of significance.

The values of the mean effect size for *Recommended punishment intensity* do point in a negative direction, especially for heavier offenses, while those from evaluation rather tend towards a positive value. Thus, whereas for punishment tendencies, a tendency for a bias in favor of the ingroup targets emerged, on evaluations, a descriptive tendency for an outgroup favoritism effect may be looming. Possibly – and this presents a worthwhile matter to examine in the future – there is a dissociation of simple evaluation of a norm-violating target on one hand and the legitimacy of harmful behavior against that violator on the other. Evaluations may be more sensitive to presentation concerns and still trigger legitimacy issues, leading to the elimination of discrimination tendencies just as in evaluations of targets in a

neutral situation, that is of which no behavior was known (see Mummendey & Otten, 1998). Descriptively, on evaluations, the difference even seems to be more consistent with outgroup favoritism than with ingroup favoritism.

Future research should refine the measure of punitive tendencies and definitely expect very small effect sizes *a priori*. As evident from the overview in Table 8, reliability of the *Recommended punishment severity* were lower than those of *Evaluation*. This is partly due to the refinement process of the measure throughout the studies in which the individual items were modified in response to skewed distributions, presented in an open-answer format, restricted in range and finally again to specific anchors as the open answer formats yielded quite large standard deviations on individual items. Possibly this is also a reason for why punitive tendencies tended into the direction of an ingroup favoritism: Assigning judgments about appropriate amounts of punishment may be a rather unusual task for lay people (low reliability is not necessarily, but could be an indication of that). It could therefore be open to the influence of affective biases such as the one hypothesized to be operating in the situations employed in the studies of this complex – more so than evaluations of individuals which are made on an everyday basis and quite well practiced. Thus, the presumed reason for the low reliability of punishment judgments in the studies presently reported – lack of experience with them –, may turn out to be an advantage for research targeting punitive tendencies and their susceptibility to bias.

The unequal number of studies using lighter and heavier offenses in the scenarios included in the meta-analysis deserves to be mentioned. Effect size estimates for *Recommended punishment severity* for heavier offenses are based on a total of $N = 442$ cases across the eight studies, while for for the lighter offenses the total number of cases is $N = 172$. A similar disparity holds for *Evaluations* ($N = 217$ versus 124). The latter is due to the fact

that materials were deliberately kept short in the first and the last two studies (which accounts for exclusion of the evaluation measures). The unequal number across the lighter versus heavier distinction and arises because the focus, especially in the later studies of this complex, was on heavier offenses to get a picture whether the difference holds for a stronger stimulus at all after the individual studies rarely found significant differences. Also, the special attention given to heavier offenses coincides with the assumption that a BSE (see Chapter 5) could emerge for milder offenses, but the predicted difference in favor of ingroup targets for heavier ones (see the hypothesis in 6.2). The effects for the study subsets of the meta-analysis which were underrepresented (i.e., lighter offense for *Recommended punishment severity*, lighter and heavier offense for *Evaluation*) were probably not non-significant because of insufficient sample size, they all have the opposite sign of the difference predicted. Therefore, notwithstanding the desirability to have an equally large evidence bases for all levels of a moderator, the general validity of the present results (which are inconsistent with the original hypothesis of a *general* difference in favor of the ingroup) is not considered to be jeopardized.

11 Studies Investigating the Reversed Black Sheep Effect Hypothesis

This chapter reports studies investigating the moderation hypothesis advanced in Chapter 5 and reiterated in Section 6.2. It is predicted that a BSE may emerge for transgressions if before learning about the transgression, the ingroup and outgroup differ in their general propensity to show the behavior in question (later operationalized in the *distinct* or *distinct plus* conditions). However, if no such difference on the group level exists (later operationalized as *non-distinct* conditions), there will be a reversed BSE, that is an intergroup bias in favor of the ingroup.

The following studies all follow a very similar scheme: The main dependent measure used is an evaluation scale which is commonly used in BSE studies in which participants rate the target on a number of bipolar adjective scales of clear valence. To test the prediction that a BSE will emerge for a distinctive norm, but not for a non-distinctive norm, distinctiveness was manipulated by informing participants that either the behavior in question was more common in the ingroup than the outgroup (distinctive norm) or that it occurred at basically the same rate in both groups. The behavior presented as having been committed by the individual target was cheating in an exam for which there is social consensus that it is a norm violation for which one deserves punishment – albeit mild forms such as an automatic failing grade instead of, say, incarceration. More details are given within the descriptions of the individual studies.

11.1 Study B-1

Participants and Design

Participants were 90 students of FSU in the state of Thüringen. One of two female experimenters approached participants while they were sitting in cafeterias or study spaces

around campus and asked them whether they would fill out a brief questionnaire for a chocolate bar. If they agreed, they were given a questionnaire from a pile containing questionnaires of all four conditions of a 2 (*Group membership*: ingroup vs. outgroup) × 2 (*Norm distinctiveness*: non-distinct vs. distinct) between participants design (see below) in a random order and left alone. When participants had finished, the questionnaire was recollected by the experimenter who thanked them and handed them the candy bar. They were later debriefed by email.

Two of the participants indicated on the last page of the questionnaire that they were students of a university in Sachsen (outgroup condition, see below). Data from these two participants was omitted from analyses, leaving a final sample of $N = 88$ (mean age: 23 years, $SD = 3$ years, 59% female). Cell sizes varied between $n = 20$ and $n = 24$.

Procedure

The questionnaire consisted of a cover sheet explaining the research in very broad terms, assuring anonymity of responses and thanking the participants, followed by the text containing the manipulation on one page and then dependent measures.

The manipulations of the independent variables were embedded in a fabricated article from a made up college student Internet portal. The article reported a study on cheating in college exams which had (supposedly) found that 12% of all students in Germany indicated in a survey that they had cheated at least once on a University exam.

Norm distinctiveness was varied by the information that students from the state of Thüringen and Sachsen differed (distinct condition) or did not differ (non-distinct condition) in their responses. In the distinct condition, it was reported that 4.3% of the students from Thüringen reported having cheated on a University exam in the past and 83.2% indicated that they categorically disapproved of cheating as a means to get good grades. For Sachsen the

percentages were 22.6% (for reporting having cheated) and 58.7% (for categorically disapproving of cheating). In the non-distinct condition the percentages for cheating and disapproving of cheating were reported not to differ between Thüringen and Sachsen (12.0% vs. 12.1% and 73.1% vs. 71.8%, respectively).

The target to be evaluated was a student who was quoted saying that he had cheated by help of a mobile telephone and his friends outside the building. *Group membership* of the violator was manipulated by describing this target either as a student from Friedrich-Schiller-Universität in Jena (ingroup condition) or Technische Universität in Dresden in Sachsen (outgroup condition).

Evaluation of the violator was measured on ten seven-point bipolar adjective rating scales with the positive anchor at 1 and the negative anchor at 7. The trait pairs used were *pleasant/unpleasant*, *good/bad*, *reliable/unreliable*, *friendly/unfriendly*, *orderly/messy*, *ethical/unethical*, *careful/reckless*, *hard-working/lazy*, *peaceful/aggressive* and *polite/impolite*[12]. Also, the participants gave a global impression of the violator on a 7-point bipolar rating scale from *positive* to *negative*. These 11 items formed a reliable scale (Cronbach's $\alpha = .86$) and were averaged to a *Person evaluation* index with higher values indicating more negative evaluation.

Evaluation of the violator's behavior was given by indicating how adequate participants found five adjectives for description of the violator's behavior (*false*, *scandalous*, *unacceptable*, *outrageous*, *mean*; German: *falsch*, *skandalös*, *unakzeptabel*, *ungeheuerlich*, *gemein*). They formed a reliable scale ($\alpha = .82$) and were averaged for an index of *Behavior evaluation* with higher values indicating more negative evaluation.

12 The german pairs used were *angenehm/unangenehm*, *gut/schlecht*, *verlässlich/unverlässlich*, *freundlich/unfreundlich*, *ordentlich/unordentlich*, *moralisch/unmoralisch*, *behutsam/rücksichtslos*, *fleißig/faul*, *friedlich/aggressiv*, *höflich/unhöflich.*

Table 9

*Means and standard deviations of dependent variables in Study B-1 (*N = *88)*

	Norm distinctiveness							
	specific				non-specific			
Target's Group membership	ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Person evaluation	4.68	1.03	4.56	0.76	4.26	0.77	5.03	0.95
Behavior evaluation	3.98	1.08	3.78	1.43	3.98	0.94	4.40	1.48

Note. Higher values on Evaluation indicate more negative evaluation. 'Ingroup' denotes a target living in Jena, 'outgroup' denotes a target living in Dresden.

Results

Person and *Behavior evaluation* were highly correlated, $r(88) = .57$, $p < .001$. They were nevertheless analyzed separately. Means and standard deviations are shown in Table 9.

Person evaluation. *Norm distinctiveness* by itself had no effect on *Person evaluations*, $F < 1$. Group membership had a marginal main effect, $F(1,84) = 2.92$, $p = .091$, $\eta_p^2 = .03$, which was however qualified by a significant interaction, $F(1,84) = 5.48$, $p = .022$, $\eta_p^2 = .06$. The outgroup violator was rated considerably more negative than the ingroup violator in the non-distinct condition, $F(1,84) = 8.59$, $p < .01$, $\eta_p^2 = .09$, while there was no difference in the distinct condition, $F < 1$. Effect sizes capturing the ingroup versus outgroup differences on *Person evaluation* for the distinct and non-distinct conditions were $d = 0.14$ and $d = -0.87$, respectively.

Behavior evaluation. Evaluation of the described behavior was affected by neither factor alone nor the interaction, all Fs$(1,84) < 1.36$, all ps $> .24$, all η_p^2s $= .02$. Effect sizes capturing the ingroup versus outgroup differences on *Behavior evaluation* for the distinct and non-distinct conditions were $d = 0.16$ and $d = -0.34$, respectively.

Discussion

In sum, there was a difference in favor of the ingroup violator on *Person evaluation* if the groups were described to be similar in their general tendency to violate the norm of cheating. No such difference was found if the ingroup was described as actually tending less to such violation in general. The latter result does not constitute a BSE as predicted for this condition, but relative to the non-distinct condition, the null effect is consistent with the idea that the threat to existing positive ingroup distinctiveness by an ingroup member rather than an outgroup member triggers processes working against ingroup favoritism in the direction of its reversal. No effect resulted for *Behavior evaluation*, but this measure was administered after *Person evaluation*. The effects could thus have dissipated by the time participants answered to the *Behavior evaluation* items. Also, as the measure in original BSE research is a person evaluation measure, the focus will remain on this index.

The aim of the next study was a replication of the present results and additionally to test whether, if the threat to positive ingroup distinctiveness is increased, the pattern of intergroup difference actually reverses to a genuine BSE. The norm violation used here, cheating in an exam, is by itself admittedly not quite severe. Evidence of this comes from control questions on the last page of the questionnaire of Study B-1: Participants indicated whether they have cheated before in an exam before and, more specifically, whether they had cheated in an exam in high school, before attending university. Answer options for both questions were 'never', 'once', and 'several times'. Over 88% of all participants admitted to have cheated at least once in general or in high school. Thus it seems that while cheating is a violation of prescriptive norms and an in principle sanctionable behavior, it is quite common and descriptively not that counter-normative after all. The value of positive ingroup distinctiveness (i.e., a weaker proclivity to cheat on an exam) may thus be limited and consequently the threat caused by

such a violation may be reduced. Therefore, in the next study one more level of *Norm distinctiveness* was added, where the gains from positive ingroup distinctiveness were increased by raising the attractiveness of the difference in cheating tendencies. The article for this condition was similar to the one presented in the distinct condition, but additionally stated that students from Thüringen were the most honest in all of Germany as they had the lowest rate of cheaters among its states. If this actually increases the value flowing from positive ingroup distinctiveness, it should also increase the threat caused by a deviating ingroup member and show in the form of a BSE in evaluations..

Another possible explanation for the lack of a BSE in the distinct condition could be that presenting the ingroup as less cheating than the outgroup in the article may have induced a positive stereotype of the ingroup compensating for the suppression of evaluation. In other words, the information that members of the ingroup as a whole cheat less may have created a situational positive schema of that ingroup to which evaluation of the individual target was assimilated.[13]

Thus, devaluation of the ingroup target and the positive ingroup stereotype could have canceled each other out, resulting in no difference between the ingroup and outgroup target. In order to control for that, after the *Person evaluation* both ingroup and outgroup as a whole were evaluated on the same adjectives. Introducing the group ratings corresponding to the targets group membership as a covariate in *Person evaluations* should then reveal a tendency for a BSE in the distinct condition.

13 Note that past research on the BSE rarely used *a priori* strongly valenced behaviors. The nature of the presently used, clearly valenced norm violation however logically implies that the ingroup as a category is perceived more positive than the outgroup if there is a non-negative intergroup distinctiveness for the ingroup.

11.2 Study B-2

Participants and Design

One hundred and twenty one students of FSU participated in this study (mean age = 23 years, *SD* = 2 years, 56% female). They were recruited around campus (e.g., in cafeterias) and filled out the questionnaire by themselves. Upon returning the questionnaire, they received a chocolate bar for compensation, were thanked and later debriefed by email.

Participants received a questionnaire from a randomized pile of questionnaires containing six different versions corresponding to a 2 (target's *Group membership:* ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct vs. distinct vs. distinct plus) between participants design.

Procedure

Participants received the same questionnaire as in Study B-1 with the following modifications. First, after rating the target (*Person evaluation*, $\alpha = .86$) and his behavior (*Behavior evaluation*, $\alpha = .85$), they rated the entire ingroup and outgroup (always in that order) on the same adjective scales ($\alpha = .93$ and $\alpha = .94$ for *Ingroup* and *Outgroup evaluation*, respectively). Also, beside the distinct and non-distinct manipulations employed in Study B-1, a third *Norm distinctiveness* condition was added, henceforth referred to as the distinct plus condition. The article in that condition was similar to the distinct condition in Study B-1 but additionally mentioned that, according to the study reported, the students from Thüringen were also those cheating least and disapproving of cheating most among all the states of Germany.

Table 10

*Means (and standard deviations in parentheses) of dependent variables in Study B-2 (*N = *121)*

	Norm distinctiveness					
	non-distinct		distinct		distinct plus	
	ingroup	outgroup	ingroup	outgroup	ingroup	outgroup
Person evaluation	4.90 (0.86)	4.50 (0.62)	4.16 (0.90)	4.70 (0.71)	4.52 (0.71)	4.92 (0.67)
Behavior evaluation	4.26 (1.78)	4.31 (1.24)	4.19 (1.19)	4.22 (0.98)	4.23 (1.27)	4.88 (1.26)
Ingroup evaluation	3.47 (0.95)	3.14 (0.88)	2.99 (0.71)	2.81 (0.73)	2.93 (0.71)	3.14 (0.94)
Outgroup evaluation	3.61 (1.08)	3.27 (0.83)	3.75 (0.65)	4.48 (0.93)	3.99 (0.72)	4.27 (0.68)

Note. Higher values indicate more negative evaluation.

Results

Means and standard deviations of dependent variables and the prospective covariates (evaluation of the entire groups) are displayed in Table 10. *Person* and *Behavior evaluations* were significantly correlated, $r(121) = .44$, $p < .001$. However, as Study B-1 revealed different patterns of results for these two measures, they were again analyzed separately.

Person evaluation. Both the target's *Group membership* and *Norm distinctiveness* main effects were not significant, $F(1,115) = 1.74$, $p = .190$, $\eta_p^2 = .02$ and $F(2,115) = 1.87$, $p = .159$, $\eta_p^2 = .03$, respectively. The interaction of these two factors was however significant, $F(2,115) = 4.63$, $p = .012$, $\eta_p^2 = .08$. While in the non-distinct condition, there was a tendency for a BSE: the ingroup target was evaluated marginally more negatively than the outgroup target, $F(1,115) = 2.89$, $p = .092$, $\eta_p^2 = .02$, the pattern was reversed in the distinct and distinct plus conditions, $F(1,115) = 5.27$, $p = .023$, $\eta_p^2 = .04$ and $F(1,115) = 2.87$, $p = .093$, $\eta_p^2 = .02$,

respectively (see Table 10 for means). Thus, when a difference in the general tendency of ingroup and outgroup to cheat in exams was reported in the article, the ingroup target profited from that positive distinctiveness in that he was evaluated less negatively than an outgroup target. Note that as far as the distinct and non-distinct conditions are concerned, this result is partly opposite to that of Study B-1. Effect sizes capturing the ingroup versus outgroup difference in *Person evaluation* for the non-distinct, distinct, and distinct plus conditions were $d = 0.54$, $d = -0.73$, $d = -0.53$, respectively.

Behavior evaluation. There were no main effects of *Group membership* or *Norm distinctiveness*, nor an interaction effect on *Behavior evaluation*, all Fs = 1.04, all ps > .31, all $\eta_p^2 < .02$. Effect sizes capturing the ingroup versus outgroup difference in *Behavior evaluation* for the non-distinct, distinct, and distinct plus conditions were $d = -0.04$, $d = -0.02$, and $d = -0.49$ respectively.

Ingroup and *Outgroup evaluations*. *Ingroup* and *Outgroup evaluations* were substantially correlated, $r(121) = .31$, $p < .001$. A 2 (target's *Group membership*: ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct vs. distinct vs. distinct plus) × 2 (*Evaluated group*: ingroup vs. outgroup) mixed model ANOVA with repeated measures on the latter factor revealed that the ingroup as a whole ($M = 3.08$, $SD = 0.83$) was in general rated less negative than the outgroup ($M = 3.89$, $SD = 0.90$), $F(1,115) = 102.73$, $p < .001$, $\eta_p^2 = .47$. There were no main effects of target's *Group membership* or *Norm distinctiveness,* both Fs < 1, both ps > .39, both η_p^2s < .02. The *Group membership* × *Norm distinctiveness* interaction was marginal, $F(1,115) = 2.43$, $p = .093$, $\eta_p^2 = .04$, all other first-order and the second-order interactions were significant (see Table 11 for details). Figure 1 shows the means in graphic format.

Table 11

Results of 2 (target's Group membership*: ingroup vs. outgroup) ×3 (*Norm distinctiveness*: non-distinct, distinct, distinct plus) ×2 (*Evaluated group*: ingroup vs. outgroup) mixed model ANOVA with repeated measures on the last factor in Study B-2 (*N = *121)*

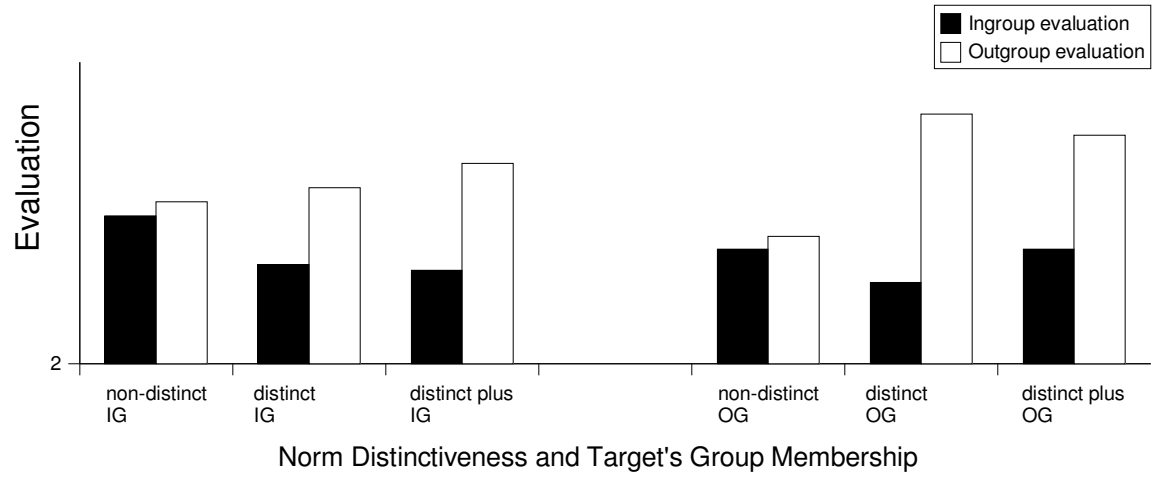
Source	df	F	p	η_p^2
Group membership	1	0.27	.606	<.01
Norm distinctiveness	2	0.93	.399	.02
Evaluated Group	1	102.73	< .001	.47
Group membership × Norm distinctiveness	2	2.43	.093	.04
Group membership × Evaluated Group	1	3.94	.049	.03
Norm distinctiveness × Evaluated Group	2	18.24	< .001	.24
Group membership × Norm distinctiveness × Evaluated Group	2	3.36	.038	.06

Note. Error terms of all F-values have df = 115.

The *Norm distinctiveness × Evaluated Group* interaction attests to the effectiveness of the *Norm distinctiveness* manipulation. Presenting participants with information that members of

Figure 1

Means of Group Evaluations *as a function of target's* Group membership *and* Norm distinctiveness



the outgroup generally tend to violate a norm more than ingroup members seems to be enough to induce a generally less positive view of that outgroup. While there was no difference in *Group evaluations* in the non-distinct condition, $F(1,115) = 0.87$, $p = .354$, $\eta_p^2 < .01$, the ingroup as a whole was evaluated considerably more positive than the outgroup in both the distinct as well as in the distinct plus conditions, $F(1,115) = 75.95$, $p < .001$, $\eta_p^2 = .40$ and $F(1,115) = 62.90$, $p < .001$, $\eta_p^2 = .35$, respectively. This sensitivity is also evidence that while participants may not view the norm-violating behavior as very negative for themselves, it still makes a difference in their perceptions of the ingroup and outgroup.The target's *Group membership* × *Evaluated Group* interaction conceptually represents the fact that participants' group ratings were sensitive to the report of an individual group member violating the norm. If the target was an ingroup member, the ingroup ($M = 3.12$, $SD = 0.82$) was evaluated more positively than the outgroup ($M = 3.79$, $SD = 0.83$), $F(1,115) = 34.06$, $p < .001$, $\eta_p^2 = .23$, but this difference was accentuated if the quoted violator was an outgroup member, $M = 3.04$ ($SD = 0.86$) for the ingroup and $M = 4.00$ ($SD = 0.97$) for the outgroup, $F(1,115) = 71.69$, $p < .001$, $\eta_p^2 = .38$. Participants hence seem to take into account the behavior of a single member of at least one of the two groups when judging the entire group.

The marginal *Group membership* × *Norm distinctiveness* interaction is also of interest here, although in a rather indirect manner. It represents the fact that group evaluations in general tended to be less positive if the individual norm violator was an outgroup member and an initial difference in the general propensity to cheat is reported, $F(1,115) = 1.56$, $p = .214$, $\eta_p^2 = .01$ and $F(1,115) = 1.32$, $p = .252$, $\eta_p^2 = .01$ for the distinct and distinct plus conditions, respectively, while in the non-distinct condition, both *Group evaluations* tended to be less positive if the individual norm violator was an ingroup member, $F(1,115) = 2.42$, $p = .137$, $\eta_p^2 = .02$. It thus seems that the behavior of an individual outgroup member triggers more

negative affect than an individual ingroup member (coloring subsequent evaluations of both ingroup and outgroup) in a situation in which the ingroup as a whole is portrayed as more norm abiding than the outgroup. In contrast, an individual ingroup member violating the norm triggers more such general negative affect than an outgroup violator if no *a priori* positive ingroup distinctiveness is established. This is inconsistent with the assumption of the theoretical treatments regarding the BSE that a violation by an ingroup member is especially aversive if it threatens an already established intergroup difference in favor of the ingroup.

In order to test the idea that the evaluation of individual violators is assimilated to the image created of their entire group by way of the *Norm distinctiveness* manipulation, a variable was created that contained the evaluation of the group of which the target in each case was a member and introduced as a covariate into the analysis of *Person* and *Behavior evaluations* reported above. This variable will henceforth be referred to as *Target group evaluation*. If indeed the group stereotype induced by the *Norm distinctiveness* manipulation caused the interactive effects of *Group membership* and *Norm distinctiveness* such that an ingroup target was evaluated more negatively than an outgroup member precisely because his entire group was rated less negatively than the outgroup, then the introduction of *Target group evaluation* should result in a pattern closer to the predicted pattern of outgroup derogation in the distinct conditions and a BSE in the non-distinct condition. However, the introduction of the covariate *Target group evaluation* for *Person evaluations*, which was itself not significant, $F(1,114) = 2.66$, $p = .105$, $\eta_p^2 = .02$, simply rendered the interaction non-significant, $F(2,114) = 2.10$, $p = .128$, $\eta_p^2 = .04$ while preserving the pattern of means (see Table 12 for estimated means). The main effects of *Norm distinctiveness* and *Group membership* were still not significant, $F(2,114) = 2.34$, $p = .101$, $\eta_p^2 = .04$ and $F < 1$, respectively. Effect sizes capturing ingroup versus outgroup differences in *Person evaluation*

Table 12

Estimated means and standard errors of Person *and* Behavior evaluation *after introduction of the* Target group evaluation *as a covariate in Study B-2 (N = 121)*

	Norm distinctiveness											
	non-distinct				distinct				distinct plus			
Target's Group membership	ingroup		outgroup		ingroup		outgroup		ingroup		outgroup	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Person evaluation	4.92	0.17	4.54	0.17	4.24	0.17	4.57	0.19	4.61	0.17	4.82	0.18
Behavior evaluation	4.28	0.29	4.39	0.29	4.36	0.29	3.95	0.33	4.41	0.30	4.66	0.31

Note. Higher values indicate more negative evaluation.

after controlling for the covariate were $d = 0.50$, $d = -0.45$, and $d = -0.28$ for the non-distinct, distinct and distinct plus conditions, respectively.

For *Behavior evaluation* the *Target group evaluation* covariate was marginally significant, $F(1,114) = 3.86$, $p = .052$, $\eta_p^2 = .03$, but there were still no significant main or interaction effects of *Group membership* or *Norm distinctiveness*, both Fs < 1. Effect sizes capturing ingroup versus outgroup differences in *Behavior evaluation* after controlling for the covariate were $d = -0.08$, $d = 0.32$, and $d = -0.34$ for the non-distinct, distinct and distinct plus conditions, respectively.

Discussion

The general interaction pattern for *Person evaluation* in the present study without control for the covariate was reversed compared to the results of Study B-1: while there was a tendency for a BSE in the non-distinct condition, the reverse was true in the distinct

conditions. In face of this blatant contradiction between results with the same materials, it seemed necessary to repeat the study.

There was weak evidence in favor of the hypothesis that individual *Person evaluations* were colored by the group stereotype induced by the *Norm distinctiveness* manipulation. This effect however did not fully eliminate the *Group membership* × *Norm distinctiveness* interaction on *Person evaluation* as expected if the created group stereotypes were responsible for the differences.

11.3 Study B-3

Materials and design of this study were identical to those in Study B-2.

Participants

Participants were 69 volunteers from two introductory lectures in statistics for economics majors and psychology for minors. They were offered the opportunity to participate and obtained a ticket for a raffle of €10 prizes at the end of the lecture. They were later debriefed by email.

There were different percentages of female volunteers in the two lectures (49% and 81%, respectively), but male and female participants were evenly distributed across conditions in both subsamples, $\chi^2(df = 5, N = 33) = 1.78$, $p > .80$, $\varphi = .23$ and $\chi^2(df = 5, N = 33) = 4.58$, $p > .47$, $\varphi = .36$. Participants were on average 22 years old ($SD = 3$ years) and mean age did not differ between subsamples, $F < 1$.

Results

Means and standard deviations of all dependent measures are shown in Table 13.

Person evaluation and *Behavior evaluation* (αs = .86 and .88, respectively) correlated significantly, $r(69) = .57$, $p < .001$.No discernible effects emerged on *Person evaluation*, all

Table 13

*Means (and standard deviations in parentheses) of dependent variables in Study B-3 (*N = *69)*

	Norm distinctiveness					
	non-distinct		distinct		distinct plus	
Target's Group membership	ingroup	outgroup	ingroup	outgroup	ingroup	outgroup
Person evaluation	4.57 (0.63)	4.23 (0.67)	4.57 (0.61)	4.44 (0.98)	4.69 (0.76)	4.80 (0.86)
Behavior evaluation	4.07 (1.71)	4.06 (1.26)	3.96 (1.24)	3.70 (1.89)	5.14 (1.08)	4.22 (1.65)
Ingroup evaluation	3.23 (0.74)	3.30 (0.54)	2.64 (0.73)	2.88 (0.83)	2.43 (0.52)	2.80 (0.60)
Outgroup evaluation	3.70 (1.16)	3.46 (0.73)	3.84 (0.78)	4.34 (0.99)	4.36 (0.62)	4.03 (0.77)

Note. Higher values indicate more negative evaluation.

*F*s < 1.16, all *p*s > .31, all η_p^2s < .04. Effect sizes capturing ingroup versus outgroup differences for the non-distinct, distinct and distinct plus conditions were $d = 0.45$, $d = 0.17$ and $d = -0.15$ respectively.

Behavior evaluation was also not subject to any statistically significant effects, $F(1,63) = 1.14$, $p = .289$, $\eta_p^2 = .02$; $F(2,63) = 1.83$, $p = .168$, $\eta_p^2 = .06$, and $F < 1$, for *Group membership*, *Norm distinctiveness* and the interaction term, respectively. Effect sizes capturing ingroup versus outgroup differences for the non-distinct, distinct and distinct plus conditions were $d = 0.01$, $d = 0.17$ and $d = 0.61$ respectively.

Ingroup and *Outgroup* evaluation (αs = .90 and .94, respectively) were uncorrelated, $r(68) = .09$, $p = .464$. A 2 (target's *Group membership*: ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct vs. distinct vs. distinct plus) × 2 (*Evaluated group*: ingroup vs. outgroup) mixed model ANOVA with repeated measures on the latter factor revealed a main

effect of *Evaluated Group*, $F(1,62) = 86.27$, $p < .001$, $\eta_p^2 = .58$. The ingroup ($M = 2.89$, $SD = 0.72$) was on the whole evaluated less negatively than the outgroup ($M = 3.97$, $SD = 0.90$). This effect was however mainly driven by the distinct and distinct plus conditions, as evidenced by a significant *Norm distinctiveness* × *Evaluated group* interaction, $F(2,62) = 11.65$, $p < .001$, $\eta_p^2 = .27$. While the difference between *Ingroup* and *Outgroup evaluations* was rather large in the distinct and distinct plus conditions, $F(1,62) = 60.48$, $p < .001$, $\eta_p^2 = .49$ and $F(1,62) = 41.47$, $p < .001$, $\eta_p^2 = .40$, respectively, it was smaller and non-significant in the non-distinct condition, $F(1,62) = 2.73$, $p = .104$, $\eta_p^2 = .04$. No other effects were significant, all Fs < 1.40, all ps > .25, all η_p^2s < .05. Thus, there was again an ingroup favoritism effect in the distinct conditions, evidence for participant's sensitivity to the *Norm distinctiveness* manipulation on *Group evaluations*.

Upon introduction of *Target group evaluation* into the analysis of *Person evaluation* as a covariate as in the previous study (this covariate was significant, $F(1,61) = 12.68$, $p < .001$, $\eta_p^2 = .17$), a main effect of the target's *Group membership* emerged, $F(1,61) = 7.46$, $p < .01$, $\eta_p^2 = .11$. The ingroup target was rated more negatively (estimated $M = 4.69$, $SE = .15$) than the outgroup target (estimated $M = 4.25$, $SE = .14$). No other effects were significant, both Fs(2,61) < 2.01, both ps > .14, $\eta_p^2 < .07$. This suggests that the group stereotype induced by the *Norm distinctiveness* manipulation indeed worked against a BSE in this study, but uniformly across *Norm distinctiveness* conditions. Effect sizes capturing ingroup versus outgroup differences in *Person evaluation* after controlling for the covariate were $d = 0.62$, $d = 1.18$, and $d = 0.79$ for the non-distinct, distinct and distinct plus conditions, respectively.

For *Behavior evaluation*, the covariate was also significant, $F(1,61) = 5.40$, $p = .023$, $\eta_p^2 = .08$. Further, the introduction of the covariate rendered the main effect of *Group membership* significant, $F(1,61) = 5.23$, $p = .026$, $\eta_p^2 = .08$. An ingroup target's behavior was then

evaluated more negatively (estimated $M = 5.74$, $SE = 0.31$) than an outgroup target's (estimated $M = 3.68$, $SE = 0.28$). The main effect of *Norm distinctiveness* reached marginal statistical significance after introduction of the covariate, $F(2,61) = 2.52$, $p = .089$, $\eta_p^2 = .08$. Evaluation of the target's behavior was more negative in the distinct plus condition than in the distinct condition, $t(61) = 2.19$, $p = .033$, $d = 0.66$ and also tended to be more negative in the non-distinct condition, although not significantly so, $t(61) = 1.55$, $p = .126$, $d = 0.46$. Evaluations in the distinct and non-distinct conditions did not differ, $t(61) = -0.73$, $p = .470$, $d = -0.21$. The interaction term remained non-significant, $F(2,61) = 1.62$, $p = .206$, $\eta_p^2 = .05$.

Effect sizes capturing ingroup versus outgroup differences in *Behavior evaluation* after controlling for the covariate were $d = 0.10$, $d = 0.83$, and $d = 1.24$ for the non-distinct, distinct and distinct plus conditions, respectively.

Discussion

Thus, while the main and interaction effects without considerations of the covariate were not significant, a group main effect consistent with a BSE prediction emerged after introduction of the covariate. This also points to the possibility that while processes leading to a BSE were operating, their results were eliminated by the positive group image created by the *Norm distinctiveness* manipulation.

Given that the sample size of this study was rather small, yet another replication with the same design and materials was attempted.

11.4 Study B-4

Design and materials employed in this study were identical to those in Studies B-2 and B-3.

Participants

Participants were 218 students of FSU who volunteered to participate in a hallway on campus. They were randomly assigned to one of the six conditions of the 2 (*Group membership)* × 3 (*Norm distinctiveness*) design. The received a chocolate bar for compensation and were later debriefed by email.

Nine participants indicated that they had already filled out the questionnaire during an earlier study, their data was therefore omitted from analyses, leaving a final sample of N = 209 (mean age = 21 years, *SD* = 2 years, 65% female). Sizes of the six cells of the design ranged from to 34 to 37.

Results

Means and standard deviations of the dependent measures and prospective covariates are shown in Table 14.

Person evaluation (α = .84) and *Behavior evaluation* (α = .85) strongly correlated, $r(208) = .55$, $p < .001$. The former was not subject to any main effects, $F < 1$ and $F(2,203) = 1.28$, $p = .289$, $\eta_p^2 = .01$ for *Group membership* and *Norm distinctiveness*, respectively, nor to an interaction effect, $F < 1$. Effect sizes capturing ingroup versus outgroup differences for the non-distinct, distinct and distinct plus conditions were $d = -0.19$, $d = 0.06$ and $d = 0.03$ respectively.

Table 14

*Means (and standard deviations in parentheses) of dependent variables in Study B-4 (*N = *209)*

	Norm distinctiveness					
	non-distinct		distinct		distinct plus	
Target's Group membership	ingroup	outgroup	ingroup	outgroup	ingroup	outgroup
Person evaluation	4.56 (0.89)	4.71 (0.74)	4.76 (0.78)	4.71 (0.74)	4.53 (0.79)	4.51 (0.84)
Behavior evaluation	4.19 (1.38)	4.51 (1.36)	4.25 (1.26)	4.45 (1.34)	4.11 (1.44)	4.21 (1.36)
Ingroup evaluation	3.52 (0.78)	3.44 (0.93)	2.96 (0.66)	2.64 (0.60)	2.87 (0.72)	2.66 (0.64)
Outgroup evaluation	3.54 (0.78)	3.54 (0.94)	4.19 (0.99)	4.12 (0.87)	4.33 (0.73)	4.36 (0.71)

Note. Higher values indicate more negative evaluation.

There were also no effects on *Behavior evaluation* ($\alpha = .85$), $F(1,203) = 1.23$, $p = .269$, $\eta_p^2 = .01$, $F < 1$, $F < 1$, for *Group membership*, *Norm distinctiveness*, and their interaction, respectively. Effect sizes capturing ingroup versus outgroup differences for the non-distinct, distinct and distinct plus conditions were $d = -0.23$, $d = -0.15$ and $d = -0.07$ respectively.

Ingroup and *Outgroup evaluations* were both reliable (αs = .89 and .93, respectively) and correlated significantly, but weakly, $r(208) = .15$, $p < .027$. A 2 (target's *Group membership*: ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct, distinct, distinct plus) × 2 (*Evaluated group*: ingroup vs. outgroup) mixed model ANOVA with repeated measures on the latter factor revealed a main effect of *Evaluated Group*, $F(1,202) = 261.97$, $p < .001$, $\eta_p^2 = .57$. The ingroup ($M = 3.01$, $SD = 0.80$) was on the whole evaluated less negatively than the outgroup ($M = 4.02$, $SD = 0.90$). This main effect did not occur uniformly across levels of *Norm distinctiveness* as evidenced by a significant *Norm distinctiveness* × *Evaluated Group*

interaction, $F(2,202) = 58.43$, $p < .001$, $\eta_p^2 = .37$. While the ingroup was evaluated significantly less negative than the outgroup in the distinct condition, $F(1,202) = 159.37$, $p < .001$, $\eta_p^2 = .44$, as well as in the distinct plus condition, $F(1,202) = 224.34$, $p < .001$, $\eta_p^2 = .53$, there was no intergroup difference in the non-distinct condition, $F(1,202) = 0.33$, $p = .568$, $\eta_p^2 < .01$ (see Table 14 for means and standard deviations). No other effects reached conventional levels of statistical significance, all Fs< 2.28, all ps > .13, all η_p^2s < .02. This pattern closely resembled that of *Ingroup* and *Outgroup evaluations* in studies B-2 and B-3 and attests to participants' sensitivity to the *Norm distinctiveness* manipulation and also to the negativity of the focal behavior of cheating in exams.

The covariate *Target group evaluation* in a model predicting *Person evaluation* scores was significant, $F(1,201) = 23.40$, $p < .001$, $\eta_p^2 = .10$. This addition of the covariate also rendered the main effect of *Group membership* significant, $F(1,201) = 4.44$, $p = .036$, $\eta_p^2 = .02$. As in Study B-3, the ingroup target ($M = 4.77$, $SE = 0.08$) was evaluated more negatively than the outgroup target ($M = 4.51$, $SE = 0.08$). Furthermore, the *Group membership* × *Norm distinctiveness* interaction was significant, $F(2,201) = 3.68$, $p = .027$, $\eta_p^2 = .04$. This reflects the fact that main effect of *Group membership* was driven by the distinct and specific norm plus conditions, $F(1,201) = 6.19$, $p = .014$, $\eta_p^2 = .03$ and $F(1,201) = 4.71$, $p = .031$, $\eta_p^2 = .02$, respectively, while there was no difference in the non-distinct condition, $F < 1$. This result is again in line with the hypothesis that the induced distinctiveness in the two distinct conditions worked against a BSE which however re-emerges once the differential group images caused by the norm manipulation is statistically controlled for. Effect sizes capturing ingroup versus outgroup differences in *Person evaluation* after controlling for the covariate were $d = -0.24$, $d = 0.57$, and $d = 0.68$ for the non-distinct, distinct and distinct plus conditions, respectively.

Introduction of the covariate into the model predicting *Behavior evaluations* did not change the result pattern for the main effects of *Group membership* and *Norm distinctiveness* and their interaction, all Fs < 1.31, all ps > .27, all η_p^2s < .01. The covariate itself however was significant, $F(1,201) = 8.31$, $p < .01$, $\eta_p^2 = .04$. Effect sizes capturing ingroup versus outgroup differences in *Behavior evaluation* after controlling for the covariate were $d = -0.27$, $d = 0.14$, and $d = 0.31$ for the non-distinct, distinct and distinct plus conditions, respectively.

Discussion

Regarding the hypothesis that there would be a BSE in the distinctive and distinctive plus conditions, but a reversal in the non-distinct condition – if the covariate was not considered –, an inconsistent picture has emerged so far. In Study B-1 one pattern was found (ingroup favoritism in the non-distinctive condition, no difference in the distinct condition), in Study B-2 somewhat of an opposite picture (BSE in the non-distinct condition and ingroup favoritism in the two distinct conditions) and studies B-3 and B-4 revealed essentially null effects (with a very small sample size in Study B-3). In the face of this ambiguous picture regarding the Reversed Black Sheep Effect Hypothesis, a last study with cell sizes $n > 20$ was undertaken to test the hypothesis and locate a more meaningful and unequivocal pattern in the data.

11.5 Study B-5

This last study in the series was again identical in materials and design to studies B-2, B-3, and B-4.

Participants

Participants were 140 students of Fachhochschule Jena (School of Applied Sciences) and FSU who volunteered to participate in a hallway on the Fachhochschule campus. The

Table 15

*Means and standard deviations of dependent variables in Study B-5 (*N = *140)*

	Norm distinctiveness											
	non-distinct				distinct				distinct plus			
Target's Group membership	ingroup		outgroup		ingroup		outgroup		ingroup		outgroup	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Person evaluation	4.44	0.73	4.20	0.87	4.43	0.78	4.65	0.69	3.86	0.76	4.68	0.87
Behavior evaluation	3.84	1.21	3.92	1.18	4.31	1.27	4.58	1.49	4.09	1.15	4.17	1.21
Ingroup evaluation	3.41	0.95	3.30	0.66	2.89	0.74	3.06	0.75	2.78	0.60	2.69	0.67
Outgroup evaluation	3.42	0.93	3.38	0.70	3.97	0.87	3.81	0.88	3.64	0.75	4.11	0.85

Note. Higher values indicate more negative evaluation.

received a chocolate bar for compensation and later debriefed by email. Their average age was 24 years (*SD* = 4 years) and 41% were female. Cell sizes were *n* = 23 or *n* = 24.

Results

Means and standard deviations of measured variables are shown in Table 15.

Person and *Behavior evaluations* correlated strongly, $r(140) = .48$, $p < .001$. *Person evaluations* ($\alpha = .82$) did not differ as a function of *Norm distinctiveness*, $F(2,134) = 1.49$, $p = .230$, $\eta_p^2 = .02$. They were however subject to a main effect of the target's *Group membership*, $F(1,134) = 4.01$, $p = .047$, $\eta_p^2 = .03$. Across all *Norm distinctiveness* conditions, the ingroup target ($M = 4.24$, $SD = 0.79$) was evaluated less negatively than the outgroup target ($M = 4.51$, $SD = 0.84$). This *Group membership* main effect was qualified by a *Group membership* × *Norm distinctiveness* interaction effect, $F(2,134) = 5.34$, $p = .006$, $\eta_p^2 = .07$. While in the

Table 16

Results of a 2 (target's Group membership*: ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct, distinct, distinct plus) × 2 (*Evaluated group*: ingroup vs. outgroup) mixed model ANOVA with repeated measures on the last factor in Study B-5*

Source	df	F	p	η_p^2
Group membership	1	0.13	.721	<.01
Norm distinctiveness	2	0.43	.654	.01
Evaluated Group	1	102.21	< .001	.43
Group membership × Norm distinctiveness	2	0.47	.628	.01
Group membership × Evaluated Group	1	0.46	.497	< .01
Norm distinctiveness × Evaluated Group	2	23.41	< .001	.26
Group membership × Norm distinctiveness × Evaluated Group	2	3.40	.036	.06

Note. Error terms of all F-values have *df* = 134.

distinct plus condition, the ingroup target was evaluated less negatively than the outgroup target, $F(1,134) = 12.73$, $p < .001$, $\eta_p^2 = .09$, there was no difference in the non-distinct and distinct conditions, Fs(1,134) < 1.09, ps > .30, η_p^2s < .01 (see Table 15 for means and standard deviations). Effect sizes capturing ingroup versus outgroup differences for the non-distinct, distinct and distinct plus conditions were $d = 0.30$, $d = -0.28$ and $d = -1.04$ respectively.

Behavior evaluation ($\alpha = .79$) was subject to a marginally significant main effect of *Norm distinctiveness*, $F(2,134) = 2.39$, $p = .096$, $\eta_p^2 = .03$. Evaluations were less negative in the non-distinct condition than in the distinct condition, $t(134) = 2.80$, $p = .031$, $d = 0.45$, with evaluations in the specific norm plus condition between these two, but not significantly different from either, |t|s(134) < 1.24, ps > .21, |d|s < .26. The main effect of *Group membership* and the interaction effect were non-significant, both Fs < 1. Effect sizes

capturing ingroup versus outgroup differences for the non-distinct, distinct and distinct plus conditions were d = -0.06, d = -0.21 and d = -0.06 respectively.

Ingroup and *Outgroup evaluation* were reliable scales (α = .87 and α = .91, respectively) and correlated considerably, $r(140) = .32$, $p < .001$. A 2 (target's *Group membership*: ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct, distinct, distinct plus) × 2 (*Evaluated group*: ingroup vs. outgroup) mixed model ANOVA with repeated measures on the latter factor revealed a main effect of *Evaluated Group*, an interaction effect of *Norm distinctiveness* and *Evaluated Group* and a three way interaction (see Table 16 for results of this ANOVA). Only these significant effects will be discussed here.

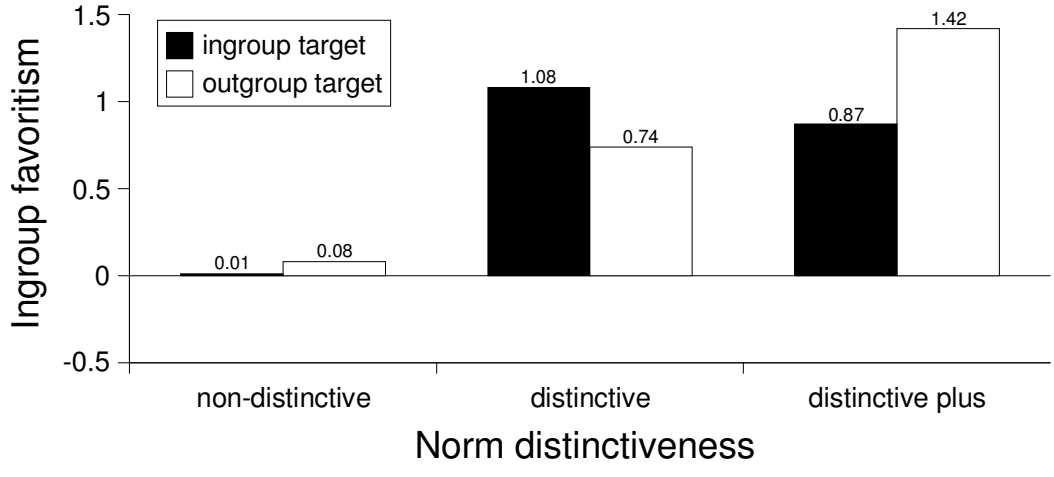
The main effect for *Evaluated group* reflected that the ingroup (M = 3.02, SD = 0.77) was on the whole evaluated less negatively than the outgroup (M = 3.72, SD = 0.86). Concerning the *Norm distinctiveness* × *Evaluated Group* interaction, as in the preceding studies, participants' group evaluations reflected the *Norm distinctiveness* manipulation in that there was a clear ingroup favoritism in the distinct and the distinct plus conditions, $F(1,134) = 91.69$, $p < .001$, $\eta_p^2 = .41$ and $F(1,134) = 56.77$, $p < .001$, $\eta_p^2 = .30$, respectively, but no difference in the non-distinct condition, $F < 1$.

In order to interpret the three-way interaction, the difference *Outgroup – Ingroup evaluations* was calculated with higher values representing higher ingroup favoritism, as higher values of the original variable indicates more negative evaluation. This difference was then treated as a dependent variable in a 2 (target's *Group membership*: ingroup vs. outgroup) × 3 (*Norm distinctiveness*: non-distinct vs. distinct vs. distinct plus) between-participants ANOVA. The means of the group difference variable for *Group Evaluations* are displayed in Figure 2.

This ANOVA revealed that while the degree of ingroup favoritism (i.e., the difference score), was not affected by the target's *Group membership* in the non-distinct condition, $F < 1$, the target's ingroup membership non-significantly increased ingroup favoritism compared to outgroup membership in the distinct condition, $F(1,134) = 1.89$, $p = .171$, $\eta_p^2 = .01$, an ingroup favoritism was attenuated in the distinct plus condition if the target was an ingroup member compared to if he was an outgroup member, $F(1,134) = 5.33$, $p = .023$, $\eta_p^2 = .04$.

Next, the covariate of the respective *Group evaluation* was introduced into the model predicting *Person evaluation* from *Group membership* and *Norm distinctiveness*. The covariate was itself significant, $F(1,133) = 60.28$, $p < .001$, $\eta_p^2 = .31$, and rendered both the *Group membership* main effect and the interaction non-significant, $F(1,133) = 1.35$, $p = .247$, $\eta_p^2 = .01$ and $F(2,133) = 1.01$, $p = .367$, $\eta_p^2 = .02$, respectively. But in this analysis, a marginally significant main effect of *Norm distinctiveness* emerged, $F(2,133) = 2.94$, $p = .057$, $\eta_p^2 = .04$. *Person evaluations* were less negative in the distinct plus condition (estimated $M = 4.25$, $SE = 0.10$) than in the distinct condition (estimated $M = 4.57$, $SE = 0.10$), $t(133) =$

Figure 2
Ingroup favoritism as a function of target's Group membership *and* Norm distinctiveness *in Study B-5*



2.33, $p = .021$, $d = 0.48$, and ratings in the non-distinct conditions were in between (estimated $M = 4.33$, $SD = 0.10$), differing marginally from those in the distinct condition, $t(133) = 1.75$, $p = .083$, $d = 0.36$, but not from those in the distinct plus condition, $t(133) = 0.59$, $p = .554$, $d = 0.12$. Effect sizes capturing ingroup versus outgroup differences in *Person evaluation* after controlling for the covariate were $d = 0.34$, $d = 0.44$, and $d = -0.13$ for the non-distinct, distinct and distinct plus conditions, respectively.

For *Behavior Evaluation,* the covariate was also significant, $F(1,133) = 16.47$, $p < .001$, $\eta_p^2 = .11$, but the results did not change compared to the analysis without the covariate: There was a marginally significant main effect of *Norm distinctiveness*, $F(2,133) = 2.92$, $p = .057$, $\eta_p^2 = .04$ with means following the same pattern as in the analysis without the covariate, but no *Group membership* main effect or interaction, $F(1,133) = 1.24$, $p = .268$, $\eta_p^2 = .01$ and $F < 1$, respectively. Effect sizes capturing ingroup versus outgroup differences in *Behavior evaluation* after controlling for the covariate were $d = -0.07$, $d = 0.18$, and $d = 0.52$ for the non-distinct, distinct and distinct plus conditions, respectively.

Discussion

In sum, a reversed BSE (i.e., ingroup favoritism) on *Person evaluation* was found in the distinct plus condition (where originally a BSE was expected), but not in the two remaining conditions. This effect was eliminated by introduction of the covariate *Target group evaluation*, but not reversed. There is thus no evidence in this individual study that a BSE and an ingroup favoritism effect canceled each other out. It seems instead that there was only an assimilation of the target evaluation to the positive image of the group in the distinct plus condition.

Next, effect sizes from the individual studies will be summarized using meta-analysis.

12 Meta-Analysis on Studies Investigating the Reversed Black Sheep Effect Hypothesis

12.1 Results

All effect sizes capturing ingroup versus outgroup target differences on both *Person* and *Behavior evaluation* were subjected to a meta-analysis using the Weighted Integration Method (Hedges & Olkin, 1985; s. above). An overview of these effect sizes is displayed in Table 17.

Person evaluation. The overall effect size across all levels of the *Norm distinctiveness* factor with weighting for different sample sizes was d_m = -0.15, SE = 0.08, p = .030. Applying correction for imperfect reliability did not change this result, d_m = -0.16, SE = 0.08, p = .022. This indicates that indeed, overall there is ingroup favoritism in the *Person evaluation* of the target: Evaluation scores are less negative for ingroup members than outgroup members.

The meta-analysis was repeated for each of the originally manipulated different *Norm distinctiveness* conditions: non-distinct norm, distinct norm and distinct norm plus. Effect sizes did indeed differ as a function of *Norm distinctiveness*: while there was virtually no effect for the non-distinct conditions, uncorrected: d_m = -0.02, SE = 0.13, p = .445, corrected: d_m = -0.02, SE = 0.13, p = .448, an ingroup favoritism effect started to emerge for the distinct condition, uncorrected: d_m = -0.12, SE = 0.14, p = .193, corrected: d_m = -0.13, SE = 0.14, p = .175, and was significant in the distinct plus conditions, uncorrected: d_m = -0.37, SE = 0.15, p < .01, corrected: d_m = -0.40, SE = 0.15, p < .01.

The contrast describing a linear pattern (coefficients +1 0 -1 for non-distinct, distinct and distinct plus conditions, respectively) was marginally significant, for uncorrected effect sizes: Z = 1.73, p = .083 (two-tailed), for corrected effect sizes: Z = 1.88, p = .060, while the

Table 17

Overview of effect sizes from ingroup versus outgroup comparisons in Person Evaluation *and* Behavior Evaluation *from studies using the cheating in an exam scenario*

Measure	Norm distinctiveness	Study	n_{ig}	n_{og}	d (ig – og)	Reliability
Person evaluation	non-distinct	B-1	22	24	-0.87	0.86
	non-distinct	B-2	20	20	0.54	0.86
	non-distinct	B-3	11	13	0.45	0.86
	non-distinct	B-4	34	35	-0.19	0.84
	non-distinct	B-5	23	24	0.30	0.82
		Mean:	d_m = -0.02, *SE* = .13, *CI:* [-0.28, 0.25]			
	distinct	B-1	22	20	0.14	0.86
	distinct	B-2	21	19	-0.73	0.86
	distinct	B-3	9	14	0.17	0.86
	distinct	B-4	35	34	0.06	0.84
	distinct	B-5	23	23	-0.28	0.82
		Mean:	d_m = -0.12, *SE* = 0.14, *CI:* [-0.39, 0.15]			
	distinct plus	B-2	21	20	-0.53	0.86
	distinct plus	B-3	13	9	-0.15	0.86
	distinct plus	B-4	34	37	0.03	0.84
	distinct plus	B-5	23	24	-1.04	0.82
		Mean:	d_m = -0.37, *SE* = .15, *CI:* [-0.66, -0.07]			
		Overall mean:	d_m = -0.15, *SE* = 0.08, *CI:* [-0.31, 0.01]			
Behavior evaluation	non-distinct	B-1	22	24	-0.34	0.82
	non-distinct	B-2	20	20	-0.04	0.85
	non-distinct	B-3	11	13	0.01	0.88
	non-distinct	B-4	34	35	-0.23	0.85
	non-distinct	B-5	23	24	-0.06	0.79
		Mean:	d_m = -0.16, *SE* = 0.13, *CI:* [-0.42, 0.11]			
	distinct	B-1	22	20	0.16	0.82
	distinct	B-2	21	19	-0.02	0.85
	distinct	B-3	9	14	0.17	0.88
	distinct	B-4	35	34	-0.15	0.85
	distinct	B-5	23	23	-0.21	0.79
		Mean:	d_m = -0.05, *SE* = 0.14, *CI:* [-0.32, 0.22]			
	distinct plus	B-2	21	20	-0.49	0.85
	distinct plus	B-3	13	9	0.61	0.88
	distinct plus	B-4	34	37	-0.07	0.85
	distinct plus	B-5	23	24	-0.06	0.79
		Mean:	d_m = -0.09, *SE* = 0.15, *CI:* [-0.38, 0.21]			
		Overall mean:	d_m = -0.10, *SE* = 0.08, *CI:* [-0.26, 0.06]			

Note. Reliability is Cronbach's α. Means are calculated according to the Weighted Integration Method (Hedges & Olkin, 1985) without correction for imperfect reliability.

residual contrast describing a curvilinear pattern (coefficients -1 2 -1) was not, for uncorrected effect sizes: $Z = 0.44$, $p = .659$, for corrected effect sizes: $Z = 0.47$, p = .636. This is evidence that contrary to the original prediction, no group difference emerges for a situation in which there is no *a priori* difference between groups regarding the norm violated by the target. For a situation with a group norm difference independent of the target to be evaluated, for which originally a BSE was expected, the opposite pattern was found. The moderator contrast analysis suggests that the higher the value of the positive distinctiveness for the ingroup the stronger the ingroup favoritism effect on *Person Evaluation*.

Recall that in the course of the studies, the covariate of *Group evaluations* was introduced to control for a possible assimilation of target judgments to the image of the entire group created by the *Norm distinctiveness* manipulation. Indeed some of the covariance analyses reported for the individual studies suggest that this may be the case: Such a group stereotype was indeed created by manipulating *Norm distinctiveness* and participants seemed to apply this stereotype to the individual target, resulting in assimilation. In order to get an integrated picture of the effects corrected for this influence, the meta-analysis was repeated with effect sizes calculated from estimated means in the covariance analyses for all but the first study with the cheating exam. These effect sizes are shown in Table 18.

The overall mean effect size of *Person evaluation* after controlling for *Target group evaluation* was indeed positive, uncorrected: $d_m = 0.26$, $SE = 0.09$, $p < .01$, corrected: $d_m = 0.28$, $SE = 0.09$, $p < .001$. The effect was smallest in the non-distinct condition, $d_m = 0.18$, $SE = 0.15$, $p = .112$ ($d_m = 0.20$, $SE = 0.15$, $p = .095$ with reliability correction), which is to be expected as no *a priori* difference between the groups was induced in this condition. The effect was significant in the distinct condition, $d_m = 0.36$, $SE = 0.15$, $p = .010$ (corrected: $d_m = 0.39$, $SE = 0.15$, $p < .01$) and marginally significant in the distinct plus condition, $d_m = 0.25$,

SE = 0.15, *p* = .052 (corrected: d_m = 0.26, *SE* = 0.15, *p* = .041). Regardless of whether corrected or uncorrected effect sizes were used, both the linear as well as the quadratic contrasts for the moderation test were not significant, all *Z*s < 0.85, all *p*s = .394. Therefore, one has to regard the means for the different *Norm distinctiveness* conditions as homogeneously consistent with outgroup favoritism or a BSE. Thus, if the positive group image created by the information of different propensities to cheat in the two populations (ingroup and outgroup) was statistically controlled for, an outgroup favoritism effect emerged, a BSE. Interestingly, there was even a tendency for a BSE across the non-distinctive conditions, where no such manipulation leading to differential group images had been made, after controlling for the group image. This could be explained by a default view of the ingroup as more positive view of the ingroup. But in only one study, B-3, was such a difference apparent in the non-distinctive condition, it was not discernible in the remaining 3 studies measuring evaluations of the entire categories (all *p*s > .35, see individual studies in Chapter 11 for details).

Behavior evaluation. According to the Weighted Integration Method (Hedges & Olkin, 1985), the mean effect size for *Behavior evaluation* over all levels of *Norm distinctiveness* was d_m = -0.10, *SE* = 0.08, *p* = .110 (after correction for imperfect reliability, d_m = -0.11, *SE* = 0.08, *p* = .088). Results from separate analyses for levels of *Norm distinctiveness* showed that across these levels, the mean effect size did not vary (see Table 17), consistent with both contrasts testing for moderation being far from significant, linear contrast *Z* = 0.34, *p* = .734, and quadratic contrast *Z* = 0.43, *p* = .670 (*Z* = 0.37, *p* = .714, and *Z* = .46, *p* = .643, respectively, with corrected effect sizes). Overall, there was thus a tendency of an ingroup favoritism effect on *Behavior evaluation*: Participants judged the behavior marginally less

negative if it was shown by an ingroup rather than by an outgroup member, homogeneously across *Norm distinctiveness* conditions.

Repeating these analyses for effect sizes on *Behavior evaluation* after controlling for *Target group evaluation* (see Table 18) yielded an overall positive effect, d_m= 0.14, SE = 0.09, p = .051 (after correction: d_m = 0.15, SE = 0.09, p = .038), indicating that after controlling for the group image induced by the *Norm distinctiveness* manipulation, the behavior was evaluated more negatively if the target was an ingroup rather than an outgroup member. This effect seemed to be mainly driven by the distinct and distinct plus conditions, d_m = 0.26, SE = 0.15, p = .041, and d_m = 0.30, SE = 0.15, p = .024, respectively (for corrected effect sizes: d_m = 0.29, SE = 0.15, p = .030 and d_m =.32, SE = .15, p = .016), the mean effect size for non-distinct conditions was in the opposite direction, but not significant, d_m = -0.13, SE = 0.15, p = .201 for uncorrected and d_m = -0.14, SE = 0.15, p = .18 for corrected effect sizes. A contrast comparing mean effect sizes from distinct and distinct plus conditions on the one hand, and those from non-distinct condition on the other (contrast coefficients: 1 1 -2, respectively), was indeed significant, Z = 2.22, p = .027 for uncorrected and Z = 2.41, p = .016 for corrected effect sizes. The residual contrast comparing the distinct and distinct plus conditions was not significant, Z = 0.17, p = .867 for uncorrected and Z = .19, p = .851 for corrected effect sizes. Thus, *Behavior evaluations* followed the same pattern as Person evaluations: once the group image was controlled for, norm violation in the distinct conditions was considered worse if committed by an ingroup rather than by an outgroup member.

Table 18

Overview of effect sizes from ingroup versus outgroup comparisons in Person Evaluation *and* Behavior Evaluation *from studies using the cheating in an exam scenario after statistically controlling for* Target group evaluation.

Measure	Norm distinctiveness	Study	n_{ig}	n_{og}	d (ig – og)	Reliability
Person evaluation	non-specific	B-2	20	20	0.50	0.86
	non-specific	B-3	11	13	0.62	0.86
	non-specific	B-4	34	35	-0.24	0.84
	non-specific	B-5	23	24	0.34	0.82
		Mean:	d_m = 0.18, *SE* = 0.15, *CI:* [-0.11, 0.48]			
	specific	B-2	21	19	-0.45	0.86
	specific	B-3	9	14	1.18	0.86
	specific	B-4	35	34	0.57	0.84
	specific	B-5	23	23	0.44	0.82
		Mean:	d_m = 0.36, *SE* = 0.15, *CI:* [0.06, 0.66]			
	specific plus	B-2	21	20	-0.28	0.86
	specific plus	B-3	13	9	0.79	0.86
	specific plus	B-4	34	37	0.68	0.84
	specific plus	B-5	23	24	-0.13	0.82
		Mean:	d_m = 0.25, *SE* = 0.15, *CI:* [-0.05, 0.54]			
		Overall mean:	d_m = 0.26, *SE* = 0.09, *CI:* [0.09, 0.43]			
Behavior evaluation	non-specific	B-2	20	20	-0.08	0.85
	non-specific	B-3	11	13	0.10	0.88
	non-specific	B-4	34	35	-0.27	0.85
	non-specific	B-5	23	24	-0.07	0.79
		Mean:	d_m = -0.13, *SE* = 0.15, *CI:* [-0.42, 0.17]			
	specific	B-2	21	19	0.32	0.85
	specific	B-3	9	14	0.83	0.88
	specific	B-4	35	34	0.14	0.85
	specific	B-5	23	23	0.18	0.79
		Mean:	d_m = 0.26, *SE* = 0.15, *CI:* [-0.03, 0.56]			
	specific plus	B-2	21	20	-0.34	0.85
	specific plus	B-3	13	9	1.24	0.88
	specific plus	B-4	34	37	0.31	0.85
	specific plus	B-5	23	24	0.52	0.79
		Mean:	d_m = 0.30, *SE* = 0.15, *CI:* [0.00, 0.60]			
		Overall mean:	d_m = 0.14, *SE* = 0.09, *CI:* [-0.03, 0.31]			

Note. Reliability is Cronbach's α. Means are calculated according to the Weighted Integration Method (Hedges & Olkin, 1985) without correction for imperfect reliability.

12.2 Discussion

It should be noted that *Behavior evaluation* was always measured after *Person evaluation* such that it is possible that reactions to the norm violation were expressed on *Person evaluation* and then dissipated by the time participants made judgments about the target's behavior. In order to assert or refute the proposition that reactions to a norm-violating individual are distinct from reactions to the shown behavior itself, future studies should control for the order of the measures. If the results for *Person* and *Behavior evaluation* then are still similar to the present ones, this would suggest that there is indeed a dissociation of the measures. Presently however, this speculation is a possible target of future research.

Introduction of the covariate *Target group evaluation* led to a reversal of the (insignificant) mean effect with effect sizes from models without the covariate. As in *Person evaluation*, the behavior in question was evaluated more negatively if the target person was an ingroup rather than an ingroup member. This is also evidence for a process leading to a BSE if descriptive norms differ in the sense that the distribution of the behavior in question creates positive distinctiveness of the ingroup. Thus the manipulated positive distinctiveness between the groups to the advantage of the ingroup seems to trigger two different processes: On the one hand it leads to more positive evaluation of the target by assimilation to the group image. On the other hand, and working in the opposite direction, the ingroup target violating precisely that positively distinguishing norm elicits negativity as he threatens this comforting distinctiveness.

The covariate which is essential in the finding that a BSE emerges for *Person* and *Behavior evaluation* after controlling for a positive group image created by the *Norm distinctiveness* manipulation was measured after target and behavior evaluation and was thus

potentially influenced by the target evaluation itself. Future research investigating norm violation in the context of group level differences in behavioral norms should secure that the generation of a group stereotype to be applied to an individual target is examined without such contamination. *Group evaluations*, the covariate, were still influenced by the information about the different propensity to cheat in the two groups (see the *Norm distinctiveness* × *Evaluated group* interactions in studies B-2 through B-5), but the joint effects of *Norm distinctiveness* and the target's *Group membership* seemed to have worn off by the time participants rated the entire groups.

Tighter control of potential order effects in the measures would help to better understand the additive or even interactive effects of negative affect from norm violation as a result of a threat to social identity (potentially stronger from an ingroup violator than from an outgroup violator) on the one hand and a differential group image stemming from a difference in norm between the groups on the other. This group image may act just like a stereotype, coloring judgments of individual targets in the direction of the group image. Note that the SGD (Abrams et al., 2004, 2005) identifies situations in which the groups as wholes differ in the propensity to show the behavior in question as especially prone to produce a BSE. The current result however suggest that this may only be the case with behaviors which are not clearly valenced. If the behavior under scrutiny is valenced, and ingroup and outgroup differ *a priori* in the tendency to show such a behavior, there will be a valence difference between the groups which in turn influences judgments of individual targets and potentially acts against the social identity threat causing a BSE.

13 Studies Investigating the Prior Positive Affect Hypothesis

In this part, several studies are reported which test the hypothesis that positive affect that a target is associated with before anything about a norm violation is known, will influence the punitive reaction upon learning about a norm violation. Specifically, recommendations of punishment for a target that is in principle regarded favorably, or evokes positive affect, should be attenuated compared to a target that does not evoke such positive affect or at least less such affect.

Empirically, as it has become apparent in Chapters 9 and 10, the moderated Intergroup Punishment Difference Hypothesis received support. Only for more severe offenses a bias to the advantage of the ingroup was apparent, no bias emerged for lighter offenses. As the automatic positive affect associated to ingroups versus outgroups – which are relatively content-free within the present context –, may be small and fragile measurement error could relatively easily influence possible effects, especially if affective reactions are relatively mild, such as in the scenarios with lighter offenses. Therefore, the more general hypothesis was tested in this complex using group memberships which did not apply to participants, but were clearly valenced *a priori* as established by a pretest and subsequent control measurements.

In this context, membership in a social category of which participants are not a member (i.e., the self is not involved) is taken as a feature of a target which may or may not evoke positive affect. More specifically, nationalities were taken as social categories that can be associated with more or less positive affect. Several countries and their inhabitants were pretested for positivity in their general evaluation without reference to any particular norm violation event. A prominently positively evaluated nationality (Norway) and one which was evaluated distinctly less positive (United States) were then chosen to describe a crime

committed by a diplomat and participants indicated various reactions to learning of that crime and recommended punitive treatments of the criminal.

In the first studies, a moderately positive nationality (Italy) was also added as well as the participants' ingroup nationality, German. Both were later dropped. Italy was dropped as results regarding that category were quite inconsistent compared to Norway and the US. The latter two were evaluated quite distinctly in the pretest and later auxiliary measures (see studies A-2 through A-4) and therefore provide clearer instances to test a hypothesis regarding affective processes with the directly measured self reports. Germany was not included in the pretest and later also dropped from the main design because it is a natural ingroup category. Such categories generally provide problematic instances to test effects of evaluative bias. They surely carry some amount of positive affect just like other nationalities, but are also probably evaluated more positively because of the mere fact that they comprise participants' self (see Hewstone, Rubin, & Willis, 2002). Therefore, self-presentation and social identity concerns likely are involved in evaluating a target from that category as well as recommending harmful treatment (i.e., punishment) for him. These concerns will be featured in the other two complexes of this report, but were presently simply ignored after the first two studies to obtain clear and unambiguous results for the two categories which were not as problematic in this respect (Norway and the United States).

13.1 Pretest

In a pretest, 49 students of Friedrich-Schiller-Universität Jena (FSU) gave general likability ratings for 12 countries and their inhabitants (Belgium, Bulgaria, France, Great Britain, Italy, Netherlands, Norway, Poland, Russia, Slovakia, Spain and the United States of America) . The countries were always presented in the same, alphabetical order with four

Table 19

Reliabilities, means, and standard deviations of country evaluations in the pretest for studies using the Diplomatic immunity scenario

	Cronbach's α	*M*	*SD*
Belgium	.75	4.98	1.46
Bulgaria	.77	4.51	4.37
France	.86	5.53	2.08
Great Britain	.83	5.41	2.02
Italy	.79	5.36	1.81
Netherlands	.85	6.28	2.12
Norway	.75	6.58	1.94
Poland	.84	3.99	1.74
Russia	.74	3.93	1.48
Slovakia	.82	3.90	1.60
Spain	.82	5.48	1.82
United States of America	.83	4.09	2.10

Note. Higher values indicate higher likability.

identical questions: 'How likable do you find this country in general?', 'How likable do you find the people in this country overall?', 'Would you consider in principle to move to that country permanently?', 'Would you consider giving up your current nationality to obtain the nationality of this country?'. They were all answered on 10 point rating scales anchored in a way such that higher values indicated higher likability of the respective country. The four items were averaged to a general evaluation index for each country as Cronbach's αs for the scale was above .74 for all of the countries. Reliabilities, means and standard deviations for the country evaluations are given in Table 19.

For the main studies, Norway was chosen as a country viewed positively, Italy for a country in the medium range of general likability and the US as one of the more negatively viewed countries. The United States were not the most negatively evaluated country – Poland,

Russia, and Slovakia had descriptively lower scores. But the United States were chosen because it is a wealthy western country with a stable democratic history just as Norway. Also, contact with people from Norway is probably rather rare among participants, just as contact with US Americans is, while contact with Russians, Poles and Slovakians may be more frequent and laden with personal experiences, historical animosities and homogeneous hostile stereotypes among participants of the prospective studies (students of FSU).

According to post-hoc unadjusted LSD-tests, the likability ratings of the chosen countries differed significantly from each other, Norway vs. Italy: $t(48) = 3.72$, $p < .001$, $d = 0.54$; Italy vs. US, $t(48) = 3.76$, $p < .001$, $d = 0.54$; Norway vs. US: $t(48) = 5.55$, $p < .001$, $d = 0.80$.

The main studies followed the same basic design which will be described in detail for Study A-1.

13.2 Study A-1

In Study A-1, participants were presented with an alleged newspaper article about a diplomat to the United Nations who had clearly been involved in criminal behavior. Participants then answered questions about their anger reactions to that case and items tapping into recommended punishment intensity and punishment goals.

This study was conducted after an originally unrelated study in which participants first completed a maze task and filled out a questionnaire containing dependent measures of that unrelated study (see the Procedure section for more details). This originally unrelated study will henceforth be referred to as the *Regulatory Focus and Powerful Groups* study (RFPG).

The combination of the RFPG and the one presently intended to be of interest (and described shortly) was in fact done for reasons of data collection economy. There was thus no intention to combine these two studies before data collection. However, results from the presently reported data collection (as reported below) suggested that the manipulation of

Regulatory focus moderated data patterns from the present study testing the Prior Positive Affect Hypothesis. It was already mentioned in the deduction of this latter hypothesis that a possible moderator of the predicted effect could be whether participants processed information spontaneously versus carefully. Knowing the results of the present data set (i.e., *Regulatory focus* did indeed moderate the effect, see below), it seemed plausible that the expected effects should be especially pronounced in a state of spontaneous processing, as participants would be most susceptible to irrelevant information (which the prior association with positive affect represents in the present context of reactions to a crime). In a state of careful deliberation and vigilance, in contrast, these effects may be eliminated or even reversed (see Chapter 6: Summary and Hypotheses). Promotion focus manipulations have indeed been found to lead to faster and less accurate processing while induced prevention focus resulted in more analytical and slower performance (e.g., Förster, Higgins, & Bianco, 2003; Seibt & Förster, 2004). Thus, in a promotion focus participants may be prone to use irrelevant information such as the affect initially associated with the category of a target, while they are not in prevention focus. In the latter state, they may simply ignore the irrelevant information or, if they are aware of it, even overcorrect (Wegener & Petty, 1997), resulting in an effect opposite to that predicted. The exact results leading to the systematic inclusion and manipulation of Regulatory focus in the design will be discussed below. But it is important to motivate the general idea behind the inclusion of *Regulatory focus* in the present study before the description of the actual study.

Participants and Design

Eighty students of FSU participated in this study (mean age: 21, *SD* = 2 years; 63% female). They were given the opportunity to partake in a study in a public hallway on the campus of the university for a chocolate bar. The current study, employing a one-factorial

design with four levels (diplomat's *Country of origin*: Norway vs. Italy vs. United States vs. Germany) was completed after an in principle unrelated study involving the manipulation of *Regulatory focus* (two levels: promotion focus vs. prevention focus, RFPG) which turned out to be of importance. Participants were however from the beginning randomly assigned to one of the eight cells resulting from complete crossing of the factors of RFPG and those from the initially interesting present study, subject to a constraint of equal cell sizes ($n = 10$).

Procedure

Participants first worked on a task to induce promotion versus prevention focus (*Regulatory focus*) adapted from Friedman and Förster (2001). This task has participants help a mouse to get to through a maze shown on paper in 150 seconds. The mouse is either said and depicted to want to reach a piece of her favorite cheese at the other end (promotion focus) or to escape an eagle circling above her at the start of the maze and to reach a safe hole in the wall at the other end of the maze (prevention focus). The cheese, the eagle and the hole are depicted quite vividly on the sheet.

Participants then completed a questionnaire regarding their views of groups of high versus low power and high versus low status unrelated to the current study. These groups were not proposed, rather participants were instructed to think of one such group themselves.

After this part, participants completed a questionnaire containing the main manipulation and dependent measures of the current study. The manipulation was embedded in a fabricated newspaper article reporting discussions about a criminal diplomat at the UN. Specifically, the article was mainly about the question whether diplomatic immunity should be revoked for the diplomat or not. The diplomat was presented as a being a member of the delegation from Norway, Italy, the US or Germany (*Country of origin*). He was said to be involved with a drug and human trafficking organization (mainly importing into the US) with certainty and

having received considerable amounts of money for his services to the ring. The article reported that there is a heating debate at the UN about whether diplomatic immunity should be revoked for him so that he could be tried in court.

Participants then answered to two items tapping into anger reactions toward the diplomat's being involved in crimes (*Anger*, 'Were you angry upon reading that the delegate was involved in drug and human trafficking' and 'Do you get upset over incidences like this one?', $\alpha = .74$), five items measuring recommended punishment intensity (*Punishment*, 'Should the delegate's immunity be revoked?', 'Should the delegate be excluded from his delegation by his government?', 'If excluded from the delegation, should the delegate loose entitlements to, for example, retirement benefits or health insurance as a state official?', 'Should the delegate be excluded from all public service by his government?', 'Which degree of a punishment do you find appropriate?', $\alpha = .56$), and one item each for the extent to which they viewed punishment to be important for retributive reasons (*Just deserts*, 'To what extent do you find it important that a punishment hurts the delegate like he hurt other people by his behavior?') and for rehabilitation and betterment of the diplomat (*Utilitarian punishment*, 'To what extent do you find it important that punishment caused bethinking and betterment in the delegate?'). The latter items were included as single item measures of *Punishment goals* for exploratory purposes. I wanted to get an idea if the difference in positive affect associated with the targets would possibly moderate the focus on different goals pursued by the punishment recommended. It could be for instance, that associated positive affect shifts the purpose of punishment from just deserts to rather utilitarian goals. If this is the case, then the *a priori* positively evaluated target should evoke less endorsement of the just deserts goal compared to the less positively evaluated one, while a reversal would be expected for utilitarian goals (see also the distinction between just deserts and utilitarian punishment goals

in Chapter 2). This dissociation would not directly speak to the intensity of recommended harmful treatment, but could have an influence on the character of measures taken to punish the perpetrator. This hypothesis was secondary however and the rather primitive measurement using two items was carried along through the following studies as the items were placed at the end of the questionnaire where they presumably could 'not do much harm'.

All items were answered on nine point rating scales with higher values indicating more negative reaction and higher endorsement of the punishment goal, respectively.

Results

Means and standard deviations of the dependent variables are shown in Table 20.

A 2 (*Regulatory focus*: promotion vs. prevention) × 4 (diplomat's *Country of origin*: Norway vs. Italy vs. US vs. Germany) on *Anger* revealed no significant main effects for *Country of Origin* and *Regulatory Focus*, $F(3,72) = 0.80$, $p = .499$, $\eta_p^2 = .03$ and $F(1,72) = 0.86$, $p = .357$, $\eta_p^2 = .01$, respectively, and no interaction effect, $F(3,72) = 1.81$, $p = .152$, $\eta_p^2 = .07$. As in the following studies, only Norway and the US will be compared in the meta-analysis, effect sizes reported here are differences of scores for Norway minus scores for the US divided by the standard deviation from the complete design, for promotion and prevention focus separately. Effect sizes capturing the differences between *Anger* towards the diplomats form Norway and the US were $d = -0.86$ in the promotion focus, and $d = 0.59$ under prevention focus.

Table 20

*Means (and standard deviations in parentheses) of dependent variables in Study A-1 (*N = *80)*

	Diplomat's Country of origin							
Country of origin	Norway		Italy		US		Germany	
Regulatory Focus	PM	PV	PM	PV	PM	PV	PM	PV
Anger	5.20 (1.70)	6.75 (1.78)	5.65 (1.86)	6.30 (1.96)	6.80 (1.81)	5.65 (2.38)	5.10 (1.73)	5.60 (1.61)
Punishment	7.04 (1.07)	7.72 (1.15)	7.66 (1.20)	6.70 (1.06)	8.06 (0.73)	7.08 (0.67)	7.52 (0.71)	7.54 (1.30)
Just Deserts	6.00 (1.63)	6.10 (2.56)	5.80 (2.35)	5.50 (2.51)	6.70 (1.57)	5.60 (1.78)	5.90 (2.18)	7.80 (2.15)
Utilitarian Punishment	6.50 (1.90)	7.90 (1.29)	7.40 (1.43)	7.80 (1.62)	7.00 (1.76)	6.90 (2.08)	7.40 (1.71)	7.10 (2.28)

Note. Higher values indicate more anger, harsher punishment and higher endorsement of the respective punishment goal. 'PM' denotes promotion focus conditions, 'PV' denotes prevention focus conditions.

For *Punishment* recommendations, there were also no main effects, $F(3,72) = 0.61$, $p = .611$, $\eta_p^2 = .03$ and $F(1,72) = 1.88$, $p = .175$, $\eta_p^2 = .03$ for *Country of origin* and *Regulatory Focus*, respectively. But an interaction effect emerged, $F(3,72) = 3.19$, $p = .029$, $\eta_p^2 = .12$. For ease of interpretation, the means are displayed in graphical format in Figure 4.

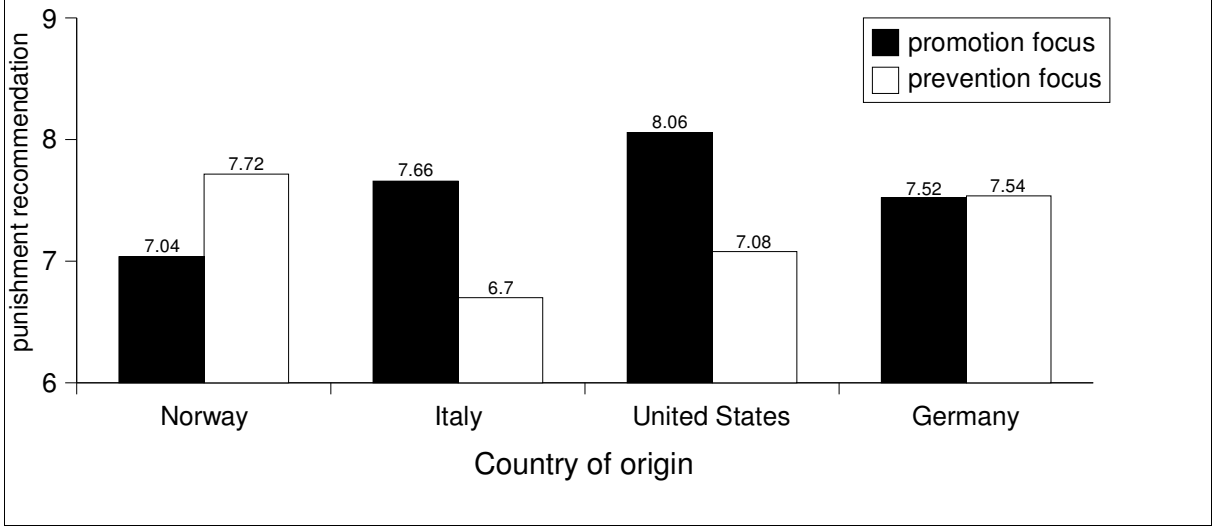
Under promotion focus the means for Norway, Italy, and the US followed the predicted pattern: The more positive the general affective attitude towards the country, the milder the punishment recommendation. Recommendations for the diplomat from the US were significantly higher than for the one from Norway, $t(72) = 2.25$, $p = .027$, $d = 1.01$, judgments for the diplomat from Italy were between those for Norway and the US but not significantly different from either, both ts$(72) < 1.37$, both ps $> .17$, both ds < 0.62. The mean for Germany is not considered here, as no pretest data is available.

Anger and *Punishment* scores correlated significantly, $r(80) = .39$, $p < .001$.

Under prevention focus, the picture was different: The diplomat from Italy received the relatively mildest *Punishment* recommendation, differing significantly from that for the diplomat from Norway, $t(72) = 2.25$, $p = .027$, $d = 1.01$, and marginally from that for the diplomat from Germany, $t(72) = 1.86$, $p = .068$, $d = 0.83$. *Punishment* for the diplomat from the United States tended to be milder than that for the one from Norway, but not significantly, $t(72) = 1.41$, $p = .162$, $d = 0.63$.

Answers to the two items tapping *Punishment goals* were not correlated, $r(80) = -.04$, p = .76. They were therefore analyzed in separate 2 (*Regulatory focus*: promotion vs. prevention) × 4 (diplomat's *Country of origin*: Norway vs. Italy vs. US vs. Germany) ANOVAs. No main or interaction effects were detected, all Fs < 1.79, all ps > .15, all η_p^2s < .07. Effect sizes (Norway – US) were calculated for the two items separately. For the *Just deserts* goal, the effect sizes were $d = -0.33$ under promotion focus, and $d = 0.24$ under prevention focus. For *Utilitarian goals*, effect sizes were $d = -0.28$ and $d = 0.56$, respectively.

Figure 4
Punishment *recommendation as a function of* Country of origin *and* Regulatory focus *in Study A-1 (N = 80)*



Discussion

The results of the present study show that under promotion focus, the likability ratings from the pretest indeed negatively predict *Punishment* recommendation. Under prevention focus, this pattern seems to be even partly reversed. For Norway and the US, the positively and the less positively viewed countries, this cross over pattern is visible for *Anger*, *Punishment*, and the single item *Punishment goals* items, albeit sometimes statistically not clear cut (presumably due to low power). The pattern under promotion focus is consistent with the hypothesis that initial positive affect buffers negative reactions towards an offense. Under prevention focus, processing should be more analytical, more careful and therefore the influence of information unrelated to the actual case in question should be disregarded or even (over-)compensated for (see Wegener & Petty, 1997). This is what happened in this study. However, to be clear, the *Regulatory focus* manipulation was not originally included in the design. Rather, the assignment to *Regulatory focus* condition in the preceding, unrelated study, which was simply combined with the present design for reasons of data collection economy, happened to be known. So clearly, it is a post hoc factor and the results of Study A-1 are in dire need for replication. The aim of A-2 was this replication with a more focused manipulation of processing style. Assuming that the difference between promotion and prevention focus driving the interactions found in Study A-1 were the scrutiny with which participants weighted and integrated information in their judgment or even self-monitored, the next study explicitly and concisely asked participants to either work through the materials casually and answer spontaneously or to very carefully think about their answers.

13.3 Study A-2

Participants

Participants were 161 students of an introductory lecture in social psychology. They were debriefed in a later session of the lecture.

One participant indicated to have participated in Study A-1, her data was excluded from analyses, leaving a final sample of $N = 160$ (mean age = 22 years, $SD = 2$ years; 83% female). Cell sizes ranged from $n = 18$ to $n = 22$. Participants were later debriefed by email.

Design and Procedure

The questionnaire was similar to the one used in Study A-1. However a manipulation of careful versus spontaneous processing was included as well as control questions at the end of the questionnaire in order to ascertain the pretest results.

Manipulation of Processing style. On the cover sheet of the questionnaire, participants were either instructed to 'spontaneously give their personal opinion' answering questions (spontaneous processing) or to 'reflect their answers carefully' (careful processing). Also, participants in the careful condition were instructed to attentively read the materials of the questionnaire while no such mention was made to those in the spontaneous condition. Finally, after the article, as an introduction to the answering of the questions, the spontaneous version read to 'simply answer the questions spontaneously' and that 'there are no right or wrong answers'. In the careful condition, the questions were introduced by the request to 'reflect answers carefully because of the sensitive nature of the issue'.

Group evaluation control questions. After the same questions regarding the newspaper article as in Study A-1, in the present study participants also indicated how likable they found the four countries of the delegation of which the diplomat from the article was a member.

Table 21

Means (and standard deviations in parentheses) of dependent variables in Study A-2

	Diplomat's Country of origin							
Country of origin	Norway		Italy		US		Germany	
Regulatory Focus	SP	CF	SP	CF	SP	CF	SP	CF
Anger	6.69 (1.49)	5.93 (1.85)	5.83 (1.79)	6.67 (1.29)	6.47 (1.02)	5.58 (1.96)	6.02 (2.20)	5.82 (1.82)
Punishment	7.60 (1.38)	7.50 (0.93)	7.42 (1.12)	7.45 (0.77)	7.81 (0.90)	7.52 (1.08)	7.79 (0.98)	7.01 (1.07)
Just Deserts	5.11 (2.72)	5.71 (1.79)	5.89 (1.61)	5.43 (2.20)	7.05 (1.68)	6.05 (1.90)	5.76 (2.14)	4.91 (1.74)
Utilitarian Punishment	7.50 (1.43)	7.14 (1.59)	7.33 (1.57)	7.33 (1.85)	7.63 (1.50)	7.95 (1.31)	7.81 (1.12)	7.73 (1.12)

Note. Higher values indicate more anger, harsher punishment and higher endorsement of the respective punishment goal. 'SP' denotes the spontaneous processing conditions and 'CF' the careful processing conditions.

Each participant answered the question 'How likable do the people from ____ seem to you, very generally speaking?' for the US, Italy, Norway, and Germany (alway in that order) on nine point rating scales anchored by 1 ('very unlikable') to 9 ('very likable'). The answers to these questions were intended to ascertain a similar ranking of likability of the countries in this main study as in the pretest (*Group evaluations*).

Results

Means and standard deviations for all dependent measures are shown in Table 21.

Anger. The anger items again formed a reliable scale ($\alpha = .82$). No reliable main effects emerged for *Anger*, both Fs < 1. There was however a tendency for a *Country of origin* × *Processing style* interaction effect, $F(3,151) = 2.02$, $p = .114$, $\eta_p^2 = .04$. Effect sizes capturing the differences in *Anger* towards the diplomats form Norway and the US were $d = 0.13$ in the *Spontaneous processing* condition, and $d = 0.20$ under the *Careful processing* manipulation.

Punishment recommendations (α = .71) differed marginally as a function of *Processing style*, $F(1,152) = 3.06$, $p = .082$, $\eta_p^2 = .02$. They tended to generally be higher under the spontaneous instruction ($M = 7.66$, $SD = 1.09$) than under the careful instruction ($M = 7.36$, $SD = 0.97$). The main effect of *Country of origin* as well as the interaction were not significant, both Fs(3,152) < 1.23, both ps > .30, both η_p^2s < .03. Effect sizes capturing the differences in *Punishment* towards the diplomats form Norway and the US were $d = -0.20$ in the *Spontaneous processing*, and $d = -0.01$ in the *Careful processing* conditions.

Punishment and *Anger* correlated moderately, but significantly, $r(159) = .22$, p < .001.

Punishment goals. Answers to the two punishment goals items were not correlated, $r(159) = .03$, $p = .734$. ANOVAs analogous to those in Study A-1 were performed on the two *Punishment goal* items. For *Just deserts*, a main effect of *Country of origin* emerged, $F(3,151) = 3.24$, $p = .024$, $\eta_p^2 = .06$. Endorsement of this punishment goal was higher for the diplomat from the US than for the diplomat from the other countries, all ts(151) > 2.02, all ps < .045, all ds > 0.66, while those for the diplomat from the other countries did not differ from each other, all ts(151) < 0.75, all ps > .45, all ds < .25. The main effect of *Processing style* as well as the interaction effect were not significant, both Fs < 1.93, both ps > .16, both η_p^2s < .03. Effect sizes capturing the differences in the *Just deserts* goal towards the diplomats form Norway and the US were $d = -1.00$ in the spontaneous processing, and $d = -0.17$ in the careful processing conditions.

No effects were significant for *Uitilitarian goals*, all Fs < 1.39, all ps > .24, all η_p^2s < .03. Effect sizes capturing the differences towards the diplomats form Norway and the US were $d = -0.14$ in the *Spontaneous processing*, and $d = -0.55$ in the *Careful processing* conditions.

Group evaluations mirrored the pretest results. Figure 3 shows the means and standard deviations of *Group evaluations* as a function of *Country of origin.* In a 4 (diplomat's *Country*

of origin: Norway vs. Italy vs. US vs. Germany) × 4 (*Evaluated group*: Norway vs. Italy vs. US vs. Germany) mixed model ANOVA with repeated measures on the latter factor revealed a main effect of *Country of origin*, $F(3,453) = 59.38$, $p < .001$, $\eta_p^2 = .28$. People from Norway were evaluated most positively ($M = 6.47$, $SD = 1.33$), followed by Italians ($M = 6.10$, $SD = 1.39$), Germans ($M = 5.66$, $SD = 1.31$) and people from the US ($M = 4.97$, $SD = 1.56$). All means were significantly different from each other, all ps < .001. The *Evaluated Group* × *Country of origin* interaction was marginal, $F(9,465) = 1.84$, $p = .059$, $\eta_p^2 = .04$. Visual inspection of the means shows that *Group evaluations* for the Country from which the diplomat originated tended to be depressed compared to conditions in which the diplomat was from a different country than the evaluated one. This suggests that *Group evaluations* were colored by the information about a negatively behaving individual from that group.

All other effects were not significant, all Fs < 1.50, all ps > .22, η_p^2s < .02.

Discussion

The effects regarding the presumed influence of initial positive affect towards people from Norway compared to people from the US were not as clear cut as in Study A-1. This may have been due to the more specific manipulation of *Processing style*. Therefore a second attempt for replication was undertaken, this time again with the regulatory focus manipulation by way of the maze task, but without the questionnaire that was administered in Study A-1 between the maze task and the questionnaire presently pertinent. Also, levels of the *Country of origin* factor were reduced to *Norway* and *United States*, as the hypothesized differences were expected to be most pronounced for these two.

13.4 Study A-3

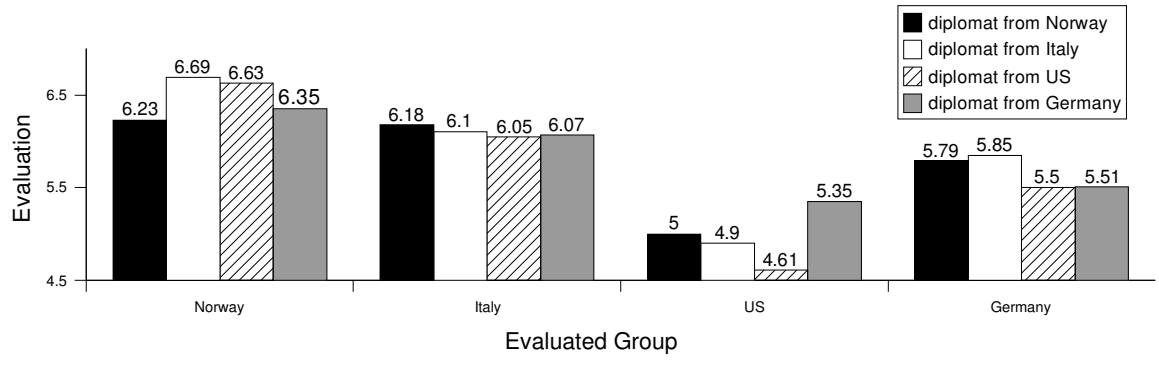
Participants and Design

Participants were 96 students of FSU who participated in a session comprising different, unrelated studies. They received €5 for 30 minutes of their time. They completed one unrelated study before any of the manipulations and materials of the presently discussed study were administered. This preceding study involved a between participants manipulation that was counterbalanced with manipulations in this study and therefore will not be discussed further. Participants were randomly assigned to one cell of a 2 (*Regulatory focus*) × 2 (*Country of Origin*) design.

Two participants indicated on a post session questionnaire that they had participated in studies A-1 or A-2. Their data were omitted from analyses, leaving a final sample of $N = 94$ (mean age = 21.6 years, $SD = 2.6$ years, 59% female). Cell sizes were $n = 23$ or $n = 24$. Participants were later debriefed by email.

Figure 3

Means of Group Evaluations *as a function of* Country of Origin *in Study A-2*



Note. Higher values indicate more positive evaluation.

Procedure

Participants again completed the maze task (Friedman & Förster, 2001, see also Study A-1) and filled out the same questionnaire as in Study A-1, but no other questionnaire was completed between these two parts. Before *Group evaluations* as in Study A-2, two more items regarding expectations were added to the questionnaire to be able to control for the possibility that an effect opposite to the one in the promotion focus condition (i.e., more negative reactions toward the more positively viewed diplomat from Norway) could be explained by differential expectancies regarding the criminal behavior (Olson, Roese, & Zanna, 1996, Bettencourt, Dill, Greathouse, Charlton, & Mullholand, 1997; Kernahan, Bartholow, & Bettencourt, 2000). This hypothesis is referred to henceforth as the *Expectancy Violation Under Prevention Focus Hypothesis* (EPH). According to such an explanation, under prevention focus, the more positively viewed target violates positive expectations and thus elicits negative affect additional to that from the behavior itself. The less positively viewed target, in turn, would be expected by participants to display such negative behavior and thus expectancies would be confirmed which does not lead to negative affect beyond the one from the behavior itself. This explanation must also assume that the hypothesized positive affect buffer for the positively viewed target is smaller than that extra negative affect from expectancy violation, or according to the moderation hypothesis advanced before, that careful processing in prevention focus eliminates this initial positive affect buffer altogether, while expectancy violation still has an effect. This is quite a complex hypothesis translating into a moderated mediation: Under prevention focus, a mediation of the effect of *Country of origin* on the dependent variables is present, while it is not (or considerably weaker) under promotion focus. It will presently be tested if there is the expected difference in favor of the

diplomat from Norway under promotion focus, the reverse pattern under prevention focus and the minimal conditions for a test of mediation are met (see below).

The items used to measure expectancy violation were 'Would you have expected that a diplomat working for <country> could be involved in drug and human trafficking?' and 'Are you surprised that a diplomat working for <country> could be involved in drug and human trafficking?', reverse scored. These items were also answered on nine point rating scales ranging from 1 = 'I would not have expected that at all' and 'No, I am not surprised at all', respectively, to 9 = 'I would have expected that very much' and 'Yes, it does surprise me'. These two items were significantly correlated, $r(94) = .63$, $p < .001$ and therefore averaged for an index of *Expectancy congruency*. *Country of Origin* was manipulated to be either Norway or the US.

Results

Means and standard deviations of dependent measures are shown in Table 22.

Anger ($\alpha = .81$) was not subject to any significant effects, $F(1,90) = 2.53$, $p = .115$, $\eta_p^2 = .03$ for the interaction; for the main effects, both Fs < 1. Effect sizes capturing the differences in *Anger* towards the diplomats form Norway and the US were $d = -0.23$ under promotion focus, and $d = 0.43$ under the prevention focus manipulation.

There were also no effects on *Punishment* ($\alpha = .72$), all Fs < 1. Effect sizes capturing the differences in *Punishment* towards the diplomats form Norway and the US were $d = 0.26$ under promotion focus, and $d = -0.08$ under the prevention focus manipulation.

Anger and *Punishment* were moderately correlated, $r(94) = .21$, $p = .043$.

Expectancy congruency was subject to a main effect of *Country of origin*, $F(1,90) = 21.86$, $p < .001$, $\eta_p^2 = .20$. Participants found the behavior more expectancy congruent if the diplomat was from the US ($M = 5.09$, $SD = 1.98$) rather than from Norway ($M = 3.27$, $SD =$

Table 22

*Means (and standard deviations in parentheses) of dependent variables in study DISS5b (*N = *94)*

	Diplomat's Country of origin			
Country of origin	Norway		US	
Regulatory Focus	PM	PV	PM	PV
Anger	5.83 (2.33)	6.63 (1.87)	6.33 (1.79)	5.72 (2.48)
Punishment	7.56 (1.34)	7.40 (1.21)	7.21 (1.40)	7.50 (1.43)
Expectancy congruency	3.13 (1.14)	3.42 (2.25)	4.83 (2.25)	5.35 (1.67)
Just Deserts	5.50 (1.87)	5.83 (2.58)	5.70 (2.27)	5.43 (2.68)
Utilitarian Punishment	7.50 (1.50)	7.17 (2.60)	7.43 (1.65)	7.17 (2.39)

Note. Higher values indicate more anger, harsher punishment, higher expectancy congruency and higher endorsement of the respective punishment goal. 'PM' denotes promotion focus conditions, 'PV' denotes prevention focus conditions.

1.77). The main effect of *Regulatory focus* and the interaction were not significant, both Fs(1,90) < 1.10, both ps > .29, both η_p^2s < .02.

Answers to the *Punishment goals* items were again not correlated, $r(94) = .12$, $p = .269$ and analyzed separately. No effects emerged for both items, all Fs < 1. Effect sizes capturing the differences in the *Just deserts* goal towards the diplomats form Norway and the US were $d = -0.08$ in the promotion focus conditions, and $d = 0.17$ in the prevention focus conditions. For *Utilitarian goals*, effect sizes were $d = 0.03$ and $d = 0.00$, respectively.

Regarding the EPH, there was an overall main effect of *Country of origin* on *Anger* and *Just deserts,* but it was not moderated. But because the pattern of means were consistent with a cross-over interaction, this hypothesis of moderated mediation will nevertheless be considered. In order for a mediation of the effect of *Country of origin* on *Anger* and *Just deserts* by *Expectancy congruency* under prevention focus to be present, at least the mediator must predict the dependent variables under statistical control for the effect of *Country of origin* (Baron and Kenny, 1986; MacKinnon, Lockwood, Hoffman, West and Sheets, 2002;

Table 23

Correlations (with p *values in parentheses) of* Expectancy congruency *with* Anger, Punishment, Just deserts, *and* Utilitarian Punishment *overall, under promotion and under prevention focus with the association of* Country of origin *with the dependent variables partialed out in Study A-3*

	Anger	Punishment	Just deserts	Utilitarian punishment
Overall	-.12 (.263)	-.22 (.035)	-.04 (.673)	< .01 (.978)
Promotion focus (N = 44)	-.17 (.255)	-.35 (.016)	-.18 (.224)	-.15 (.309)
Prevention focus (N = 44)	-.07 (.636)	-.11 (.460)	.05 (.722)	.09 (.535)

Note. * CO = Country of origin; EC = Expectancy congruency; DV = Dependent variable (*Anger* or *Punishment*).

Shrout and Bolger, 2002). Partial correlations of *Expectancy congruency* and Anger, Punishment, Just deserts and Utilitarian punishment controlled for *Country of origin* are shown in Table 23. From them, it is already evident that the minimal condition just mentioned is not met. If anything, *Expectancy congruency* was negatively related to *Punishment* under promotion focus, but no relationships emerged under prevention focus. Therefore, the test for mediation within the prevention focus condition will not reject the null hypothesis: There is no evidence of mediation by *Expectancy congruency* under prevention focus. Further steps to test the EVP are thus omitted, the hypothesis receives no support.[14]

Group evaluations again showed the pattern consistent with the pretest. There was a main effect of *Evaluated Group*, $F(3,270) = 32.69$, $p < .001$, $\eta_p^2 = .27$. Again, all ratings differed from each other, all ps < .004. Evaluations of Norway were most positive ($M = 6.80$, $SD = 1.22$), followed by Italy ($M = 6.12$, $SD = 1.52$), Germany ($M = 5.47$, $SD = 1.57$) and the US ($M = 4.91$, $SD = 1.84$). There was a theoretically uninteresting marginally significant *Evaluated Group* × *Regulatory focus* interaction, $F(1,90) = 3.34$, $p = .071$, $\eta_p^2 = .04$, but no other main or interaction effects, all Fs < 1.41, all ps > .23, all η_p^2s < .02.

14 This test for a moderated mediation was repeated using a more sophisticated procedure suggested by Muller et al. (2005), which will be described in more detail on p. 160 below. It did not lead to different conclusions.

Discussion

Again, the results of Study A-1 were not replicated. But the study was also carried out in a different setting (laboratory) than before. The first study (A-1) took place in the separated corner of a hallway on campus and participants were passers-by, whereas this current Study A-3 was carried out in a laboratory with participants who had signed up days before their actual participation. Thus, the opportunity for the *Regulatory focus* manipulation to influence the effect small affective biases could have on reactions may have been jeopardized in the lab. It is conceivable that during studies in a regular laboratory, participants are, to begin with, in a generally more careful and vigilant mind set than if they fill out a questionnaire sitting in a cafeteria (e.g., between two cups of coffee, see Study A-1) or during a lecture where it is casually introduced as an illustration example to be discussed later (see Study A-2). Thus, the manipulation by way of the maze task may not have been strong enough to counteract the scrutinous laboratory mind set to an extent that it influenced reactions to norm violation in the way hypothesized here. Therefore the study was carried out once more in the original setting with the same design of the study, but no other study mixed into the materials.

The EPH received no support at all. There was no evidence whatsoever that *Expectancy violation* played a mediating role any more under prevention focus than under promotion focus.

13.5 Study A-4

Participants and Design

Design and materials were identical to those of Study A-3. However, participants were eighty students of FSU passing by a separated corner of a hallway who volunteered to participate in the present study for a chocolate bar as compensation. They underwent no other

Table 24

*Means (and standard deviations in parentheses) of dependent variables in Study A-4 (*N = *52)*

	Diplomat's Country of origin			
Country of origin	Norway		US	
Regulatory Focus	PM	PV	PM	PV
Anger	4.35 (2.40)	4.58 (2.08)	6.32 (1.87)	3.81 (1.85)
Punishment	7.37 (1.43)	6.72 (1.80)	7.79 (1.44)	6.12 (1.10)
Expectancy congruency	3.62 (1.80)	4.25 (1.63)	5.57 (2.25)	6.04 (1.38)
Just Deserts	6.33 (2.06)	6.08 (2.15)	5.43 (2.71)	5.46 (1.94)
Utilitarian Punishment	7.08 (2.25)	6.58 (2.15)	7.29 (1.94)	7.00 (1.78)

Note. Higher values indicate more anger, harsher punishment and higher endorsement of the respective punishment goal. 'PM' denotes promotion focus conditions, 'PV' denotes prevention focus conditions.

manipulations or procedure except for the ones of the present study and were randomly assigned to conditions. Twenty-eight persons indicated that they had participated in one of the previous studies using the diplomat scenario, therefore their data was omitted from analyses, leaving a final sample of $N = 52$ (mean age = 22 years, $SD = 3$ years, 65% female) who were evenly distributed across the cells of the design, $n = 12$ or $n = 13$. Participants were later debriefed by email.

Results

Means and standard deviations of dependent variables are shown in Table 24.

Anger scores ($\alpha = .85$) were subject to a marginal main effect of *Regulatory focus*, $F(1,48) = 3.97$, $p = .052$, $\eta_p^2 = .08$. Participants reported to be more angry in the promotion focus condition ($M = 5.37$, $SD = 2.33$) than in the prevention focus condition ($M = 4.18$, $SD = 1.96$). This effect was however moderated by a significant interaction, $F(1,48) = 5.79$, $p = .020$, $\eta_p^2 = .11$. Consistent with the hypothesis, under promotion focus anger was higher toward the diplomat from the US than the one from Norway, $t(48) = 2.49$, $p = .016$, $d = -0.96$; under prevention focus, the difference tended to be reversed, although not statistically

Table 25

Results from regression analyses and Sobel tests testing the EPH in Study A-4

	Effects from regression analyses					
	CO → EC *		EC → DV *		Sobel test	
Dependent measure (DV)	*β*	*p*	*β*	*p*	*t*	*p*
Anger						
Regulatory focus						
promotion focus (N = 26)	.44	.020	-.24	.242	1.08	.291
prevention focus (N = 26)	.53	.007	-.02	.937	0.08	.469
Punishment						
Regulatory focus						
promotion focus (N = 26)	.44	.020	.10	.665	0.43	.670
prevention focus (N = 26)	.53	.007	-.38	.114	1.44	.162

Note. * CO = Country of origin; EC = Expectancy congruency.

significant, $t(48) = 0.94$, $p = .351$, $d = 0.38$. There was no main effect of *Country of origin* by itself, $F(1,48) = 1.10$, $p = .299$, $\eta_p^2 = .02$.

Punishment recommendations ($\alpha = .84$) were only influenced by *Regulatory focus*, $F(1,48) = 8.20$, $p < .01$, $\eta_p^2 = .15$. Under promotion focus ($M = 7.59$, $SD = 1.42$) they tended to be higher than under prevention focus ($M = 6.41$, $SD = 1.48$). The main effect of *Country of origin* as well as the interaction were not significant, Fs(1,48) < 1.57, ps > .21, η_p^2s < .04.

Effect sizes capturing the differences in *Punishment* towards the diplomats form Norway and the US were $d = -0.29$ under promotion focus, and $d = 0.41$ under the prevention focus manipulation.

Anger and *Punishment* were significantly correlated, $r(52) = .28$, $p = .046$.

The two items measuring *Expectancy congruency* were again highly correlated, $r(52) = .53$, $p < .001$, and therefore averaged and analyzed as a combined index. Again, there was only a main effect of *Country of origin*, $F(1,48) = 13.89$, $p < .001$, $\eta_p^2 = .22$, the main effect of

Regulatory focus as well as the interaction were not significant, both Fs(1,48) < 1.21, both ps > .27, both η_p^2s < .03. Seperate mediation analyses for promotion focus and prevention focus conditions were performed for the dependent variables *Anger* and *Punishment*, as descriptively, for these measures, the pattern was the characteristic cross-over interaction found in Study A-1 (see means in Table 24). Results of these mediation analyses are shown in Table 25. For *Anger*, if anything, there is a descriptive tendency for the mediation by *Expectancy congruency* under promotion focus, but not under prevention focus. For *Punishment* however, the descriptive pattern is consistent with the EPH: While there is no mediation by *Expectancy congruency* under promotion focus, there is a tendency for that mediation under prevention focus, but also – possibly due to low power – far from statistical significance.[15]

Answers to the *Punishment goals* items were correlated, but not significantly, $r(51) = .19$, $p = .18$. No significant effect emerged on either, Fs(1,48) < 1.46, ps > .23, η_p^2s < .04. Effect sizes capturing the differences in the *Just deserts* goal towards the diplomats form Norway and the US were $d = 0.39$ in the promotion focus conditions, and $d = 0.28$ in the prevention focus conditions. For *Utilitarian goals*, effect sizes were $d =$- 0.10 and $d =$ -0.20, respectively.

Discussion

No significant main effects of *Country of origin* emerged nor any two-way interactions of this factor and *Regulatory focus.* This may be due to low power. Originally $n = 20$ participants per cell were intended; but the fact that data from a solid 35% of participants had to be excluded from analyses because they had filled out the questionnaire in one of the earlier studies was unfortunate. But this problem is of relatively lesser concern here, as individual

15 The analysis was also carried out using the method proposed by Muller et al. (2005), see footnote 14. Results did not differ from the current analysis.

effect sizes are incorporated into the meta-analysis where their level of significance is less relevant.

There was weak evidence for the EPH on *Punishment* in that indeed under prevention focus, the less positively viewed US diplomat's transgression was more expectancy congruent than that of the Norwegian diplomat and this difference showed a slight tendency to mediate the difference in punishment to the advantage of the less positively viewed US diplomat. However, considering that sample size was unexpectedly low in this study and that this tendential result is at odds with the complete lack of evidence for the EPH in the preceding study, a possible expectancy violation mechanism operating only under prevention focus and responsible for the reversal of the effect under promotion focus should be taken up by future research. Presently, the main focus is on the negative difference between negative reactions to the diplomat from Norway and that from the US emerging under promotion focus, but not under prevention focus. Why the difference would even be *reversed* under prevention focus instead of simply attenuated will not and cannot be examined further here[16], but I will come back to this in the Discussion of the meta-analysis over studies A-1 through A-4 (Chapter 14) below.

16 The items measuring *Expectancy violation* were only contained in the questionnaires of studies A-3 and A-4 and yielded quite inconsistent, even contrary results. Therefore an examination in the meta-analysis to follow is refrained from.

14 Meta-Analysis on Studies Investigating the Prior Positive Affect Hypothesis

14.1 Results

All effect sizes from studies using the diplomat scenario are shown in Table 26.

Reported *Anger* was over all studies and *Regulatory focus* conditions not different for the diplomat from Norway and the US, d_m = -0.01, SE = 0.12, p = .467 for effect sizes not corrected for imperfect reliability, and d_m = -0.01, SE = 0.12, p = .465 for corrected effect sizes. The analysis was repeated for the prevention and promotion focus conditions separately. It turned out that under promotion focus, there was indeed a difference in *Anger* to the disadvantage of the diplomat from the US, d_m = -0.34, SE = 0.18, p = .027, while this difference was reversed under prevention focus conditions, d_m = 0.36, SE = 0.18, p = .020. With correction for imperfect reliability, these results did not change much, d_m = -0.38, SE = 0.18, p = .018 and d_m = 0.40, SE = 0.18, p = .011, respectively. The test for moderation suggest by DeCoster (2004) confirmed that the mean effect sizes for promotion versus prevention focus were indeed different from each other, Z = 2.38, p = .005 for uncorrected, and Z = 3.13, p = .002 for corrected effect sizes.

Punishment recommendations were overall also virtually identical for the two diplomats independently of whether correction for imperfect reliability was applied or not, both $|d_m|$s < 0.02, both SEs = 0.12, both ps > .43. Separate analyses for effect sizes from prevention versus promotion focus conditions revealed that in the former, *Punishment* tended to be milder towards the diplomat from the US, d_m = 0.13, SE = 0.17, p = .228 for uncorrected and d_m = 0.16, SE = 0.17, p = .185 for corrected effect sizes. In the promotion focus conditions however, the opposite tended to be the case, d_m = -0.15, SE = 0.17, p = .192 for uncorrected

Table 26

Overview of effect sizes (differences between the diplomats from Norway and the US) in Anger, Punishment, Just Deserts, *and* Utilitarian goals *from studies using the diplomat scenario (A-1, A-2, A-3, A-4)*

Measure	Regulatory Focus or Processing style	Study	n_{ig}	n_{og}	*d* (ig – og)	α		
Anger	Prevention focus	A-1	10	10	0.59	.74		
	Careful Processing	A-2	21	19	0.20	.82	d_m	0.36
	Prevention focus	A-3	24	23	0.43	.81	*SE*	0.18
	Prevention focus	A-4	12	13	0.38	.85	*CI*	[0.02, 0.71]
	Promotion Focus	A-1	10	10	-0.86	.74		
	Spontaneous Processing	A-2	18	19	0.13	.82	d_m	-0.34
	Promotion Focus	A-3	24	23	-0.23	.81	*SE*	0.18
	Promotion Focus	A-4	13	14	-0.96	.85	*CI*	[-0.69, 0.01]
		Overall Mean:			d_m = -0.01, *SE* = 0.12, *CI:*			[-0.25, 0.23]
Punishment	Prevention focus	A-1	10	10	0.63	.56		
	Careful Processing	A-2	21	20	-0.01	.71	d_m	0.13
	Prevention focus	A-3	24	23	-0.08	.72	*SE*	0.17
	Prevention focus	A-4	12	13	0.41	.84	*CI*	[-0.21, 0.47]
	Promotion Focus	A-1	10	10	-1.01	.56		
	Spontaneous Processing	A-2	18	20	-0.20	.71	d_m	-0.15
	Promotion Focus	A-3	24	23	0.26	.72	*SE*	0.17
	Promotion Focus	A-4	13	14	-0.29	.84	*CI*	[-0.50, 0.19]
		Overall Mean:			d_m = 0.02, *SE* = 0.12, *CI:*			[-0.23, 0.26]
Just Deserts	Prevention focus	A-1	10	10	0.24	n.a.		
	Careful Processing	A-2	21	19	-0.17	n.a.	d_m	0.18
	Prevention focus	A-3	24	23	0.17	n.a.	*SE*	0.24
	Prevention focus	A-4	12	13	0.28	n.a.	*CI*	[-0.30, 0.66]
	Promotion Focus	A-1	10	10	-0.33	n.a.		
	Spontaneous Processing	A-2	18	19	-1.00	n.a.	d_m	-0.15
	Promotion Focus	A-3	24	23	-0.08	n.a.	*SE*	0.24
	Promotion Focus	A-4	13	14	0.39	n.a.	*CI*	[-0.63, 0.33]
		Overall Mean:			d_m = 0.02, *SE* = 0.17, *CI:*			[-0.32, 0.36]
Utilitarian Punishment	Prevention focus	A-1	10	10	0.56	n.a.		
	Careful Processing	A-2	21	19	-0.55	n.a.	d_m	0.15
	Prevention focus	A-3	24	23	0.00	n.a.	*SE*	0.24
	Prevention focus	A-4	12	13	-0.20	n.a.	*CI*	[-0.33, 0.63]
	Promotion Focus	A-1	10	10	-0.28	n.a.		
	Spontaneous Processing	A-2	18	20	-0.14	n.a.	d_m	-0.06
	Promotion Focus	A-3	24	23	0.03	n.a.	*SE*	0.24
	Promotion Focus	A-4	13	14	-0.10	n.a.	*CI*	[-0.54, 0.42]
		Overall Mean:			d_m = 0.05, *SE* = 0.17, *CI:*			[-0.29, 0.39]

Note. α is Cronbach's α reliability coefficient. Means are calculated according to the Weighted Integration Method (Hedges & Olkin, 1985) without correction for imperfect reliability.

Table 27

Mean effect sizes, standard errors and p-values from studies A-1 through A-4

	Processing Style								
	Prevention focus/ careful			Promotion focus/ spontaneous			Overall		
Measure	d_m	*SE*	*p*	d_m	*SE*	*p*	d_m	*SE*	*p*
Anger	0.34	0.17	.09	-0.34	0.18	.02	-0.01	0.12	.47
Punishment	0.13	0.17	.23	-0.15	0.17	.19	0.02	0.12	.45
Just Deserts	0.18	0.24	.23	-0.15	0.24	.27	0.02	0.17	.46
Utilitarian P.	0.15	0.24	.27	-0.06	0.24	.41	0.05	0.17	.40

Note. Means effect sizes are from differences (Norway – US) integrated according to the Weighted Integration Method without correction for imperfect reliability.

and d_m = -0.18, *SE* = 0.17, *p* = .150 for uncorrected effect sizes. These separate mean effect sizes individually are not significantly different from zero and their confidence intervals overlap: The moderation test was not significant, *Z* = 1.15, *p* = .249 and *Z* = 1.38, *p* = .167 for uncorrected and corrected effects sizes, respectively. But the tendency of the results is consistent with the hypothesis and also the result for *Anger*.

Partly similar pictures as for *Anger* and *Punishment* emerged for both punishment goals. As Table 27 summarizes, overall mean effect sizes for them were essentially equal to zero. For the *Just deserts* punishment goal, separate mean effect sizes for promotion and prevention focus showed again a pattern partly consistent with the results for *Anger* and *Punishment.* For *Utilitarian goals* the effect size was positive in the prevention focus, but only slightly negative in the promotion focus condition. The diplomat from the US tended to evoke less intense reactions than the one from Norway under prevention focus and the opposite tended to be true, at least for *Just Deserts* goals, under promotion focus. Confidence intervals of the two mean effect sizes however overlapped considerably for both variables (see Table 26).

14.2 Discussion

In sum, it is clear that reactions on the *Anger* measure, which presumably taps most strongly into initial negative affective reactions to the behavior described, show the hypothesized basic effect of derogation of target who is, by category membership, less likable. Under prevention focus, this difference is reversed. This latter finding may be preliminarily explained by overcorrection (Wegener & Petty, 1997), or, as formulated in the EPH, by *Expectancy congruency* being different for the different targets and thus responsible for the reversal under prevention focus. However, the first alternative cannot be tested here and the second one only received very weak support.

Thus, the current set of studies is indeed consistent with the idea that negative affective reactions towards a perpetrator are influenced by an initial difference in unspecific positive affect toward the perpetrator's social category. This is however only the case if the expression of such a reaction is spontaneous and rather uncontrolled such as under promotion focus. If made in a careful and scrutinizing mind set, such as prevention focus, the opposite difference emerges. Then, an initial advantage in positive affect associated with the category even seemed to work against the perpetrator in the reported studies. Why this is the case must remain a matter of conjecture here. One possibility is that, as tentatively suggested above, participants over-compensate in a careful mind set; they may have a notion of a potential bias they could show and then over-correct for this bias, resulting in a bias against the member of an otherwise rather positively viewed social category. However, if answers to items asking for anger reactions are given spontaneously and without much deliberation, these reactions seem to be buffered by an initial positive affect attached to the category of the perpetrator.

Results on other measures are less clear. One reason may be the inherent flaw of scenario studies asking self-reports in questionnaires, namely that the answering of items earlier in the

package can influence answers to later items and contribute to a steady attenuation of initial effects. This may, in the present case, have led to dissipation of initial effects and thus ambiguity regarding the interpretation of the tendencies on variables after *Anger* which, in future research, could be resolved by systematically varying the order of the items.

An alternative to this dissipation hypothesis can be tested here, namely that there is the same effect on *Punishment* as on *Anger*, but it is mediated by *Anger* and therefore weaker in its appearance on the measures taken after *Anger* measures. This issue will be examined in the following chapter.

15 The Causal Path to Punishment via Anger

Recall that while the results of the meta-analysis showed a clear *Country of origin* × *Regulatory focus* interaction on *Anger* but not on *Punishment* although the pattern of means for *Punishment* was – on a descriptive level – similar to that of *Anger*. Also, *Anger* and *Punishment* correlated significantly across the studies reported in the last complex (*r*s = .39, .22, .21, .28, see the report of the individual studies for details). Thus, while the theory of punishment proposed in Chapter 2 suggests that punitive tendencies are mediated by negative affect (here: anger), no direct effect of *Country of origin* in combination with *Regulatory focus* was found. This could be tentatively explained post-hoc as an outcome of the initial negative affective reactions to the deed dissipating rather rapidly. Alternatively however, it is possible that an effect of the manipulated variables on *Punishment* is rather small because the mediating process tapped into here allows for the intrusion of noise, that is additional variance blurring the effect. This is all the more plausible as in a usual everyday case, perceivers do not answer items regarding their affective reactions before determining appropriate punishment measures. The event of being asked and answering about an experience normally flowing into judgments in a rather unconstrained fashion may introduce undue variability which attenuates the relationship between the crime and the punishment judgment, while the experience is relatively unaffected in the probing of *Anger*.

The conjecture here is that responses to the items measuring *Punishment* are an extension of those to the *Anger* items. To strengthen this argument, one would have to present evidence that indeed *Punishment* is interactively influenced by *Country of origin* and *Regulatory focus* via *Anger*. Such an indirect effect of an independent variable on a dependent variable by way of a third variable is commonly known as mediation (Baron & Kenny, 1986). The original theory of punishment advanced it in Chapter 2 and empirical evidence supporting it lead to

the prediction of precisely such a mediation. The specific mediation hypothesis here is one of mediated moderation, as the total path which is proposed to be mediated is the moderation of the influence of *Country of origin* by *Regulatory focus* (i.e., the interaction of these two independent variables).

The test of this mediated moderation hypothesis will be achieved by a mediation analysis over all cases from studies A-1 through A-4, testing whether *Anger* mediates the non-significant *Country of origin* × *Regulatory focus* interaction effect on *Punishment reactions*. As the materials were all identical up to the measurement of *Punishment* recommendations across the studies, this pooling of all individual cases is deemed conceptually permissible. It remains to empirically legitimize the pooling of individual cases across studies achieved next.

15.1 MANOVA Over Pooled Cases From Studies A-1 Through A-4

Only cases in which the diplomat was from Norway or the US were included (*N*s = 40, 78, 94, 52 for studies A-1 through A-4, respectively). Occasionally, degrees of freedom are inconsistent with *N*s from individual studies because of missing data.

Over all the N_{total} = 264 cases, the *Anger* and *Punishment* indices constructed as in the individual studies had Cronbach's α = .84 and .75, respectively. A 2 (*Country of origin*: Norway vs. US) ×2 (*Regulatory focus*: promotion/spontaneous processing vs. prevention/careful processing) × 4 (*Dataset*: A-1 vs. A-2 vs. A-3 vs. A-4) MANOVA on *Anger* and *Punishment* scores was conducted. Results are tabulated in Table 28.

Consistent with the meta-analysis above, Country of origin had virtually no main effect on the dependent measures. However, *Regulatory focus* exerted a main effect: Both *Anger* ($M_{\text{promotion focus}}$ = 6.00, *SE* = 0.18 and $M_{\text{prevention focus}}$ = 5.58, *SD* = 0.18) as well as *Punishment*

Table 28

Results of a 2 (*Country of origin*: Norway vs. US) × 2 (*Regulatory focus*: promotion vs. prevention) × 4 (*Dataset*: A-1 vs. A-2 vs. A-3 vs. A-4) *MANOVA on* Anger *and* Punishment *over pooled cases from studies A-1 through A-4.*

Source	*df*	Error *df*	*F*	*p*	η_p^2
Country of origin	2	246	0.61	.941	< .01
Regulatory focus	2	246	2.98	.052	.02
Dataset	6	494	3.77	.001	.04
Country of origin × Regulatory focus	2	246	6.33	.002	.05
Country of origin × Dataset	6	494	0.47	.827	.01
Regulatory focus × Dataset	6	494	1.99	.066	.02
Country of origin × Regulatory focus × Dataset	6	494	1.63	.137	.02

($M_{\text{promotion focus}}$ = 7.55, *SE* = 0.11 and $M_{\text{prevention focus}}$ = 7.20, *SE* = 0.11) were higher under promotion focus than prevention focus (all means presently reported are estimated means).

There was also a main effect of *Dataset*, as reactions were generally lower in Study A-4 than in Studies A-1 through A-3 (all simple comparison *p*s < .002), which in turn did not differ from each other (all simple comparison *p*s > .85). This effect is a theoretically uninteresting difference in samples or population from which the samples are drawn and therefore will not receive further attention.

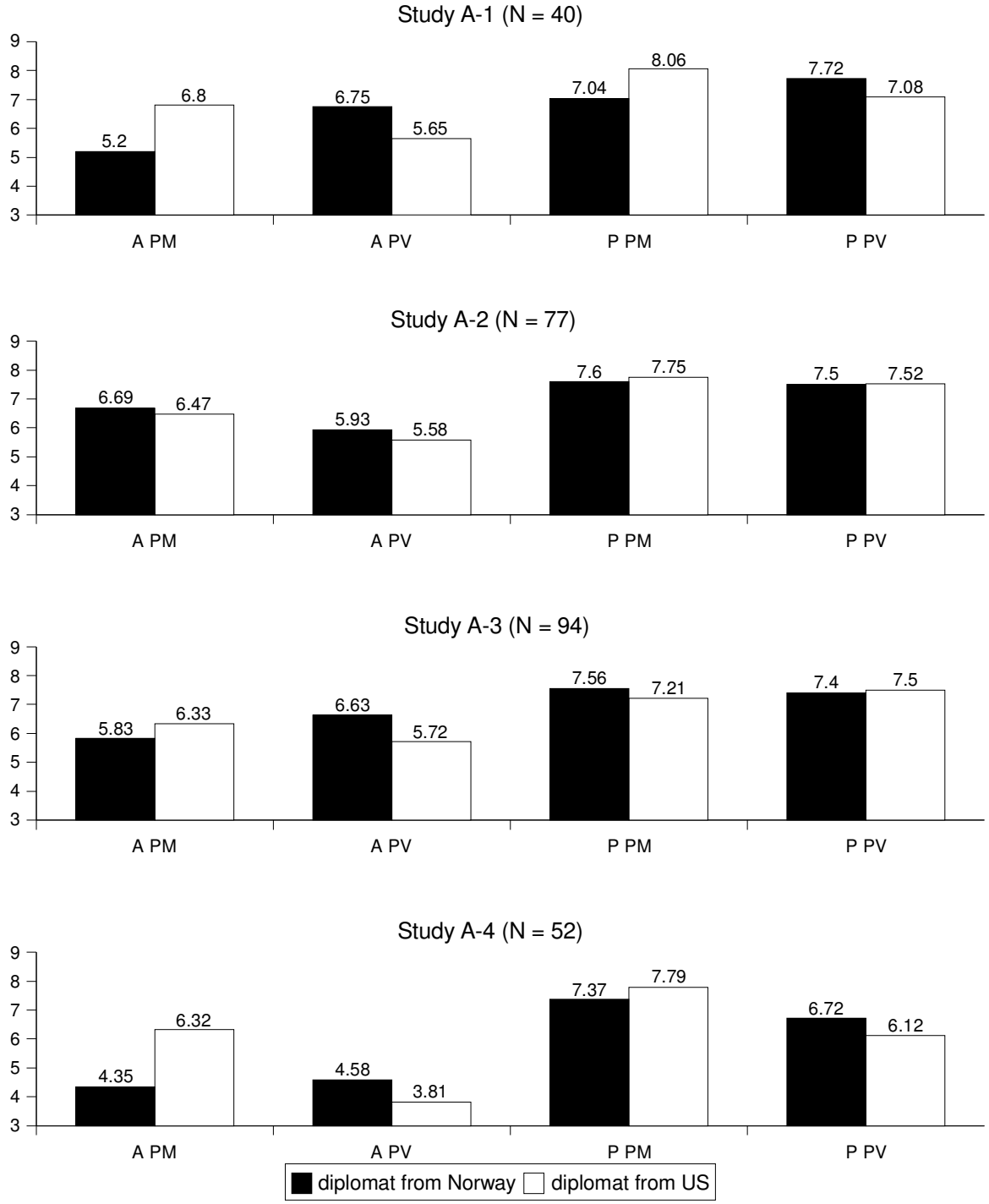
Of equally lesser theoretical importance here, there was also a marginal *Regulatory focus* × *Dataset* interaction, indicating that the main effect of *Regulatory focus* was mainly driven by the data in Study A-4, multivariate $F(2,246) = 6.45$, $p = .002$, $\eta_p^2 = .05$, and somewhat that of Study A-2, multivariate $F(2,246) = 1.73$, $p = .179$, $\eta_p^2 = .01$, while in the first and third studies, there was no main effect at all, both *F*s < 1.

Of more relevance, and in line with the meta-analysis and the (descriptive) patterns obtained in studies A-1, A-3, and A-4, the *Country of origin* × *Regulatory focus* interaction also emerged in the MANOVA. Reactions were stronger toward the US diplomat (M_{anger} = 6.48, SE = 0.25 and $M_{punishment}$ = 7.70, SE = 0.16) than toward the one from Norway (M_{anger} = 5.52, SE = 0.26 and $M_{punishment}$ = 7.39, SE = 0.16) under promotion focus, multivariate $F(2,246) = 3.81$, $p = .023$, $\eta_p^2 = .03$. Under prevention focus the reverse pattern tended to hold, $M_{US,anger} = 5.19$, $SE = 0.26$, $M_{US,punishment} = 7.06$, $SE = 0.16$, $M_{\text{Norway,anger}} = 5.97$, $SE = 0.26$, and $M_{Norway,punishment} = 7.34$, $SE = 0.16$, multivariate $F(2,246) = 2.59$, $p = .077$, $\eta_p^2 = .02$.

Most important for the present purpose, no significant three-way interaction of *Country of origin*, *Regulatory focus*, and *Dataset* emerged. While for *Anger* there was no *Country of origin* × *Regulatory focus* interaction whatsoever in Study A-2 (but the remaining studies showed this pattern at least descriptively) and for *Punishment* the characteristic interaction pattern (while possibly not significant) only showed in studies A-1 and A-4 (see Figure 5), overall, these deviations from the overall pattern were not as pronounced as to seriously jeopardize the legitimacy of pooling cases from studies A-1 to A-4. If anything, the inclusion of the studies worked against the hypothesis of a mediated moderation of the relationship between the *Country of origin* × *Regulatory focus* interaction and *Punishment* via *Anger*.

Figure 5

Means of Anger *and* Punishment *as a function of* Country of Origin, Regulatory Focus, and Dataset in studies A-1 through A-4



Note. Higher values indicate more anger and more punishment. 'A' and 'P' denote the measures *Anger* and *Punishment,* 'PM' stands for promotion focus/spontaneous processing, 'PV' for prevention focus/careful processing, and A-1 through A-4 for the individual studies with the same name.

15.2 Testing the Mediated Moderation Hypothesis

According to Muller et al. (2005), to establish mediated moderation, three models have to estimated:

$$Y = \beta_{70} + \beta_{71} X + \beta_{72} Mo + \beta_{73} XMo + \varepsilon_7 \quad (7)$$

$$Me = \beta_{80} + \beta_{81} X + \beta_{82} Mo + \beta_{83} XMo + \varepsilon_8 \quad (8)$$

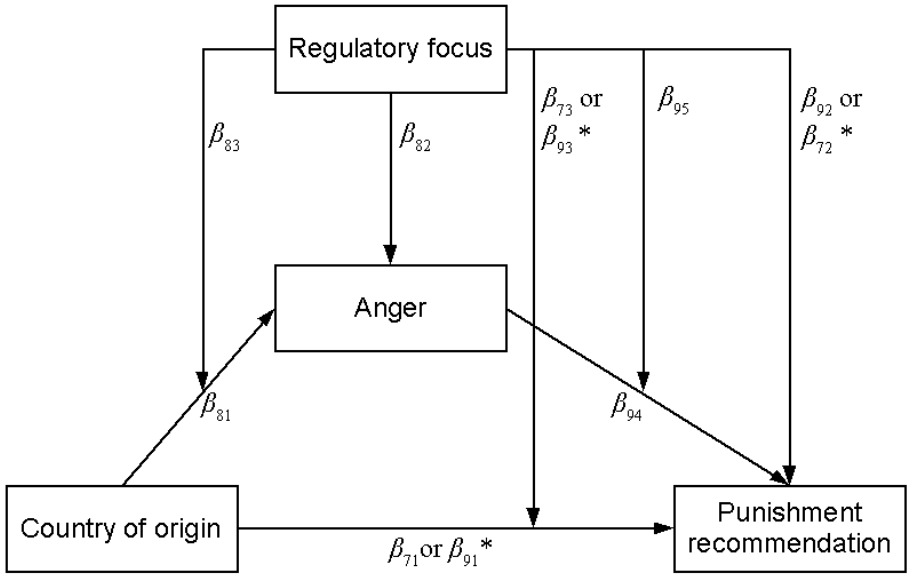
$$Y = \beta_{90} + \beta_{91} X + \beta_{92} Mo + \beta_{93} XMo + \beta_{94} Me + \beta_{95} MeMo + \varepsilon_9 \quad (9)$$

where *X* is the independent variable (here: *Country of origin*) *Y* is the dependent variable (here:*Punishment*), *Mo* the proposed moderator (here: *Regulatory focus*), and *Me* the proposed mediator (here: *Anger*). The models are visualized in Figure 6.

Muller et al. (2005) suggest that mediated moderation is present if b_{43} is significant and one or both of the following conditions hold:

Figure 6

Model of mediated moderation of the interactive effect of Country of Origin *and* Regulatory Focus *on* Punishment *via* Anger



Note. Paths parameters marked with an asterisk depend on the model (7) through (9) being estimated.

Table 29

Estimates of models (7) through (9) suggested by Muller et al. (2005) over pooled cases from studies A-1 through A-4.

Model	β_{k1} (p)	β_{k2} (p)	β_{k3} (p)	β_{k4} (p)	β_{k5} (p)
(7) $N = 263$	.00 (.978)	-.11 (.072)	-.07 (.230)		
(8) $N = 262$	.00 (.980)	-.09 (.124)	-.19 (.002)		
(9) $N = 262$	.01 (.854)	-.08 (.179)	-.02 (.798)	.29 (< .001)	.09 (.149)

Note. k in expressions β_{kn} stands for the respective model (7) through (9).

- β_{83} and β_{94} are significant, or
- β_{81} and β_{95} are significant.

It is likely that β_{73} is not significant, as the meta-analysis found an only marginal *Country of origin* × *Regulatory focus* interaction. Indeed, in the estimation of model (7), the results of which, along with that of the models (8) and (9), are shown in Table 29, no significant regression weight for the interaction term emerged. However, the main goal is not to establish the total effect of the *Country of origin* × *Regulatory focus* interaction on *Punishment per se* but rather to examine the plausibility of the causal structure linking that interaction to *Anger* and, in turn, *Anger* to *Punishment*. Therefore the lack of a significant interaction is ignored here, recognizing that descriptively, such a pattern holds (see also Kenny, Kashy, & Bolger, 1988; MacKinnon et al., 2002, for similar arguments that one may ignore the lack of a significant total effect in simple, unmoderated mediation analysis).

As for the second condition, Muller et al. (2005) specify in the first alternative that the interaction of the independent variable and the moderator exert a similar influence on the mediator as on the dependent variable (i.e., β_{83} in the model predicting the mediator from the independent variable, the moderator and their interaction, be significant) while the path from the mediator to the dependent variable simply linearily 'transmit' that effect (i.e., β_{94}, the 'main

effect'[17] of the mediator on the dependent variable in the model predicting the independent variable from the dependent variable from the mediator while controlling for direct influences of the dependent variable, the moderator and their interaction as well as possible influence of the mediator on the dependent variable moderated by the moderator). Alternatively, the relationship of the independent variable with the mediator may be a simple linear relationship (i.e., β_{81} is significant) rather than being moderated, but that the moderator determines the strength of the relationship between the mediator and the dependent variable (i.e., β_{95} is significant, representing the moderation of the effect of the mediator on the dependent variable by the moderator, while controlling for direct influences of the dependent variable, the moderator and their interaction as well as a possible 'main effect' of the mediator on the dependent variable). In the present case, the combined evidence so far from the meta-analysis and the MANOVA suggest that there is an interactive effect of the dependent variable and the moderator (i.e., a *Country of origin* × *Regulatory focus* interaction) on the mediator (*Anger*), and the consistent bivariate correlation of *Anger* and *Punishment* in the individual studies speaks for the simple linear relationship, i.e., a 'main effect' of the mediator on the dependent variable (*Punishment*). Therefore it is expected that the first of the alternative conditions holds, namely that β_{83} and β_{94} will be significant.

The estimates for the three models are displayed in Table 29. Indeed the predicted pattern of coefficients emerges: β_{83} and β_{94} are both significant. Thus, *Regulatory focus* moderates the influence of *Country of origin* on *Anger* (β_{83}), as already found in the meta-analysis as well as the MANOVA over the pooled cases, and the relationship between *Anger* and *Punishment*

17 The term main effect is used in quotes here, as in the presence of a significant interaction term the zero-order term representing the influence of one single variable by itself is, technically speaking, not a main effect but the effect if all other centered predictors have their mean value (see Aiken & West, 1991). As a side note, this is also often true with main effects in common ANOVAs. While the falsity of such terminology is admitted and the more accurate term 'conditional effect' (Aiken & West, 1991) is recognized, the imprecise terminology is presently preserved because the term main effect more familiarly conveys the conceptual idea.

remains significant if direct influences of the *Country of origin*, *Anger*, their interaction as well as the interaction of *Anger* and *Regulatory focus* are controlled for (β_{94}). Thus, even though the direct effect of the *Country of origin* × *Regulatory focus* interaction on *Punishment* is not significant, a causal path describing the intensity of recommended *Punishment* to be determined by *Anger* which itself is affected by the independent variable and the moderator is quite plausible. Further work measuring *Anger* differently from how it was achieved here (self-report questionnaire items) or even omitting it could nevertheless provide more certainty about a statistically secured effect of a valenced social category (or other valence-relevant characteristics) and processing style on *Punishment.*

16 General Discussion

This concluding overview will proceed in a sequence different from that in which the hypotheses and empirical results were discussed. First, because of the close relatedness in terms of theory as well as in predictions and empirical results, the studies pertinent to the Intergroup Punishment Difference and the Prior Positive Affect Hypothesis will be discussed along with their consequences for the hypotheses and conclusions to be drawn for future research. In a similar manner, the second part will sum up results of the studies testing the Reversed Black Sheep Effect Hypothesis.

16.1 Intergroup Punishment Difference and Prior Positive Affect Hypotheses

From a theory of punishment tendencies (Chapters 2 and 3) it was predicted that punitive tendencies towards a criminal offender from from a social category pretested to be relatively positive should be less intense than toward one from a less positively pretested category.

Additional assumptions about a prior association of ingroups with automatic general positive affect (Chapter 4) together with the foregoing reasoning lead to the more specific prediction that an ingroup offender should be reacted to less negatively than an outgroup offender.

Intergroup Punishment Difference Hypothesis

Regarding the studies manipulating ingroup versus outgroup membership of the target in the first complex (Chapter 9), the hypothesis of a simple main effect on punitive reactions received partial support. Across eight empirical studies using diverse intergroup distinctions (hometown of university students versus a different university town, German born versus not German born, college student versus full-time worker, young versus old), a small marginal intergroup difference as formulated in the original hypothesis of a simple group membership

main effect emerged (d_m = -0.09). Looking at lighter and heavier offenses separately reveals that this difference is relatively clear and consistent with predictions for heavier offenses (d_m = -0.15), but for lighter offenses, the difference is descriptively in the opposite direction and does not significantly differ from zero (d_m = 0.06). A test for moderation (i.e., a difference between the two mean effect sizes) was however not significant.

In sum, the present evidence clearly suggests that mere categorization into ingroup and outgroup results in differential intensity of recommended punishment for a severe transgression.

Contrary to the results for punitive tendencies in the summarized studies in this complex, which at least partially supported the hypotheses, evaluation measures did not yield any evidence consistent with predictions. If anything, effect sizes were positive, indicating that participants tended to evaluate ingroup targets less positively than outgroup targets. This tendency is far from conventional levels of significance and therefore the discussion here is overshadowed by utmost tentativity.

One might, as it is really appropriate in the logic of significance testing, treat these effect sizes as essentially equal to zero and claim that evaluation measures as those used here do not pick up intergroup differences in the context of norm violations. This is all the more plausible in light of the fact that the adjective scales were bipolar: they were commonly anchored by a positive adjective and its negative opposite. As mentioned earlier, research by Wenzel and Mummendey (1996) and Otten et al. (1998) found elimination of intergroup differences if evaluation scales comprised negative dimensions. Thus, the adjective scales here may have triggered normative concerns and lead to unbiased evaluations even in this case of a negative behavior of the target. On the other hand, this may seem far-fetched considering the fact that the targets clearly behaved in a norm-violating way and therefore, according to the general

punishment theory advanced throughout this report, should have attracted considerable negativity[18] and negative evaluation should be legitimate. However, occasional commentaries from participants during the data collection suggested that the evaluation of a person of which one only knows one single behavior on traits as those used seemed unusual and 'not possible'. This may not have been the case for punishment recommendations which are entirely appropriate and normal in the context of a criminal offense. Thus, it might make a difference whether one evaluates a criminal (or more specifically his personality in a global sense) or recommends treatment for him or her. The latter judgment may very well be colored by positive affect associated with an ingroup and lead to an intergroup bias, as that judgment is made without much hesitation. Evaluation on the other hand, which could be subject to doubts about its legitimacy, may well be made more carefully and thus potential biases on them are eliminated. This latter idea is clearly consistent with research on the positive-negative asymmetry (Chapter 4.1), while results regarding treatment recommendations are reminiscent of true outgroup derogation.

From a different perspective on this disparity between evaluations and punishment recommendations, more speculative and oriented toward future research, one may hypothesize that there is indeed a tendency for a BSE (Chapter 5) on evaluations while the reversed pattern (consistent with the 'classic' intergroup bias) emerges for punishment recommendations and thus for more treatment (or even behaviorally) oriented measures. Incidentally, a result from Marques et al. (2001) concurs with such an argument: In Studies 2 and 3 of this article, the more negative evaluation of an ingroup deviant than that of an outgroup deviant is accompanied by a higher intention to influence an ingroup versus an outgroup deviant to change his deviating opinion back to the descriptively normative one.

18 The relatively high means of evaluation indices (with high values indicating more negative evaluation), well above the mid-point of the seven point scales in most cases, suggest that indeed such negativity was elicited.

This intention could be the expression of higher benevolence towards ingroup deviants, just as lower punishment recommendation is. Such attempts at persuasion (and consequently potential re-inclusion) may be conceived of as lenient in nature, while recommendations of more punishment amount to the infliction of more harm. Thus, opening another path for future research, it is worthwhile to examine whether the BSE is restricted to evaluations, but generally reversed when it comes to measures of intended behavior or behaviorally oriented treatment recommendations.

One may prefer either approach to the disparity of evaluation and punishment recommendation across the studies in this complex, but it seems that this issue holds an interesting line of research – especially considering that evaluations, where corresponding measures were administered, were made *before* punishment recommendations and thus the lack of an effect on them (or even effects contradicting those on punishment recommendations) cannot be dismissed as a mere dissipation phenomenon.

In sum, the presently reported work started from the theoretical proposition that treatment recommendations and thereby judgments relevant for future behavior (at least that of actually punishing third parties[19]) will be more favorable for ingroup targets than for outgroup targets. The presently reported empirical studies provide evidence for this proposition regarding relatively severe transgressions. For less severe transgressions as well as on evaluations, no similar difference was found. Thus, the present work successfully identified a context (severe offenses) in which true outgroup derogation in the sense of more negative treatment of a

19 It should be noted that common positive dimensions on which intergroup differences are found are also often in their outcome relatively inconsequential (e.g., small amounts of money allegedly given to alleged others by the experimenter, evaluations of a group product resulting in a small price or award for those with the best evaluations) or measures of intention rather than actual behavior. The present measure of punishment recommendations is admittedly far from even tapping into intentions for own behavior, but they are clearly related to the appropriateness of harm being inflicted on the offender, possibly by third parties.

target from an outgroup than one from the ingroup occurs, based on – in the present context – rather trivial categorization.

Outgroup members may not be met with hostility 'out of the blue', but they are met with more intense negativity in treatment recommendation if they have done something very bad. An ingroup transgressor on the other hand, behaving in the same way, may be accorded some amount of goodwill as far as his treatment is concerned – even if he might be evaluated more negatively.

The following discussion of results concerning the Prior Positive Affect Hypothesis will expand the present discussion of intergroup differences to a more general level. In its course, another possible explanation for the disparity between lighter and heavier offenses in the Intergroup Punishment Difference Hypothesis studies will be provided (see p. 170 below) and light will be shed on the process hypothesized to underly the difference found for heavier offenses (i.e., an initial association with positive affect).

The Prior Positive Affect Hypothesis

It was found that membership in a third party category pretested to be evaluated generally more positive buffered anger towards an offender compared to membership in a category initially evaluated less positively. This was however only the case in promotion focus and thus presumably in a spontaneous processing mind set. Under prevention focus, which is characterized by careful and analytic processing, the opposite pattern emerged. This pattern on anger was preserved in punishment recommendations, although not significantly. Furthermore, it could be shown that punishment recommendation is a plausible causal consequence of anger and anger is determined by the target's social category (and thus presumably by its positive affect associated with that category) and regulatory focus interactively.

Thus indeed, consistent with the main hypothesis, if processing is spontaneous and not much deliberated, emotional reactions (here: anger) to a norm violation are affected by positive affect associated with third party categories, in that they buffer these reactions. These anger reactions in turn determine punitive reactions. Consistent with the moderation hypothesis, however, under a careful and scrutinizing mind set, the difference in anger is even reversed. This reversal could plausibly (but post-hoc and thus suggesting further research) be explained by a correction process: If information is processed carefully, a perceiver may well become aware that he or she has a bias against one of the offenders' social categories. But if he or she is not aware of the fact that this bias is based on the affect associated with the category (and is thus unable to accurately correct for it) or overestimates that bias and the amount of correction is larger than a spontaneous bias would have been (Wegener & Petty, 1997), the correction can go awry. Spontaneous bias against one target can turn into bias in favor of that target. Such an overcorrection hypothesis however also awaits more focused empirical testing.

Inspection of the studies shows that the pattern for anger was unequivocal for studies A-1, A-3, and A-4, while it was somewhat different (no simple main effects of category) in Study A-2. The reason for this may plausibly be the different type of mind set manipulation. Recall that while in the remaining three studies, the mind set was activated by the Friedman and Förster (2001) maze task, originally designed to induce promotion versus prevention focus, in Study A-2, in an attempt to isolate the specific facet of regulatory focus relevant to spontaneous versus careful processing, participants were directly instructed to either give their spontaneous opinion or to carefully think about their answers because of the sensitive nature of the topic. One may take this as an indication that the aspect of the regulatory focus

manipulation by means of the maze task is not actually the described mind set central to theorizing here.

Alternatively (and this interpretation is presently favored), the direct instruction may not be effective in manipulating a spontaneous processing mind set while the indirect manipulation with the seemingly unrelated maze task is. Such evidence comes from Sassenberg, Kessler and Mummendey (2006, see also Sassenberg & Moskowitz, 2005). They found that compared to mind sets (relatively complex collections of procedural knowledge, Bargh & Chartrand, 2000; Gollwitzer, Heckhausen, & Steller, 1990), the operation of which people may not be aware of, direct instructions to behave in line with a mind set (in their case: creativity) were unsuccessful while specific manipulation of the concepts and rules, unbeknownst to participants, may be more promising. Thus, the maze task used in the studies reported here may have provided a reliable manipulation while direct instruction failed to induce the desired processing mode. Instruction could even have induced a careful processing mind set for all participants and thus interrupted automaticity in general. Knowledge of procedures and strategies of mind sets could be truly implicit in the sense that people know how to act, but this knowledge is not accessible to a degree to which they can themselves verbalize it or be successfully told to apply the knowledge. Future research may thus extend the present results by examining the differences of instruction versus task manipulation more concisely. It may turn out that indeed a spontaneous processing mind set is activated most effectively and accurately by acting in it (as in the maze task) rather than by being explicitly instructed to apply it.

Referring back to results from the studies on the Intergroup Punishment Difference Hypothesis, the current moderation by processing style or mindset also suggests yet another possible explanation for why the effect on punishment was found for heavier offenses, but not

for lighter ones. While a heavier offense may be an unequivocally punishable behavior for which participants made their judgments spontaneously and without much deliberation, processing of lighter offenses could have been more deliberative and thus be subject to correction processes similar to those outlined here which lead to elimination of the difference (or even overcorrection). Similarly, evaluations in these studies may have been made in a more careful and reflective manner (as the anecdotal evidence of participant's hesitation to make these judgments mentioned earlier suggests), while punishment recommendations, possibly eliciting less suspicion, were made spontaneously and thus more influenced by biases, at least if the offense was clearly heavy and thus unequivocally punishable. In the same vein, the tendencies of BSEs found for evaluations and lighter offenses may be indicative of the heightened salience of presentation concerns due to a deliberative mind set (see also Chapter 16.2).

It is also, in a more general sense, possible that a number of participants within each of the studies investigating the Intergroup Punishment Difference Hypothesis were in a deliberative mind set (or prevention focus) and the remaining ones in a less deliberative one (or promotion focus). If there was a moderation of the effects of offender group membership and offense severity by mind set, effects in the subgroups (of differential mind set) may well have canceled each other out or at least attenuated each other. But unfortunately, there was no measurement nor manipulation of mind set in these studies. Such an idea should definitely be pursued in the future, as it is relatively easily testable by combining the manipulations employed here with the ingroup versus outgroup and the lighter versus heavier offense distinctions made in the studies testing the Intergroup Punishment Difference Hypothesis.

It may also be noted here that this reasoning opens up a new possibility of delineating boundary conditions for a BSE (see Chapter 5). If indeed initial association with positive

affect works in the ways hypothesized throughout this dissertation regarding punishment reactions and group membership, a BSE could be the result of participants answering questions in a mind set which is more similar to prevention focus than promotion focus. They would then process and answer in a careful, scrutinizing manner, avoiding or even overcorrecting for a judgment bias. This would be consistent with literature on the positive-negative asymmetry in intergroup evaluation: it has been argued – and there is empirical evidence consistent with that claim (Mummendey & Otten, 1998) –, that allocating negative resources alerts participants to the possibility of illegitimate discrimination and may motivate them to avoid it by more careful processing. In line with the SGD (Abrams et al., 2004, 2005), evaluation of a norm violation by an ingroup member may trigger more scrutiny, to ensure homogeneity and the descriptive norm within the ingroup – a careful mind set, in which ingroup favoritism is eliminated or even reversed. The BSE could be the outcome of a difference in information processing more distally caused by social identity concerns rather than being a direct outcome of the desire to protect positive ingroup distinctiveness and group norm legitimacy.

All in all, it has been argued above that in the context of punishment reactions, there are less legitimacy concerns with inflicting or recommending negative treatment and therefore discrimination should reoccur. The present results concerning the Prior Positive Affect Hypothesis suggest that this may be true, but only if it is ensured that processing occurs heuristically. The present research also provides evidence for the general process hypothesized to be operating in a difference in punishment recommendations between ingroup and outgroup. If participants make a spontaneous judgment, a target associated with an *a priori* more positively valenced category will elicit less negative and offensive feelings (anger) and probably less punitive tendencies than a target lacking such a buffer.

The effect just summarized is not large (mean effect sizes $d_m = 0.36$ for anger and 0.13 for punishment tendencies), but it may nevertheless develop considerable impact in a causal chain involving other processes pertinent to reactions to norm-violation (e.g., the ultimate attribution error, Pettigrew, 1979; Hewstone, 1990, which is not discussed further here).

The effect sizes in the studies testing the Intergroup Punishment Difference Hypothesis were also very small. But it must be kept in mind that the categorizations used in these studies were almost all relatively poor in content. The affective buffer working for an ingroup target may thus, to begin with, be smaller than that of a target from a category which was evaluated so decisively more positively than another in a pretest.

In sum, the current results are promising for further work on interactive functioning of affective buffers (be they general or in the form of ingroup versus outgroup distinctions) and states of processing style. Such research should employ more sensitive measures and particularly experimental manipulations of the small differences in automatic positive affect association. It is however presently clear that very small differences in automatic affect can have important effects on responses and behaviors, be those latter ones carefully and scrutinously thought through or not.

16.2 The Reversed Black Sheep Effect Hypothesis

The complex to be discussed lastly here is the one concerning the hypothesis that a BSE would be reversed under certain conditions. Specifically, the stronger derogation of a norm-violating ingroup member compared to an outgroup member showing the same behavior might well emerge if the behavior distributions within the two groups as a whole differed *a priori* (distinct norms), constituting positive ingroup distinctiveness. But if no such difference in the distributions is present, the reversal, that is ingroup favoritism, was predicted to re-

emerge. This hypothesis was tested using only evaluation as a dependent measure, as it is the common most pertinent variable in research on the BSE.[20]

For the simple comparison of ingroup and outgroup target evaluations, no BSE emerged across all studies. To the contrary, if the descriptive norm was manipulated to be distinct, under which condition the BSE was expected, ingroup favoritism was apparent. In the conditions without norms distinguishing between ingroup and outgroup, for which a reversal of the BSE was predicted, virtually no differences were found. For behavior evaluations, which were taken after person evaluations, there was also a tendency for ingroup favoritism, homogeneously over all norm distinctiveness conditions. Thus, overall, the results are inconsistent with the original hypothesis which must therefore be rejected. There is no reversal of a BSE under conditions of equal descriptive norms for a transgression and there is no BSE if that descriptive norm provides for positive ingroup distinctiveness which could be protected by a derogation of the ingroup target.

However, in the course of the individual studies it became apparent that the norm distinctiveness manipulation also had an effect on global evaluations of the entire groups of which the target offender was a member. Therefore it seemed plausible that a situational group stereotype was created by this manipulation and that target evaluations were assimilated to this stereotype. Thus, while the original hypothesis proved not viable with the currently reported experimental design, the results suggest an interesting extension of SGD (Abrams et al., 2005, 2005): Threats to positive social identity and ingroup distinctiveness are of concern to perceivers judging violators of overarching norms from an ingroup versus an outgroup and therefore and trigger negativity for the ingroup target. However, a clear valence

20 Future work may however fruitfully extend BSE research by treatment recommendation measures. This is interesting in the light of the disparity of results regarding evaluation and punitive tendencies discussed in the preceding section: Effects of an intergroup distinction showed generally different patterns as a function of the measure (evaluation vs. punishment recommendations).

of the behavior in question and a descriptive norm difference between the groups may counteract this tendency and lead to a reversal. This could be the reason why BSE phenomena reported in published studies often emerge for undesirable behaviors which are however not strongly valenced. Current data point to the possibility that a mild transgression by an ingroup member may elicit negative affect which a transgression by an outgroup member does not if the distributions of the behavior are perceived as distinct to the advantage of the ingroup. But at the same time, ingroup members profit from an assimilation of their evaluation to a positive ingroup stereotype which, in sum, eliminates ingroup derogation (or even a reverses it).

It is not entirely clear what consequences may be drawn from the present research in relation to research and theorizing about the BSE. Clearly, valence and group membership of deviants interact such that, after all, ingroup deviants may not necessarily be derogated relative to outgroup deviants as published research on the BSE suggests. However, the probably intricate interplay of norm distinctiveness, consequences of that distinctiveness for group and target valence, the severity of the offense and thus the appropriateness of punishment opens up a promising new line for future research.

16.3 Conclusions

Regarding the main topic of the present dissertation – differences in lay people's punishment recommendation for transgressors as a function of the valence connected to the target's social category – the present results are important in several ways.

For one thing, results show that ingroup members in fact enjoy an advantage over outgroup members in the recommendation of harmful treatment in response to severe norm violations. Thus, for intergroup research, there is a key here to understand outgroup derogation and how it occurs in the real world: While harm may not be inflicted onto outgroups and their members

because of their membership only, they will be responded towards more intensely if they have violated norms.

It is of course true, especially with larger social categories, that within every such category there are a few 'bad apples', people behaving in despicable ways and breaking laws which are to be obeyed by both members of the ingroup as well as outgroups. As current results show, the reactions toward outgroup transgressors are however a little less constrained than those against ingroup transgressors. In the eyes those then meting out punishment, maltreatment of outgroup members seems to be not driven by an intergroup bias: after all, the target has done something wrong and does deserve punishment.

Moreover, even if such punitive responses are not carried out by a perceiver him or herself, the appropriateness and legitimacy of 'a little rougher handling' of an outgroup villain will be considered appropriate by observers, too, and thus intervention on behalf of a victim of over-retribution may come later for an outgroup than an ingroup member.

Secondly, and on a more general level, current results alert to the undue influence of affect associated to any offender on punishment recommendations. Of course, participants of the current studies were no lawyers. Judges of law, for example, undergo extensive training (in written law and procedures designed not to allow for such influences) and usually probably process information relevant to their cases in a careful and scrutinizing manner. Therefore, it is questionable whether they would be influenced by affective biases in the same way as lay people, possibly spontaneously processing information. However, given that there are many cases for which punishment is subject to some latitude of judgment, relatively meaningless social categorization may play an important role, even without strong stereotypes and deep rooted prejudice working as, for example, in the case of white versus black offenders.

Also in lay courtroom juries, positive affect may have an impact on judgment and in some severe cases tip the balance in favor of a guilty verdict for an offender who is not associated with an affective buffer like the one proposed here. To be sure, a guilty verdict is different from the recommendation of punishment, but beyond considerations of responsibility (attributions), the full intensity of an affective reaction to a severe transgression may lead to the desire to punish one defendant for it (which will only happen once he or she is found guilty), while a buffer for another defendant, like the one proposed here, could ease that desire and lead to a more thorough weighing of evidence leading to acquittal (Sargent, 2004).

Finally, the processes for which evidence has presently been offered may have an impact on the treatment of offenders and norm violators more generally. For those deserving higher punishment in the eyes even of lay people who are not on a jury, more severe punishment is also more appropriate and legitimate. As already mentioned above, for an *a priori* affective bias to have considerable consequences, it is not even necessary that all or even most individuals in whose judgment the bias operates actually perform the punishment. It may just be enough if a few perform a harsher punishment which however is tolerated and accepted by those observing as appropriate and right. Then, a small difference in originally harmless preference for one but not another offender – of which a perceiver may not even be aware–, could have dramatic consequences for equal treatment under the law.

17 SUMMARY

The present dissertation tested three interrelated hypotheses regarding the influence of targets' group membership on evaluative and punitive reactions to norm violation.

Firstly, from a theory of punishment highlighting the role of negative affect in the recommendation of punishment (i.e., aversive treatment) and an approach characterizing ingroup versus outgroup distinctions by automatic positive affect being associated with the former, but not the latter, a bias on evaluation as well as recommended punishment severity in favor of an ingroup criminal offender was predicted (*Intergroup Punishment Difference Hypothesis*).

Inconsistent data patterns from a series of similar studies were integrated using meta-analysis. Results indeed provided evidence for a punishment bias, but only on punishment recommendation and for heavier offenses. For lighter offenses as well as on evaluation measures, no statistically significant overall effect was found. Descriptively however, there were differences opposite to that predicted on evaluations and punishment recommendation for lighter offenses, possibly indicating a *Black Sheep Effect* (Abrams et al., 2004, 2005). The possible disparity between punishment recommendation and evaluation is recommended as a topic for future research.

Also, following up on the idea that automatic association of a target with positive affect is responsible for relative leniency, similar studies were conducted testing the prediction that for third party groups (i.e., two outgroups), general positivity of evaluation would buffer anger (a central mediator in the punishment theory presently proposed) as well as punitive reactions toward offenders from these categories (*Prior Positive Affect Hypothesis*). This should especially be the case if information processing is spontaneous whereas an elimination or reversal of the effect was expected if information processing was deliberative. These

predictions were confirmed in the studies after meta-analytical integration: An offender of a previously rather positively evaluated nationality category elicited less anger and punishment if participants were in the spontaneous mind set of promotion focus, whereas a reversal of this effect was found under prevention focus, which is characterized by careful and analytical processing. The latter effect is here tentatively explained as an overcorrection effect (Wegener & Petty, 1997), but should be subjected to further research.

Finally, an objection to the Intergroup Punishment Difference Hypothesis may come from research on the *Black Sheep Effect* (Marques & Paez, 1994) and the theoretical development explaining it (*Subjective Group Dynamics*, Abrams et al., 2004). The *Black Sheep Effect* consists in relative derogation of an ingroup norm violator compared to a comparable outgroup violator. However, it is argued that the behaviors for which the *Black Sheep Effect* is commonly found (i.e., mildly received undesirable behaviors for which there are differential descriptive norms within the ingroup and outgroup, setting the ingroup positively apart from the outgroup) are different from strongly negatively received transgressions against overarching prescriptive norms discussed here. A set of studies examined whether indeed the prediction of the Intergroup Punishment Difference Hypothesis on evaluation ratings for a target would hold if norms are not descriptively different between ingroup and outgroup, but a reversal (i.e., a *Black Sheep Effect*) will occur if there is such a difference in descriptive norms (*Reversed Black Sheep Effect Hypothesis*).

A meta-analysis of a third set of studies however provided no support for the Reversed Black Sheep Effect Hypothesis. Instead, when ingroup and outgroup were described as differing in the descriptive norm, ingroup favoritism occurred, while no difference was found if the groups were described as not differing. Additionally, beyond evaluations of the target transgressor, evaluations of the entire groups (ingroup and outgroups) were measured.

Logically inherent in the intergroup difference in descriptive norms regarding norm violation, the manipulation of the descriptive norm difference induced a positive ingroup stereotype, which apparently colored ingroup target judgments positively, thus counteracting a *Black Sheep Effect*. If these group evaluations were controlled for, a *Black Sheep Effect* emerged, particularly in the conditions with a difference in descriptive norms between ingroup and outgroup.

In sum, the present work reports evidence consistent with the assumption that initial positive affect associated with a social category buffers against punishment recommendations if the transgression is severe and if perceivers are not in a careful, scrutinizing mind set as they make judgments. Also, as results regarding the Reversed Black Sheep Effect Hypothesis suggest, valence of a norm-violating behavior and its consequences for group stereotypes should be taken into account in research investigating the *Black Sheep Effect*.

18 ZUSAMMENFASSUNG

In der vorliegenden Dissertation werden drei Hypothesen getestet zum Einfluß von Gruppenmitgliedschaft auf Beurteilung und Strafempfehlung für ein Target, welches gegen eine Norm verstoßen hat.

Zunächst wurde vorhergesagt, dass ein Eigengruppenmitglied weniger negativ beurteilt werde und die Strafempfehlung milder ausfalle als für ein Fremdgruppenmitglied (*Intergruppen-Bestrafungsunterschied-Hypothese*). Diese Vorhersage wurde abgeleitet aus einer Bestrafungstheorie, die die Rolle von negativem Affekt in Bestrafungsempfehlungen betont, sowie der Annahme, dass Eigengruppen automatisch mit positivem Affekt assoziiert sind, Fremdgruppen jedoch nicht.

Inkonsistente Datenmuster aus einer Serie ähnlich strukturierter Studien zu dieser Hypothese wurden in einer Metaanalyse integriert. Das Ergebnis dieser Analyse stimmten teilweise mit der Vorhersage überein, allerdings zeigte sich der Unterschied nur für Bestrafungsempfehlungen für schwerere Vergehen. Für leichtere Vergehen sowie Beurteilungen fanden sich lediglich deskriptive Unterschiede, die jedoch in die umgekehrte Richtung deuteten. Sie sind eher konsistent mit einem *Black Sheep Effect* (Abrams et al., 2004, 2005). Die mögliche Dissoziation von Maßen zur Beurteilung und solchen zur Bestrafungsempfehlung wird zur weiteren Untersuchung empfohlen.

Der Idee folgend, dass automatische Assoziation mit positivem Affekt verantwortlich ist für relative Milde in Bestrafungsempfehlungen, wurde in weiteren Studien getestet, ob auch bei zwei Gruppen, von denen Beobachter kein Mitglied sind (d. h. zwei Fremdgruppen), die sich allerdings in global positiver Bewertung unterscheiden, positiver Affekt eine Ärgerreaktion (als zentrale mediierende Variable bei Bestrafungsimpulsen) sowie darauf folgende Bestrafungsempfehlungen mildert (*Positiver-Affekt-als-Puffer-Hypothese*). Dies

sollte insbesondere dann der Fall sein, wenn die Versuchsteilnehmenden die Information zum Vergehen spontan verarbeiten, während eine genaue und reflektive Verarbeitung zur Abschwächung oder gar zum umgekehrten Muster führen sollte.

Diese Vorhersagen fanden Bestätigung in einer ebenfalls metaanalytischen Zusammenfassung mehrerer Einzelstudien. Ein Täter im Pretest als positiver bewerteter Nationalität rief weniger Ärger und mildere Bestrafungsempfehlung hervor als einer weniger positiver Nationalität, wenn die Teilnehmenden Informationen spontan verarbeiteten. Unter genauer Verarbeitung hingegen zeigte sich ein umgekehrter Effekt. Diese Umkehr wurde einstweilen als Überkorrektur-Effekt erklärt (Wegener & Petty, 1997), sollte jedoch Gegenstand weiterer Forschung bleiben.

Schließlich mag aufgrund der Forschung zum *Black Sheep Effect* (Marques & Paez, 1994) und der theoretischen Erklärung dieses Effekts (*Subjective Group Dynamics*, Abrams et al., 2004) der Vorhersage und den Befunden zur Intergruppen-Bestrafungsunterschied-Hypothese widersprochen werden. Dieser *Black Sheep Effect* bezeichnet das Phänomen der Abwertung von normverletzenden Eigengruppenmitgliedern relativ zu vergleichbaren Fremdgruppenmitgliedern.

Allerdings unterscheiden sich die Normverletzungen, für die für gewöhnlich ein *Black Sheep Effect* gefunden wird (relativ milde Vergehen gegen für Eigengruppe und Fremdgruppe unterschiedliche deskriptive Normen, die für die Eigengruppe positive Distinktheit konstituieren) von den hier betrachteten: Das Augenmerk liegt gegenwärtig auf schweren Vergehen gegen übergreifende präskriptive Normen.

Eine weitere Reihe von Studien untersuchte, ob ein *Black Sheep Effect* zwar auftritt, wenn a priori die Normverletzung in Eigen- versus Fremdgruppe unterschiedlich häufig auftritt (deskriptive Norm), sich das Muster aber – im Einklang mit der Intergruppen-

Bestrafungsunterschied-Hypothese –, umkehrt, wenn kein deskriptiver Unterschied besteht (*Black-Sheep-Effect-Umkehrungs-Hypothese*). In einer dritten Metaanalyse über diese Einzelstudien bewährte sich die Black-Sheep-Effect-Umkehrungs-Hypothese nicht. Vielmehr wurde Eigengruppenfavorisierung gefunden, wenn die deskriptive Norm in Eigen- und Fremdgruppe als unterschiedlich beschrieben wurde. Wenn kein solcher Unterschied wahrgenommen wurde, unterschieden sich Beurteilungen von Eigen- und Fremdgruppenmitgliedern nicht.

Außer den Beurteilungen der Normverletzer wurden jedoch Beurteilungen der Gesamtgruppen (Eigen- und Fremdgruppe) erfragt. Es zeigte sich, dass der Unterschied in deskriptiven Normen offenbar ein positives Eigengruppen-Stereotyp induzierte, an welches die Beurteilung des Einzelverletzers aus der Eigengruppe assimiliert wurde – und somit gegen einen *Black Sheep Effect* wirkte. Nachdem die Targetbeurteilungen statistisch für dieses Stereotyp bereinigt worden waren, zeigte sich tatsächlich ein *Black Sheep Effect*, besonders dann, wenn ein Unterschied in deskriptiven Normen manipuliert worden war.

Zusammenfassend berichtet diese Arbeit Evidenz, die konsistent ist mit der Annahme, dass ursprüngliche gruppenbasierte Assoziation mit automatischem positivem Affekt als Puffer wirkt in Bestrafungsreaktion für schwerere Vergehen und wenn die Beurteilenden die Informationen zur Tat nicht sorgfältig verarbeiten. Darüberhinaus legen die Ergebnisse zur Black-Sheep-Effect-Umkehrungs-Hypothese nahe, in zukünftiger Forschung zum *Black Sheep Effect* die Valenz des normverletzenden Verhaltens sowie deren Konsequenzen für Gruppenstereotype zu berücksichtigen.

19 REFERENCES

Abrams, D., Marques, J. M., Bown, N., & Dougill, M. (2002). Anti-norm and pro-norm deviance in the bank and on the campus: Two experiments on Subjective Group Dynamics. *Group Processes and Intergroup Relations, 5*, 163-182.

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.

Abrams, D., Marques, J. M., Randsley de Moura, G., Hutchinson, P., & Bown, N. J. (2004). The maintenance of entitativity: A Subjective Group Dynamics approach. In V. Yzerbyt, C. M. Judd, & O. Corneille (Eds.), *The psychology of group perception. Perceived variability, entitativity, and essentialism* (pp. 361-379). New York, NY: Psychology Press.

Abrams, D., Randsley de Moura, G., Hutchinson, P., & Viki, G. T. (2005). When bad becomes good (and vice versa): Why social exclusion is not based on difference. In D. Abrams, M. A. Hogg, & J. M. Marques (Eds.), *The social psychology of inclusion and exclusion* (pp. 161-189). New York, NY: Psychology Press.

Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development, 74*, 1840-1856.

Abrams, D., Rutland, A., Cameron, L., & Marques, J. M. (2003). The development of subjective group dynamics: When in-group bias gets specific. *British Journal of Developmental Psychology, 21*, 155-176.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks: Sage.

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

Bargh, J. A. (1996). Automaticity in social psychology. In T. E. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169-183). New York, NY: Guilford.

Bargh, J. A., & Chartrand, T. L. (2000). Studying the mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253-285). Cambridge: Cambridge University Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. I, pp. 680–740). Boston, MA: McGraw-Hill.

Bègue, L. (2001). Social judgment of abortion: A Black-Sheep Effect in a catholic sheepfold. *The Journal of Social Psychology, 141*, 640-649.

Bentham, J. (1962). Principles of penal law. In J. Bowring (Ed.), *The works of Jeremy Bentham* (p. 396). Edinburgh: W. Tait (original work published in 1843).

Bettencourt, B. A., Dill, K. E. Greathouse, S., Charlton, K., & Mullholland, A. (1997). Predicting evaluations of ingroup and outgroup members: The role of category-based expectancy violation. *Journal of Experimental Social Psychology, 33*, 244-275.

Blanz, M., Mummendey, A., & Otten, S. (1997). Normative evaluations and frequency expectations regarding positive and negative outcome allocations between groups. *European Journal of Social Psychology, 27*, 165-76.

Bodenhausen, G. V. (1993). Emotions, arousal, and stereotypic judgments: A heuristic model of affect and stereotyping. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping* (pp. 13-37). San Diego, CA: Academic Press.

Bodenhausen, G. V., Mussweiler, T., Gabriel, S., & Moreno, K. N. (2001). Affective influences on stereotyping and intergroup relations. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 319-343). Mahwah, NJ: Erlbaum.

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). Ingroup or outgroup extremity: Importance of the threatened identity. *Personality and Social Psychology Bulletin, 19*, 381–388.

Brewer, M. B. (1999). The psychology of prejudice: ingroup love or outgroup hate? *Journal of Social Issues, 55*, 429–444.

Brewer, M. B. (2003). *Intergroup relations* (2nd edition). Buckingham: Open University Press.

Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology, 70,* 661–677.

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology, 42*, 437-451.

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.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 201-234). San Diego, CA: Academic Press.

Clement, R. W., & Krueger, J. (2002). Social categorization moderates social projection. *Journal of Experimental Social Psychology, 38,* 219–231.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Crocker, J., & Luhtanen, R. (1990). Collective self-esteem and ingroup bias. *Journal of Personality and Social Psychology, 58,* 60–67.

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.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*, 324-336.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305*, 1254-1258.

DeCoster, J. (2004). *Meta-analysis notes*. Retrieved June 26, 2006, from www.stat-help.com/meta.pdf.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Donnerstein, M., & Donnerstein, E. (1973). Variables in interracial aggression: Potential group censure. *Journal of Personality and Social Psychology, 43*, 143-150.

Donnerstein, M., & Donnerstein, E. (1975). The effects of attitudinal similarity on interracial aggression. *Journal of Personality, 43*, 485-502.

Donnerstein, M., & Donnerstein, E. (1978). Direct and vicarious censure in the control of interracial aggression. *Journal of Personality, 48*, 162-175.

Donnerstein, M., Donnerstein, E., Simon, S., & Ditrichs, R. (1972). Variables in interracial aggression: Anonymity, expected retaliation and a riot. *Journal of Personality and Social Psychology, 22*, 236-245.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*, 785-791.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*, 1-25.

Fein, S., Hoshino-Browne, E., Davies, P. G., & Spencer, S. J. (2003). The role of self-image maintenance in stereotype activation and application. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated social perception: The Ontario Symposium* (Vol. 9, pp. 21–44). Mahwah, NJ: Erlbaum.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*, 39-66.

Förster, J., Higgins, T. E., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes, 90*, 148-164.

Friedman, R. S., & Förster, J. (2001). The effects of promotion and prevention cues on creativity. *Journal of Personality and Social Psychology, 81*, 1001-1013.

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*, 212-228.

Gollwitzer, P.M., Heckhausen, H., & Steller, B. (1990). Deliberative and implemental mind-sets: Cognitive tuning toward congruous thoughts and information. *Journal of Personality and Social Psychology, 59*, 1119-1127.

Gordon, R. A. (1990). Attributions for blue-collar and white-collar crime: The effects of subject and defendant race on simulated juror decisions. *Journal of Applied Social Psychology, 20*, 971-983.

Gordon, R. A., Bindrim, T. A.,& McNicholas, M. L. (1988). Perceptions of blue-collar and white-collar crime: The effect of defendant race on simulated juror decisions. *Journal of Social Psychology, 128*, 191-197.

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.

Gramzow, R. H., Gaertner, L., & Sedikides, S. (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.

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.

Hewstone, M. (1990). The "ultimate attribution error"? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology, 20*, 311-335.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*, 575–604.

Hodson, G., Hooper, H., Dovidio, J. F., & Gaertner, S. L. (2005). Aversive racism in Britain: The use of inadmissible evidence in legal decisions. *European Journal of Social Psychology, 35*, 437-448.

Johnson, J. D., Whitestone, E., Jackson, L. A., & Gatto, L. (1995). Justice is still not colorblind: Differential racial effects of exposure to inadmissible evidence. *Personality and Social Psychology Bulletin, 21*, 893–898.

Jones, J. T., Pelham, B. W., Mirenberg, M. C., & Hetts, J. J. (2002). Name-letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology, 38*, 170–177.

Kant, I. (1990). *Metaphysik der Sitten.* Ditzingen: Reclam (original work published 1797).

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233-265). Boston, MA: McGraw-Hill.

Kernahan, C., Bartholow, B. D., & Bettencourt, B. A. (2000). Effects of category-based expectancy violation on affect-related evaluations: Toward a comprehensive model. *Journal of Applied Social Psychology, 22*, 85-100.

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin, 23,* 736–742.

Krueger, J. (1998). On the perception of social consensus. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 163–240). San Diego, CA: Academic Press.

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24*, 563–574.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104.

Marques, J. M. (1990). The black sheep effect: outgroup homogeneity in social comparison settings. In D. Abrams & M. Hogg (Eds.), *Social Identity Theory: Constructive and critical advances* (pp. 131-151). London: Harvester Wheatsheaf.

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.

Marques, J. M.& Paez, D. (1994). The black sheep effect: Social categorization, rejection of ingroup deviates, and perceptions of group variability. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 5, pp. 37-68). New York, NY: Wiley.

Marques, J. M., Robalo, E. M., & Rocha, S. A. (1992). Ingroup bias and the 'black sheep' effect: Assessing the impact of social identification and perceived variability on group judgments. *European Journal of Social Psychology, 22*, 331-352.

Marques, J. M., & Yzerbyt, V. Y. (1988). The black sheep effect: Judgemental extremity towards ingroup members in inter- and intra-group situations. *European Journal of Social Psychology, 18*, 287-292.

Marques, J. M., & Yzerbyt, V. Y., & Leyens, J.-Ph. (1988). Extremity of judgments towards ingroup members as a function of ingroup identification. *European Journal of Social Psychology, 18*, 1-16.

McFatter, R. M. (1978). Sentencing strategies and justice: Effects of punishment philosophy on sentencing decisions. *Journal of Personality and Social Psychology, 36*, 1490–1500.

McFatter, R. M. (1982). Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness. *Journal of Applied Psychology, 67*, 255–267.

Moreno, K. N., & Bodenhausen, G. V. (2001). Intergroup affect and social judgment: Feelings as inadmissible information. *Group Processes Intergroup Relations,* 4, 21-29.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When Moderation Is Mediated and Mediation Is Moderated. *Journal of Personality and Social Psychology, 89*, 852-863.

Mummendey, A., & Otten, S. (1998). Positive-negative asymmetry in social discrimination. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 9, pp. 107-143). Chichester, UK: Wiley.

Mummendey, A., & Otten, S. (2001). Aversive discrimination. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 112-132). Malden, MA: Blackwell.

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-negative asymmetry in social discrimination: Valence of evaluation and salience of categorization. *Personality and Social Psychology Bulletin, 26*, 1258-1270.

Newman, L. S., & Erber, R. (Eds). (2002). *Understanding genocide: The social psychology of the Holocaust*. New York, NY: Oxford University Press.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100*, 530-542.

Nuttin, J. M. (1985). Narcissism beyond gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15,* 353–361.

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211-238). New York, NY: Guilford.

Otten S. (2002). 'Me' and 'us' or 'us' and 'them'? — The self as heuristic for defining novel ingroups. In W. Stroebe & M.. Hewstone (Eds.), *European review of social psychology* (Vol. 13, pp.1–33). New York: Psychology Press.

Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-based spontaneous trait inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology, 36*, 77-89.

Otten, S., Mummendey, A., & Buhl, T. (1998). Information processing and social discrimination: Accuracy in information processing and positive-negative asymmetry in social discrimination. *Revue Internationale de Psychologie Sociale, 11*, 69-96.

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29*, 1049-1071.

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.

Pettigrew, T. F. (1979) The ultimate attribution error: Extending Allport's cognitive analysis of prejudice, *Personality and Social Psychology Bulletin, 5*, 461-476.

Rabkov, I. (2006). *Aussiedler in der Bundesrepublik Deutschland. Migrationserfahrungen und Kriminalitätsrisiken von ethnischen Migranten im Kontext der bundesdeutschen Zuwanderungspolitik* [Ethnic Germans in the Federal Republic of Germany. Migration experiences and crime risk of ethnic immigrants in the context of German immigration policy]. Freiburg: Unpublished Dissertation. Retrieved August 07, 2006 from http://www.freidok.uni-freiburg.de/freidok/volltexte/2006/2508/pdf/Rabkov_Dissertation.pdf

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54,* 890–902.

Robbins, J. M., & Krueger, J. I. (2005) Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review, 9*, 32-47.

Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law.* Boulder, CO: Westview Press.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59-82.

Rucker, D., Polifroni, M., Tetlock, P. E., & Scott, A. L. (2004). On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin, 30*, 673–684.

Sargent, M. J. (2004). Less thought, more punishment: Need for Cognition predicts support for punitive responses to crime. *Personality and Social Psychology Bulletin*, 30, 1485-1493.

Sassenberg, K., Kessler, T., & Mummendey, A. (2006). *When creative means different. Activating creativity as a strategy to initiate the generation of original ideas*. Jena: Unpublished manuscript.

Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology, 41*, 506-514.

Schubert, T., & Otten, S. (2000). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self & Identity, 1*, 353-376.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.

Schwarzer, R. (1989). *Computer Programs for Meta-Analysis 5.3* [Computer software and manual]. Retrieved May 26, 2006, from http://web.fu-berlin.de/gesund/gesu_engl/meta_e.htm.

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87*, 38-56.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445.

Sidanius, J. & Pratto, F. (2001). *Social Dominance: An intergroup theory of social hierarchy and oppression*. Cambridge: Cambridge University Press.

Sommers, S. R., & Ellsworth, P. C. (2000a). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin, 26*, 1367-1379.

Sommers, S. R., & Ellsworth, P. C. (2000b). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law, 7*, 201–229.

Spears, R., Jetten, J., & Scheepers, D. (2002). Distinctiveness and the definition of collective self: A tripartite model. In A. Tesser & D. A. Stapel (Eds.), *Self and motivation: Emerging psychological perspectives* (pp. 147-171). Washington, DC: American Psychological Association.

Specht, S. F. (1975). The role of affect in party discipline. J*ournal of Huschel Politics, 31*, 23-27.

Strack, F., & Hannover, B. (1996). Awareness of influence as a precondition for implementing correctional goals. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 579-596). New York: Guilford Press.

Strack, F., Schwarz, N., Bless, H., Kübler, A., & Wänke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology, 23*, 53-62.

Tajfel, H., Flament, C., Billig, M. G., & Bundy, R. P. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 149 – 178.

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

Trafimow, D., Bromgard, I. K., Finlay, K. A., & Ketelaar, T. (2005). The role of affect in determining the attributional weight of immoral behaviors. *Personality and Social Psychology Bulletin, 31*, 935-948.

Uleman, J. S., Hon, A. K., Roman, R. J., & Moskowitz, G. B. (1996). Reaction time evidence for spontaneous trait inferences on-line. *Personality and Social Psychology Bulletin, 22*, 377–394.

van den Bos, K. (2003). On the subjective quality of social justice: The role of affect as information in the psychology of justice judgments. *Journal of Personality and Social Psychology, 85*, 482-498.

Vidmar, N. J. (2000). Retribution and revenge. In J. Sanders & V. L. Hamilton (Eds.), *Handbook of justice research in law* (pp. 31-63). New York, NY: Kluwer.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141-208). San Diego, CA: Academic Press.

Weiner, B. (1984). Psychological colloquia as exercises in impression management. *American Psychologist, 39*, 926.

Weiner, B. (1995). *Judgments of responsibility : A foundation for a theory of social conduct.* New York, NY: Guilford.

Weiner, B., Graham, S., & Reyna, C. (1997). An attributional examination of retributive versus utilitarian philosophies of punishment. *Social Justice Research, 10*, 431-452.

Wenzel, M., & Mummendey, A. (1996). Positive-negative asymmetry of social discrimination: A normative analysis of differential evaluations of in-group and out-group on positive and negative attributes. *British Journal of Social Psychology, 35*, 493-507.

Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). CyberOstracism: Effects of being ignored over the Internet. *Journal of Personality and Social Psychology, 79,* 748-762.

20 ACKNOWLEDGMENTS

This work has been done in memory of my parents, I hope they would be proud.

This dissertation would have not been possible without the support and help of Kai Sassenberg and Sabine Otten. I am deeply indebted to them.

I also thank Maya Machunsky, Thomas Schubert, Ilga Vossen, Katharina Fuchs-Bodde, Frank Riedmann, and Barbara Schauenburg for their support throughout the years this work has taken. Tobias Raabe, Claudia Schneider, Karoline Hölzer, Sandra Preißler, Kristin Wenzel and Tina Treml have provided invaluable help in collecting data. Doreen Menzel, Steffen Giessner, and Angelina Elia have been a pleasure to share an office with.

Important and fascinating moments were given to me by Gayannée Kédia.

John R. Cash has made me rich.

I have to thank the entire *squadra azzurra* and particularly Fabio Grosso who saved the world from an unleashed patriotic frenzy which I could not have stood.

This manuscript has been typeset in OpenOffice.org 2.0.3.

Longest live Israel. May it resist its enemies who love death and hate life. And defend itself by all means deemed appropriate.

Finally, a special thank you is due to Claudie whom I am very happy to hang out with.

Lebenslauf

geboren am 23. August 1975 in Düsseldorf, ledig

1981-1985	Schulbesuch *Grund- und Hauptschule am Eisenberg* in St. Ingbert-Hassel (Saarland)
1985-1992	Schulbesuch *Deutsch-Französisches Gymnasium Saarbrücken*
1992-1993	Schulbesuch *C. W. Baker High School*, Baldwinsville, NY, USA; Austauschschüler im Rahmen des Parlamentarischen Patenschaftsprogrammes des *Deutschen Bundestages* und des *Congress of the United States* mit *Deutsches Youth for Understanding Komitee e. V.*
1993-1995	Schulbesuch *Deutsch-Französisches Gymnasium Saarbrücken*
Juni 1995	Deutsch-Französisches Abitur, Note: 1.0
September 1995- September 1996	Zivildienst *Psychagogisches Kinderheim Rittmarshausen e. V.*, Rittmarshausen bei Göttingen
WiSe 1996/97- SoSe 1999	Studium der Psychologie (Dipl.), *Georg-August-Universität Göttingen*
WiSe 1999/00- SoSe 2000	Auslandsstudium an der *University of California, Irvine*, CA, USA
WiSe 2000/01- SoSe 2003	Studium der Psychologie (Dipl.), *Georg-August-Universität Göttingen*
11. 08. 2003	Hochschulabschluß Diplom-Psychologe, *Georg-August-Universität Göttingen* Titel der Diplomarbeit: *What Time Is It? – It Depends On Where You're Going. Relevance and Precision in Telling the Time* (Gutachterinnen: Prof. Dr. Margarete Boos, Prof. Dr. Michael Waldmann)
01. 09. 2003- 31. 08. 2006	wissenschaftlicher Mitarbeiter in der DFG-Forschergruppe *Discrimination and Tolerance in Intergroup Relations* (www.uni-jena.de/svw/rgroup/), *Friedrich-Schiller-Universität Jena*, Teilprojekt *Soziale Identitäten und Aggressive Interaktionen: Die Rolle von Gruppenzugehörigkeiten bei der Wahrnehmung, Interpretation und Handlungsauswahl in sozialen Konflikten.*
seit 13. 10. 2003	Promotionsstudent an der *Friedrich-Schiller-Universität Jena*, *Fakultät für Sozial- und Verhaltenswissenschaften*

Jena, den 14.08.06

Johann Jacoby

Veröffentlichungen

Jacoby, J. (2003). *What Time Is It? – It Depends On Where You're Going. Relevance and Precision in Telling the Time*. Unveröffentlichte Diplomarbeit, Georg-August-Universität, Göttingen.

Jacoby, J., & Otten, S. (2006). *I Don't Like You, But I Still Want To Play With You. Retaliation Effects In Response To Exclusion In The Cyberball Game?* Unveröffentlichtes Manuskript, Friedrich-Schiller-Universität Jena.

Otten, S., & Jacoby, J. (2006). *Aggression from ingroup and outgroup members: differential retaliation effects?* Manuskript in Vorbereitung, Rijksuniversiteit Groningen.

Sassenberg, K. Moskowitz, G. B., Jacoby, J., & Hansen, N. (in Druck). The carry-over effect of competition: The impact of competition on prejudice towards uninvolved outgroups. *Journal of Experimental Social Psychology*.

Ehrenwörtliche Erklärung zur Dissertationsschrift

Punishing 'Them' Harder?
An Investigation of Group Differences in Reactions to Norm Violation

Durch meine Unterschrift versichere ich

1. dass mir die geltende Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena bekannt ist,
2. dass ich die Dissertation selbst angefertigt, insbesondere die Hilfe eines Promotionsberaters nicht in Anspruch genommen, sowie alle von mir benutzten Hilfsmittel und Quellen in meiner Arbeit angegeben habe,
3. dass mir Tobias Raabe, Claudia Schneider, Karoline Hölzer, Sandra Preißler, Kristin Wenzel und Tina Treml bei der Durchführung der Studien und der Dateneingabe assistiert haben im Rahmen ihrer Tätigkeit als studentische Hilfskräfte im Projekt *Soziale Identitäten und Aggressive Interaktionen* der DFG-Forschergruppe *Discrimination and Tolerance in Intergroup Relations,* während an der Auswahl und Auswertung des Materials außer mir niemand beteiligt war,
4. dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
5. dass ich die vorgelegte Arbeit noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe,
6. dass ich nicht die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule bzw. anderen Fakultät als Dissertation eingereicht habe sowie
7. dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Jena, den 14.08.06

Johann Jacoby