American Psychological Association

Journal of Personality and Social Psychology:
Interpersonal Relations and Group Processes

0022-3514/19/$12.00

2019, Vol. 117, No. 1, 124–141
http://dx.doi.org/10.1037/pspi0000165

Compassion Magnifies Third-Party Punishment

Stefan Pfattheicher
Aarhus University and Ulm University

Claudia Sassenrath and Johannes Keller
Ulm University

The last decades of research have provided overwhelming evidence that compassion fosters a vast range of behaviors toward reducing suffering of others. In this regard, compassion has been described as a prosocial tendency par excellence, fostering helping behavior across a variety of social situations. With the present contribution, we apply a differentiated perspective on compassion. Building on just deserts theory, we argue that when other individuals suffer from unjust actions, compassion for the suffering individuals can foster *harmful* tendencies toward those who caused the suffering (i.e., third-party punishment). In Studies 1a to 1f, we examined a rich variety of situations in which unjust suffering occurs (i.e., terrorist attacks, sexual assaults, rape of children, and war) and documented a positive relation between compassion for suffering victims and punishment inclinations toward those who caused the suffering. Applying an experimental approach using various paradigms in Studies 2 through 6, compassion was shown to increase third-party punishment. Additional analyses revealed that (a) this increase occurs because compassion intensified moral outrage, which in turn predicted third-party punishment (Studies 2 to 6), and (b) compassion only fosters third-party punishment when suffering was caused by high (vs. low) unjust acts (Study 5). Overall, the present research discusses compassion in a different light in that harmful consequences of compassion are considered. Implications are discussed from a perspective of basic research on compassion and third-party punishment as well as from a societal perspective.

Keywords: compassion, empathy, harm, injustice, third-party punishment

Remember the Paris attacks of November 2015, when several mass shootings and a suicide bombing took place in the heart of the city. Many were hurt and killed, and substantial suffering resulted. Strong reactions were elicited based on this suffering, ranging from compassion for the injured to retaliatory tendencies toward those who caused the harm. In this contribution, the main idea put forward is that compassion for suffering individuals can motivate harmful tendencies toward those responsible for the suffering. Specifically, we argue that when suffering is caused by profound unjust behavior, compassion fosters punishment toward the perpetrator(s).

The present work has two central aims. The first is to apply a differentiated perspective on compassion. Compassion has been conceptualized as a prosocial tendency par excellence, fostering helping behavior across a variety of social situations (Dalai Lama & Ekman, 2008; Goetz, Keltner, & Simon-Thomas, 2010). In this contribution, it is argued and shown that in situations where individuals do not have the option to help a victim, compassion can foster harmful consequences for other individuals (i.e., third-party punishment of suffering-causing individuals). As such, the present work reveals compassion in a different light. This perspective from basic research is complemented by a societal perspective. Specifically, the second aim is to understand and to document the interpersonal consequences that result from the omnipresent confrontation with suffering. Indeed, people today are frequently exposed to suffering, notably when terrorist attacks happen, when viewing media reports about the war in Syria, or when perpetrators violate fundamental moral standards and cause suffering (e.g., the New Year's Eve sexual assaults in Cologne, Germany in 2015). In this regard, we tested the idea of compassion fostering third-party punishment in a variety of contexts of high societal relevance (i.e., terrorist attacks, sexual assaults, rape of children, and war). In the following, the current state of research on compassion is outlined. Subsequently, the exploratory framework is presented about why compassion fosters third-party punishment.

The Essence of Compassion

The core theme of compassion refers to being moved by another's suffering and possessing concerned feelings that are elicited in response to this suffering (Haidt, 2003; Lazarus, 1991; for a compelling review on compassion, see Goetz et al., 2010). As a consequence of compassion, actions are elicited that aim to reduce the negative state of the suffering individual, especially when the suffering other deserves help and one has the resources to help (Goetz et al., 2010; Haidt, 2003; Ministero, Poulin, Buffone, & DeLury, 2018). In this regard, it is shown that compassion leads to prosocial actions such as helping and supporting children in need,

This article was published Online First April 4, 2019.

Stefan Pfattheicher, Department of Psychology and Behavioral Sciences, Aarhus University, and Department of Social Psychology, Ulm University; Claudia Sassenrath and Johannes Keller, Department of Social Psychology, Ulm University.

This research was supported by a grant from the Baden-Württemberg Foundation to Stefan Pfattheicher.

Correspondence concerning this article should be addressed to Stefan Pfattheicher, Department of Psychology and Behavioral Sciences, Aarhus University, 8000 Aarhus C, Denmark. E-mail: sp@psy.au.dk

aiding people who suffer from a catastrophe or a personal loss, assisting homeless or poor people, and being supportive toward a romantic partner (Ashar et al., 2016; Batson et al., 1989, 1997; Cialdini, Brown, Lewis, Luce, & Neuberg, 1997; Crocker, Canevello, & Lewis, 2017; Eisenberg, McCreath, & Ahn, 1988; Eisenberg et al., 1994).[1]

One main motivational tendency emerging from compassion is the goal of relieving pain from suffering others (Goetz et al., 2010). Building on the notion that relieving pain is one central element of morality (Haidt, 2003; Schein & Gray, 2018), compassion is considered a moral emotion (Cameron & Payne, 2012; Haidt, 2003; Nussbaum, 1996; Zaki, 2018). Beyond that, compassion entails a specific sensitivity regarding *negative* social information, for instance, regarding suffering caused by unjust behavior (Goetz et al., 2010; Keller & Pfattheicher, 2013). When compassionate individuals observe unjust suffering, they "moralize" the event and then possess an intensified moral concern in this regard (Bekkers & Ottoni-Wilhelm, 2016; Horberg, Oveis, & Keltner, 2011). As a consequence of their moral concern, compassionate individuals' resulting actions aim to reestablish justice and morality, for instance by means of helping a suffering victim (Goetz et al., 2010).

In the following, we outline the rationale about why compassion fosters third-party punishment. Third-party punishment in this work is conceptualized broadly as any behavior that includes negative sanctions and consequences or the support of negative sanctions and consequences targeted at a third party who has caused harm. These negative sanctions and consequences include a broad range of tendencies such as physical or psychological harm or inflicting financial costs for the third party. Examples of third-party punishment are support for retribution of terror attacks, support for the death penalty of rapists, and third-party punishment of unfair individuals in economic games. As such, the theoretical and empirical basis of the present work is not restricted to a single and specific manifestation of third-party punishment but rather makes claims regarding a greater range of third-party punishment tendencies.

Why Compassion Fosters Third-Party Punishment: Just Deserts Theory

We build on a prominent theory in the domain of criminology and morality, that is, *just deserts theory*, to argue why compassionate individuals are motivated to punish those who caused suffering. Just deserts theory deals with the basic question of why people punish and why they want violators of morality to be harmed. Its basic foundation goes back to Immanuel Kant (1785/1998), who argued that humans act as rational agents; consequently, individuals must consider the consequences of their actions, and therefore would deserve punishment in case their acts violated justice (cf. Barton, 2004; Darley, Carlsmith, & Robinson, 2000). The essence of just deserts theory is that harm targeted at an offender should be proportionate to the strength of the moral violation (Von Hirsch, 1976); its justification lies in the reestablishment of justice and in righting morally wrong past actions. Empirically, Carlsmith, Darley, and Robinson (2002) showed that individuals follow the just deserts approach when they observe a moral violation. They documented that judging an action as morally wrong led to moral outrage, that is, negative affect (anger, contempt) provoked by the perception of a moral violation (Batson et al., 2007; Hoffman, 1990; Montada & Schneider, 1989; Tetlock, 2002). Moral outrage, in turn, led to the inclination to harm the offender (Carlsmith et al., 2002; Eriksson, Strimling, Andersson, & Lindholm, 2017).

It is important to note that just deserts theory includes the idea that contextual factors pronounce or attenuate individuals' perception of a moral violation and consequently their moral outrage and punishment inclinations (Carlsmith et al., 2002; Robinson & Darley, 1995). This implies that the more an action is judged as morally wrong (because of certain contextual factors), the more moral outrage emerges, leading to stronger tendencies to punish the perpetrator (Fetchenhauer & Huang, 2004). Carlsmith and colleagues (2002) found empirical support for this idea. In a series of studies, they showed that higher seriousness of a morally wrong action elicits more moral outrage, and as a response, stronger motivation and support to punish the perpetrator (see also Fiske & Tetlock, 1997; Tetlock, Kristel, Elson, Green, & Lerner, 2000).

We argue that compassion reflects such a contextual factor that intensifies individuals' moral outrage and punishment inclinations. Compassion makes one sensitive toward the cared-for (Bekkers & Ottoni-Wilhelm, 2016; Goetz et al., 2010; Haidt, 2003). In such a state, unjust harm toward the cared-for is then perceived as more profound, intensifying moral outrage (see Thulin & Bicchieri, 2016, and Rothschild & Keefer, 2018, for initial empirical evidence). Moral outrage, in turn, should lead to the tendency to punish the perpetrator to reestablish justice. In other words, because one cares for and feels compassion for a suffering person, unjust harm done to this person is perceived as more morally profound, resulting in intensified moral outrage and magnified third-party punishment. We therefore expect that compassion for individuals suffering from unjust behavior magnifies third-party punishment because compassion intensifies moral outrage, which in turn increases punishment inclination toward those who caused unjust suffering. To summarize, we argue that compassion reflects an intensifier of individuals' perception of a moral violation and the downstream consequences (moral outrage and third-party punishment); therefore, we expect that, given that suffering is caused by unjust behavior, compassion intensifies moral outrage about unjust suffering, which in turn increases third-party punishment toward those responsible for the suffering.

In the studies of the present contribution, we first tested whether compassion is positively correlated to third-party punishment across a variety of situations in which suffering was caused by unjust behavior (Studies 1a–1f). In Studies 2 through 6 we applied an experimental approach and tested whether compassion increases third-party punishment. We further showed that (a) this increase occurs because compassion intensifies moral outrage, which in turn predicted third-party punishment (Studies 2–6), and

[1] It is noteworthy that other researchers use "empathic concern" instead of compassion to describe concerned feelings for suffering others. In fact, empathic concern is typically used interchangeably with compassion (cf. Batson, 2009; Condon & DeSteno, 2017; Goetz et al., 2010; Singer & Klimecki, 2014; Singer & Steinbeis, 2009; Stellar & Keltner, 2014; Stellar et al., 2015). As the term 'empathy' creates misunderstanding (Batson, 2009; DeSteno, 2015), the term compassion is used to make explicit that the present research focuses on concerned feelings for *suffering* others, not for others in general.

(b) compassion fosters third-party punishment only when the action that caused the suffering was unjust, not when the action was not unjust (Study 5).

The Value of the Present Research

The present work makes important contributions to the literature, both from the perspective of research on compassion as well as from the perspective of research on third-party punishment: First, the present contribution shapes our basic understanding of compassion in that harmful consequences of compassion are considered. So far, research on compassion has focused clearly on its ability to reduce suffering (Goetz et al., 2010). By showing that compassion actually motivates tendencies that entail harmful consequences for other individuals (i.e., third-party punishment), we point to a so far overlooked consequence of compassion.

In this regard, the present research is in line with previous work from Buffone and Poulin (2014) and Keller and Pfattheicher (2013) in showing negative interpersonal tendencies resulting from compassion. Buffone and Poulin (2014) document that compassion for a distressed person increases harming behavior toward a third individual who faced an upcoming competition with the distressed person. In this way, compassionate individuals helped the suffering (distressed) other by means of handicapping the nonsuffering individual. Thus, compassion leads to harmful behavior that *directly* helps a suffering person. Hence, the work by Buffone and Poulin (2014) fits well in the overall picture that the main tendency that results from compassion is to help a suffering other. The present work goes one crucial step further, arguing and showing that in situations in which it is difficult for compassionate individuals to help a victim, they are motivated to harm the suffering-causing individual. As such, we show that compassion is not only about helping a victim. It is more than that; it is also about punishing a perpetrator for causing harm to a person someone cares for. Of course, punishment of a perpetrator can carry an indirect helping function; past research has shown that (third-party) punishment of unfair individuals reduces future unfair behavior to the same and other victims (Fehr & Fischbacher, 2004; Rand & Nowak, 2013). Our main point here is that compassion not only fosters actions that improve the state of a suffering victim; it can also foster actions that impair another individual's state.

Further relating the present contribution to other relevant work, Keller and Pfattheicher (2013) showed that compassion for a suffering individual who was harmed by others leads to a hostile worldview that humans are bad in general given that individuals were prevention-focused. As an extension of Keller and Pfattheicher (2013), the present work shows that compassion can foster negative consequences that really affect other individuals, not only by increasing a general negative belief about humans.

We can extend existing literature on compassion in an additional meaningful way. So far, research on compassion has focused mainly on compassion emerging from suffering that is *not* caused by third parties (cf. Goetz et al., 2010). We document that harmful tendencies (i.e., punishment) result from compassion when suffering is caused by third parties, given that the suffering-causing actions are unjust. In the few existing studies that include consideration of suffering caused by third parties, it is argued that anger emerges when seeing that a person for whom one cares is undeservedly harmed (termed "empathic anger"; Batson et al., 2007; Hoffman, 2000; Vitaglione & Barnett, 2003). This anger fosters the motivation to punish the one who is responsible for the observed suffering (Batson, 2009). This perspective fits the line of argument of the present work. We argue that being in a compassionate state (what Batson et al., 2007, call "care about the welfare of a suffering other") intensifies the perception of injustice (i.e., intensified moral outrage). We suggest that moral outrage and empathic anger on the basis of undeserved (i.e., unjust) harm are closely related, as moral outrage includes anger (besides contempt; Tetlock, 2002; Tetlock et al., 2000). In this regard, we aim to extend research on empathic anger as our research suggests why empathic anger emerges (because of compassion for someone who suffers from injustice).

The present work also advances understanding of the emergence of third-party punishment. Past research shows that the moral emotion of anger drives third-party punishment of individuals that intentionally behaved unfairly toward an interaction partner (Nelissen & Zeelenberg, 2009; see also Jordan, McAuliffe, & Rand, 2016). Approaches that explain the emergence of third-party punishment assume that unfair behavior reflects the violation of a fairness norm; this perception of a norm violation elicits anger which in turn motivates punishment of the unfair third party (Falk, Fehr, & Fischbacher, 2003; Fehr & Fischbacher, 2004; Singer & Steinbeis, 2009). We rely on this research, arguing that moral outrage (which includes anger about unfair behavior) motivates third-party punishment. Moreover, we provide an *additional* piece to the puzzle on how third-party punishment emerges. Here, we build on just deserts theory (as outlined in detail above). Just deserts theory includes the idea that contextual factors pronounce or attenuate individuals' perception of a moral violation and consequently their moral outrage and punishment inclinations (Carlsmith et al., 2002; Robinson & Darley, 1995). We argue that compassion reflects such a contextual factor in that it intensifies the perception of unfairness (i.e., intensifies moral outrage), leading to increased third-party punishment. In sum, with the empirical research program reported in the following, we aim to increase knowledge and understanding about so far overlooked consequences of compassion and to point to a crucial factor in the emergence of third-party punishment.

Studies 1a–1f

The aim of the first set of studies was to put the idea of compassion-based third-party punishment to a correlational test. Moreover, we aimed at documenting a positive relation of compassion and third-party punishment in many different situations of high societal relevance. To do so, we assessed compassion for suffering individuals from various terrorist attacks (Paris 2015; Istanbul 2015; Brussels 2016) in Studies 1a–1c; compassion for suffering individuals from the war in Syria in Study 1d; compassion for women suffering from sexual assaults perpetrated by groups of young foreign men on New Year's Eve, 2015 in Cologne, Germany, in Study 1e; and compassion for three children suffering from rape by a man in Study 1f. In each of the studies, we also assessed third-party punishment inclinations toward those responsible for the suffering, that is, how much participants agree to inflict harm on the suffering causing individuals (e.g., support of retributive actions for terror attacks and support for death penalty for the rapist).

Method

Research ethics statement. The studies reported in the present contribution were conducted in full accordance with the Ethical Guidelines of the German Association of Psychologists (DGPs) and the American Psychological Association (APA). Institutional review boards or committees are not mandatory at Danish and German universities. Data, all items, and dropout rates of all studies presented in this contribution are available on the Open Science Framework (see https://osf.io/fqnjm/).

Procedure. Each study was conducted online. Studies 1a–1d were conducted via Amazon Mechanical Turk (cf. Buhrmester, Kwang, & Gosling, 2011); Study 1e was distributed via a mailing list of German individuals who are regularly informed about new psychological online studies; and Study 1f was conducted via an online platform providing access to German individuals who are willing to participate in studies in exchange for a small amount of money. All studies followed the same procedure. First, participants read an introductory statement that reminded them of an event that included suffering. For instance, the introductory statement in the context of the Paris attacks read, "The following statements refer to the recent terrorist attacks in Paris, France."

After the introductory sentence(s), participants completed three items assessing compassion toward the suffering individuals (the items are reported below). Then they responded to one filler item (e.g., "How I dress is important to me."). The filler item was not significantly related to either compassion or third-party punishment in any of the studies (with a single exception of minor empirical relevance in Study 1f). After the filler item, participants completed an attention check item (i.e., "This is an attention check item. Please answer: I completely agree."). Finally, they responded to three items assessing third-party punishment toward those responsible for the suffering. Cronbach's αs, mean values, and standard deviations of all measures are provided in Table 1.

Compassion. To assess compassion, three items were formulated based on the empathic concern scale by Davis (1983). The compassion items in the context of the Paris attacks read, "I feel compassion for the people suffering from the attacks in Paris"; "I am quite touched by the people suffering from the attacks in Paris"; and "I have concerned feelings for the people suffering from the attacks in Paris." Responses were given on a 7-point Likert scale ranging from *strongly disagree* to *strongly agree* in Studies 1a–1e and on a 5-point Likert scale in Study 1f. The items assessing compassion in the other contexts (terrorist attacks in Instanbul and Brussels, the war in Syria, the sexual assaults in Cologne, and sexual assaults of children) were formulated exactly as the items of the Paris attacks study, but the context differed (e.g., "I feel compassion for people suffering in Syria" or "I feel compassion for the children.").

Third-party punishment. To assess third-party punishment, three items were formulated in each study. The items in the context of the Paris attacks read, "I support attacks on those who caused the suffering of the attacks in Paris"; "Those responsible for the attacks in Paris should be attacked"; and "If I could, I would attack those who caused the suffering of the attacks in Paris." The items assessing third-party punishment in the other contexts were formulated exactly as the items of the Paris attacks study, adapted to the respective context (e.g., "If I could decide I would attack those who cause suffering in Syria," or "The rapist deserves the death penalty").

Participants. Using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), a power analysis was conducted for a two-tailed test to detect at least a small to medium effect ($r = .25$); power was set to .80 (Cohen, 1992). This power analysis revealed a required sample size of $N = 120$ to detect a significant effect (alpha level of .05) given there is a true effect. However, because we did not know the "true effect," we oversampled the studies when possible (in terms of financial opportunities and accessibility to participants). Only in Study 1e, in which sample size ($N = 72$) was determined by a restricted pool of participants, did we fall below the threshold set by the power analysis. See Table 1 for how many participants completed each study as well as for means and standard deviations regarding age and percentage of sexes. Sample sizes in Table 1 refer to how many participants completed all items and responded correctly to the attention check item.

Results and Discussion

An overview concerning the assessed variables is given in Table 1. It should be noted that the observed means in reported compassion are relatively high, suggesting that the applied contexts in which suffering occurred elicited compassion to a substantial extent. Yet, as indicated by the relatively large standard deviations, some individuals in the studies did not report strong compassionate tendencies. Regarding third-party punishment, means differed; they were higher when terrorist attacks happened compared with when persons suffered from sexual assaults. The large standard

Table 1
Overview Concerning the Characteristics and Variables of Studies 1a–1f as well as Correlations Between Compassion and Third-Party Punishment

		Sample size	Age	Sex	Compassion		Third-party punishment		Correlation	
Study	Context	*N*	Mean (*SD*)	% women	α	Mean (*SD*)	α	Mean (*SD*)	*r*	*p* value
Study 1a	Paris attacks	188	35.6 (10.9)	47.3%	.73	6.31 (0.95)	.90	4.99 (1.78)	.22	.002
Study 1b	Istanbul attacks	203	36.1 (12.2)	51.2%	.86	6.07 (1.12)	.87	4.73 (1.84)	.23	.001
Study 1c	Brussels attacks	238	36.4 (12.1)	48.7%	.90	6.13 (1.16)	.92	4.65 (1.94)	.24	.001
Study 1d	War in Syria	170	36.1 (12.0)	57.1%	.94	5.40 (1.70)	.91	4.61 (1.54)	.20	.010
Study 1e	Sexual assaults Cologne	72	27.4 (11.1)	72.9%	.90	5.13 (1.35)	.87	2.03 (1.37)	.34	.003
Study 1f	Rape of children	198	37.2 (11.6)	39.9%	.81	4.77 (0.44)	.97	2.96 (1.56)	.18	.014

Note. Responses were given on a 7-point Likert scale in Studies 1a–1e and on a 5-point Likert scale in Study 1f.

deviations in third-party punishment indicate that participants differ to a large degree in their punishment inclinations.

Most important for the present contribution are the correlations between compassion and third-party punishment. In all six studies, across a variety of contexts of high societal relevance, a positive relation was found, indicating that higher compassion is accompanied by higher third-party punishment inclinations toward those who caused the suffering. All relations remained robust in terms of effect size and significance levels when controlling for sex and age. The lowest correlation was found in the context of the death penalty for a man who raped three children, potentially due to the low variance on the compassion measure. The highest correlation was found in the smallest sample; accordingly, the confidence interval of this correlation is larger than in the other samples. Overall, an internal meta-analysis including all samples, using fixed effects in which the mean effect size was weighted by sample size (cf. Goh, Hall, & Rosenthal, 2016), revealed a positive relation of small to medium size between compassion and third-party punishment ($M\ r = .22$, $SE = 0.03$, $Z = 7.39$, $p < .001$, two-tailed).

We acknowledge that these studies were of correlational nature, thus we do not know whether compassion fosters third-party punishment (in a causal sense). In addition, in these studies, we have not tested whether other emotions (e.g., anger) might be responsible for the observed correlations. Yet if we had included other emotions, the studies would still be of correlational nature. That is, if we really want to show that compassion is driving the effects instead of other emotions, we need experimental studies in which only compassion is manipulated. This was done in the experimental studies that follow (Studies 2 to 6).

Studies 2a and 2b

Method

Procedure. Study 2a was conducted in the laboratory at a German University, and Study 2b was conducted via Amazon Mechanical Turk (cf. Buhrmester et al., 2011). In both studies, compassion was manipulated first, then a manipulation check was implemented to assess state compassion and other emotions. *After that*, participants learned that suffering was caused by unjust behavior. This aspect is important, because in this way we manipulated compassion prior to imparting (and thus independent from) the information about unjust behavior and injustice-related emotions. Finally, individuals in both studies report on their third-party punishment inclinations.

Compassion manipulation. Participants read a short text about George (70 years of age) who reports:

> Last week I broke my shoulder. As you can imagine, at my age you don't recover quickly, it takes longer than usual. Every time I move my body it hurts . . . although I frequently take aspirins. I do go for short walks though. But walking is difficult. After that I'm so exhausted that I have to lie down.

Before reading the text, a standard procedure was used to manipulate compassion (Batson et al., 1997; Batson, Chang, Orr, & Rowland, 2002). Participants were randomly assigned either to a high- or a low-compassion condition. In the high-compassion condition, participants read these compassion-promoting instructions for how they should approach the text before reading the text (cf. Batson et al., 1997, Study 3; Batson et al., 2002): "While reading, imagine how the described person feels about what has happened and how it has affected his life. Try to feel the full impact of what he has been through and how he feels as a result." In the low-compassion condition, participants read, "While reading, take an objective perspective toward what is described. Try not to get caught up in how the described person feels; just remain objective and detached." Thus, we had a low versus high-compassion condition in the studies.

Manipulation check. State compassion (Study 2a $\alpha = .83$, Study 2b $\alpha = .91$) was assessed with the items "moved," "sympathetic," "concerned," and "compassionate" (cf. Batson et al., 1997, 2002). Please note that we refer to this index as "state compassion" in the following text. In addition, the six basic emotions of anger, sadness, fear, surprise, happiness, and disgust (Ekman, 1992) were assessed along with the moral emotion of contempt (Haidt, 2003), as it has been argued that moral outrage is reflected by the combination of anger and contempt (Tetlock, 2002; Tetlock et al., 2000).[2] Responses were given on a 7-point Likert scale ranging from *not at all* to *extremely*. Anger and contempt were highly correlated (Study 2a $r = .49$, $p < .001$, Study 2b $r = .61$, $p < .001$) and averaged to compute an index of state outrage. We refer to this index as "state outrage" in the following text.

Third-party punishment. Before individuals reported on their punishment inclinations, they learned that George was intentionally harmed, as *intentional* harm is one main elicitor of injustice perception (e.g., Mikula, 1993). Specifically, participants read: "George continues: Maybe you are interested in how I broke my shoulder. It was an accident. A young man deliberately bumped into me. Then I fell. This is how it happened." After reading this text, third-party punishment inclinations toward the young man were assessed. The items (Study 2a $\alpha = .73$, Study 2b $\alpha = .93$) read, "The young man deserves hard treatment"; "I want to teach the young man a lesson"; and "I want the young man to be severely punished."

Moral outrage. We have argued that compassion for unjust suffering motivates third-party punishment because compassion intensifies moral outrage, which in turn fosters third-party punishment. In Study 2b (but not in Study 2a) we assessed moral outrage right before punishment. The items ($\alpha = .90$) read, "I am angry at the young man"; "I am outraged by the young man's behavior"; and "The behavior of the young man is absolutely immoral." Please note that we refer to this index as "moral outrage" (as opposed to the index of "state outrage" described above) in the following text.

Participants. Using G*Power (Faul et al., 2009), a power analysis was conducted for a two-tailed test to detect at least a medium effect ($d = .50$); power was set to .80 (Cohen, 1992). This power analysis revealed a required sample size of $N = 128$ to detect a significant effect (alpha level of .05) given there is a true

[2] We acknowledge that there is disagreement about whether or not to include disgust in a measure of moral outrage (for a discussion see Gervais & Fessler, 2017). Our *a priori* approach was to follow Tetlock and colleagues (2000; Tetlock, 2002), arguing that moral outrage is reflected by the combination of anger and contempt. We note, however, that the results do not change when disgust is included in the moral outrage index.

effect. In Study 2a, we obtained 105 German participants (53.3% women; M_{age} = 21.0, SD_{age} = 3.4). The set threshold determined by power analysis was not reached because in the laboratory the time the study could be conducted was restricted. In Study 2b, 171 U.S.-American individuals completed the study (64.3% women; M_{age} = 39.1, SD_{age} = 13.4). Dropouts were not significantly uneven by condition in any of the experimental studies reported in the present contribution (cf. Zhou & Fishbach, 2016).

Results and Discussion

Manipulation check. First, we examined in Study 2a whether the manipulation increased (only) state compassion. This was the case. Analysis revealed a significant difference, $t(103) = 3.05$, $p < .01$; Cohen's $d = 0.59$ in state compassion between the low-compassion condition ($M = 3.74$, $SD = 1.39$) and the high-compassion condition ($M = 4.52$, $SD = 1.24$). Next, there was no significant difference in state outrage, $t(103) = 0.34$, $p = .74$; Cohen's $d = 0.07$ between the low-compassion condition ($M = 1.59$, $SD = 0.99$) and the high-compassion condition ($M = 1.52$, $SD = 1.05$). The low means in state outrage suggest that state outrage did not play a role in this manipulation. The two conditions also did not differ regarding the other assessed emotions (all ts $< |1|$), except for sadness which was (nonsignificantly) increased in the high-compassion condition, $t(103) = 1.46$, $p = .15$; Cohen's $d = 0.28$. The same pattern in terms of significance levels, effect sizes, and direction of effects could be found in Study 2b. State compassion was significantly increased in the high-compassion condition, $t(169) = 3.22$, $p < .01$; Cohen's $d = 0.50$ but state outrage was not, $t(169) = 1.49$, $p = .14$; Cohen's $d = 0.23$. These results, in both studies, indicate a successful and specific induction of compassion in both studies, with sadness also increased. We note that one can argue that sadness must be included in the state compassion index (e.g., Batson et al., 1989; Sassenrath, Pfattheicher, & Keller, 2017).[3]

Third-party punishment. In Study 2a, analysis revealed a significant difference, $t(103) = 2.10$, $p = .04$; Cohen's $d = 0.41$, in third-party punishment between the low-compassion condition ($M = 3.66$, $SD = 1.48$) and the high-compassion condition ($M = 4.24$, $SD = 1.35$). This finding was replicated in Study 2b; analysis revealed a significant difference, $t(169) = 2.74$, $p < .01$; Cohen's $d = 0.40$ in third-party punishment between the low-compassion condition ($M = 3.24$, $SD = 2.02$) and the high-compassion condition ($M = 4.02$, $SD = 1.68$).

Mediation analysis in Study 2b. For testing mediation, we used the PROCESS macro provided by Hayes (2013) and ran Model 4. First, moral outrage (which was assessed *after* participants learned about unjust suffering) was regressed on the experimental conditions. This analysis revealed higher moral outrage in the high-versus the low-compassion condition ($B = 1.04$, $SE = 0.28$, $t = 3.69$, $p < .001$), which was in turn positively related to punishment ($B = 0.84$, $SE = 0.04$, $t = 19.73$, $p < .001$) when the experimental conditions were included in the regression model ($B = -0.09$, $SE = 0.16$, $t = -.59$, $p = .55$). The indirect effect was also significant, that is, the bootstrapped 95% confidence interval excluded zero [0.41; 1.34] (Hayes, 2013). In other words, moral outrage mediated the effect of experimentally induced compassion on third-party punishment in that compassion increased moral outrage, which in turn predicted third-party punishment.

Additional analyses. In additional analyses, we predicted third-party punishment with state compassion and state outrage (to recapitulate, these were assessed right after the manipulation of compassion but *before* participants learned about unjust suffering) together in a simultaneous regression model. In this way, we further tested whether state compassion is driving the effect of the applied manipulation on third-party punishment and not *directly* induced state outrage. Multiple regression analyses revealed that state compassion was significantly predictive of third-party punishment (Study 2a: $B = 0.37$, $SE = 0.10$, $t = 3.78$, $p < .001$; Study 2b: $B = 0.40$, $SE = 0.09$, $t = 4.60$, $p < .001$) and relevant in terms of effect size (Study 2a $\Delta R^2 = .13$; Study 2b $\Delta R^2 = .12$). In contrast, state outrage assessed prior to the intentional harm information, expectedly, was not significantly related in Study 2a ($B = 0.04$, $SE = 0.13$, $t = 0.33$, $p = .74$) although statistically significant in Study 2b ($B = 0.26$, $SE = 0.13$, $t = 2.10$, $p = .04$). However, in both studies, the effect size of state outrage was negligible (Study 2a $\Delta R^2 = .00$; Study 2b $\Delta R^2 = .02$), and the significant effect of Study 2b was not replicated in the following studies.

Discussion. In Studies 2a and 2b, we induced compassion without any reference to injustice or a moral violation by a third party. It is shown that the manipulation induced compassion and not moral outrage. We then showed that compassion intensified moral outrage in response to the subsequently presented information on intentional harm, which in turn predicted third-party punishment. Overall, Studies 2a and 2b provide first evidence for the idea that compassion increases third-party punishment in an online study and a study conducted in the laboratory, using participants from the U.S. and Germany.

Additional Studies

We ran additional studies which are, because of redundancy and methodological shortcomings, not reported in detail in this paper; detailed information as well as data can be found on the OSF (see https://osf.io/fqnjm). This package includes one online study in which we induced compassion for a homeless person who suffered from unjust action by her husband and a lab study in which we induced compassion for a 12-year old student who was beaten up by an older schoolmate. We found significantly increased third-party punishment inclinations in the high- compared with the low-compassion condition. However, the two studies suffer from a main shortcoming. Specifically, we have manipulated compassion for another individual who suffered from injustice by a third person; this situation confounded compassion (elicited by suffering) and moral outrage (elicited by injustice). As such, we could have directly induced moral outrage in addition to compassion. The shortcomings are not present in Studies 2a and 2b reported above and Studies 4 to 5 reported below; in these studies, we induce compassion without any reference to injustice or a moral violation by a third party.

[3] We note that there were no problems in the present set of studies with multicollinearity in terms of the assessed emotions when looking at the correlations and the Variance Inflation Factor (VIF), expect for sadness, which can arguably be included in the state compassion index (Batson, 2011; Olderbak et al., 2014; Sassenrath et al., 2017).

We also note that we ran an additional study in which we first manipulated compassion for a woman suffering from AIDS. Thereafter, participants learned that AIDS was unjustly caused by a man (the scenario was adapted from Batson et al., 1997). In this study, we did not find a significant effect of the compassion manipulation on third-party punishment of the man, $t(139) = 1.27$, $p = .21$, $d = .22$. There can be many reasons for a nonsignificant effect (e.g., Simonsohn, 2015): in the present case, it is possible that the scenario was not realistic, that (a proportion of) participants blamed the victim (e.g., Lerner, 1980) rather than felt compassion, or that there is no true effect. We included this study in the internal meta-analysis reported below.

In the next studies (Studies 3a–3c), we aimed (a) to conceptually replicate the basic effect of compassion on third-party punishment, (b) to replicate the found mediation in Study 2b, and (c) to apply a paradigm to obtain a *behavioral* measure of third-party punishment. To do so, we used the classic third-party punishment paradigm (Fehr & Fischbacher, 2004).

Studies 3a–3c

Method

Procedure. Studies 3a to 3c were conducted online via Amazon Mechanical Turk (cf. Buhrmester et al., 2011). Participants first read the instructions of the third-party punishment paradigm (Fehr & Fischbacher, 2004). Compassion was then manipulated. In Studies 3a and 3b we measured state compassion after the manipulation (as a manipulation check); in Study 3c, no manipulation check was used to get a dependent variable unbiased by the measurement of the manipulation check. After that, participants had the opportunity to engage in third-party punishment. In Study 3c, we further had a measure of moral outrage (the mediator) after the dependent variable; moral outrage was measured after the dependent variable to get an unbiased and valid measure of third-party punishment (Spencer, Zanna, & Fong, 2005). Please note that in Studies 2, 5, and 6, the mediator was measured in-between the independent and the dependent variable.

Third-party punishment paradigm. Participants were first told that two other participants (called Player A and Player B) have worked together with equal effort to fulfill a task. In this task, they had together earned \$10. Next, participants learned that Player A was given full control of the distribution of these \$10. They were told that Person A took \$7 while giving Person B \$3, thus harming Person B financially. Player B had no choice but to accept the decision of Player A. As such, participants observed an unjust distribution of money (Eriksson et al., 2017; Nelissen & Zeelenberg, 2009).[4]

To assess punishment behavior, participants were given \$10 as endowment and provided with the possibility to punish Person A by reducing Person A's income out of their own endowment of \$10. Specifically, each dollar invested to punish Person A decreased Person A's income by \$1 (e.g., when a participant invested \$3, the participant's income was decreased by these \$3; the income of Person A also decreased by \$3). In this way, we received a behavioral measure of third-party punishment, ranging from 0 (when null dollars were invested) to 7 (when \$7 were invested).

At the beginning of the instructions, participants were informed that one of 20 participants in the game will be paid out, as regularly done in the field of experimental economics (i.e., random-lottery incentive scheme; see, e.g., Bardsley et al., 2010). This makes it possible to play the game with high and relevant stakes while assuring that participants' choice had, to the given probability, real financial consequences for participants. Payment of one of 20 participants was made after data collection was completed via the Amazon Mechanical Turk system.

Compassion. After participants learned about the unfair distribution of money, but before participants were allowed to punish, compassion was manipulated in line with Study 2, adapted to the present study's context (see also Batson et al., 1997; Batson et al., 2002). Participants were either randomly assigned to a high- or a low-compassion condition; specifically, participants in the high-compassion condition read, "Before you can make your decision, imagine how Player B feels about the decision by Player A and how it has affected Player B. Try to feel the full impact of what the decision could mean to Player B and how s/he feels as a result." Participants in the low-compassion condition read, "Before you make your decision, please take an objective perspective toward the decision by Player A. Try not to get caught up in how Player B feels; please remain objective and detached." Then participants were asked to write down their thoughts and feelings. After this instruction, participants decided how much to reduce Player A's income.

Manipulation check. After the manipulation of compassion in Studies 3a and 3b, but before participants could punish, they respond to three items assessing their current compassionate feelings for Player B. Participants were asked to indicate, on a 5-point Likert scale, the degree to which they are currently feeling each emotional state toward Player B. The items read (Study 3a: $\alpha = .80$; Study 3b: $\alpha = .82$), "compassionate," "moved," "touched."[5]

Moral outrage. In Study 3c, moral outrage was measured (after third-party punishment). The items used in Study 2b were adapted to the present study's context. The items read, "I think that the decision by Player A was immoral"; "I think the split by Player A was immoral"; and "I think that Player A's behavior was immoral." Responses were given on a 6-point Likert scale ranging from *strongly disagree* to *strongly agree*. In addition, we assessed individuals' emotional response toward the (unjust) distribution decision of Person A; specifically, individuals had to indicate their emotional state when they thought about the distribution decision of Person A. We assessed the six basic emotions of anger, sadness, fear, surprise, happiness, and disgust (Ekman, 1992) along with the moral emotion of contempt (Haidt, 2003). Responses were given on a 5-point Likert scale ranging from *not at all* to *extremely*. Anger and contempt were highly correlated, $r = .48$, $p < .001$ and averaged to compute an index of anger and contempt. This index and the first three items were highly correlated, $r = .57$, $p < .001$,

[4] Actually, Player A and Player B were fictitious. Participants were debriefed in this regard at the end of the study.

[5] In these studies, we have not assessed other emotions. Please note that in Study 4, we used the same game theoretical paradigm as in Studies 3a–3c (the third-party punishment paradigm); in Study 4, we found that other emotions (except for sadness) were not affected.

z-standardized because of the different Likert scale ranges,[6] and averaged to compute an index of moral outrage. The results reported below were exactly the same when we used structural equation modeling (SEM) and modeled a latent variable affecting the first three items and the index of anger and contempt, and then used this latent variable as the mediator.

Participants. Power analysis was equivalent to Study 2, so we aimed to obtain at least $N = 128$ participants. We sampled $N = 133$ participants in Study 3a (41.4% women; $M_{age} = 33.7$, $SD_{age} = 11.3$), $N = 126$ participants in Study 3b (50.0% women; $M_{age} = 34.3$, $SD_{age} = 10.2$) who completed all measures, and, to adequately power path analysis, $N = 229$ participants in Study 3c (43.7% women; $M_{age} = 33.9$, $SD_{age} = 10.8$).

Results and Discussion

Manipulation check. Analysis in Study 3a of state compassion toward Player B revealed higher state compassion in the high-compassion condition ($M = 2.77$, $SD = 0.88$) compared with the low-compassion condition ($M = 2.41$, $SD = 1.04$). This difference was significant, $t(131) = 2.15$, $p = .03$; Cohen's $d = 0.37$. The same pattern emerged in Study 3b: State compassion was higher in the high-compassion condition ($M = 2.73$, $SD = 1.10$) compared with the low-compassion condition ($M = 2.22$, $SD = 1.01$). This difference was significant, $t(124) = 2.74$, $p < .01$; Cohen's $d = 0.40$. State compassion was further related to third-party punishment (Study 3a $r = .33$, $p < .001$; Study 3b $r = .43$, $p < .001$).

Main results. Next, it was tested whether the experimental manipulation of compassion increased third-party punishment. This was not the case in Study 3a, $t(131) = 0.03$, $p = .97$; Cohen's $d = 0.01$. In contrast, analyses revealed a significant difference in Study 3b, $t(124) = 2.95$, $p < .01$; Cohen's $d = 0.47$ between the low-compassion condition ($M = 2.16$, $SD = 1.46$) and the high-compassion condition ($M = 3.06$, $SD = 1.93$), as well as in Study 3c, $t(227) = 2.56$, $p = .01$; Cohen's $d = 0.43$ between the low-compassion condition ($M = 1.84$, $SD = 1.45$) and the high-compassion condition ($M = 2.39$, $SD = 1.78$). Given the inconsistent results across the three studies, we ran an internal meta-analysis using fixed effects in which the mean effect size was weighted by sample size (cf. Goh et al., 2016). The effect was significant, $M\ d = .29$, $Z = 3.22$, $p < .01$, two-tailed, such that overall, compassion increased third-party punishment. Here, a small-to-medium effect was found.

Mediation analysis. Next, mediation analysis was executed in Study 3c (Hayes, 2013). Analysis revealed that moral outrage was higher in the high- compared to the low-compassion condition ($B = 0.39$, $SE = 0.11$, $t = 3.38$, $p < .001$); this in turn was positively related to third-party punishment ($B = 0.87$, $SE = 0.11$, $t = 8.02$, $p < .001$) when the experimental condition was controlled for ($B = 0.21$, $SE = 0.19$, $t = 1.08$, $p = .28$). This indirect effect was significant; the bootstrapped 95% confidence interval (based on 5000 resamples) excluded zero [0.15; 0.59].

We also tested, controlling for moral outrage, whether the other assessed emotional reactions toward the distribution decisions (i.e., sadness, fear, surprise, happiness, and disgust) functioned as mediators. This was not the case: In fact, none of the indirect effects turned out to be significant.

In sum, moral outrage mediated the effect of compassion on third-party punishment, thus replicating Study 2b. Moreover, we applied a causal analytic test to the relation of compassion and third-party punishment, including punishment *behavior*, in our analyses. However, we want to acknowledge a shortcoming of Studies 3a–3c: As we have manipulated compassion for another individual who suffered from injustice, we could have *directly* induced moral outrage in addition to compassion. Therefore, in the next study, we manipulated compassion first and did so independently from the injustice information.

Study 4

Method

Procedure. The overall procedure followed Studies 3a–3c. Study 4 was conducted in the laboratory at a German University. Participants first read the instructions of the classic third-party punishment paradigm. Compassion for Player B was then manipulated without giving any information about the unfair distribution of money. Only *after* having successfully induced compassion (which is confirmed by an independent manipulation check study, see below) did participants learn that the individual for whom they feel compassion (Player B) received only $3 from Player A. That is, in this study, we manipulated compassion independently from injustice. Finally, participants could engage in third-party punishment toward Player A.

Compassion. We adapted the procedure of Batson, Batson, et al. (1995) and told our participants that participants in the role of Player A and Player B had to write something interesting that happened to them recently. All participants were then told that they will read what Player B has written. The text of Player B read (cf. Batson, Batson, et al., 1995):

> I'm supposed to write about something interesting that's happened to me lately. Well, I don't know if this will be interesting to anybody else, but the only thing I can think of is that since two days I'm not together with my boyfriend any more. We've been going together since our junior year in high school and it's been great. Now we split up and I've been real down. It's all I think about. My friends all tell me that I'll meet other guys and they say all I need is for something good to happen to cheer me up. I guess they're right, but so far that hasn't happened.

Before reading the text, as in the previous studies, participants read compassion promoting instructions for how they should approach the text (Batson, Batson, et al., 1995; Batson, Chang, Orr, & Rowland, 2002; Batson, Polycarpou, et al., 1997).

In an independent online manipulation check study to test this manipulation ($N = 127$), we measured state compassion ($\alpha = .89$) as in Study 3a and 3b using the items "moved," "sympathetic," "concerned," and "compassionate."[7] In addition, the six basic

[6] The different Likert scale ranges (5 versus 6) applied in this study were a mistake.

[7] The manipulation check was not included in the main study (Study 4), because we also wanted to report on a study where there is no measurement (manipulation check and mediator) in between the manipulation of the independent variable and the assessment of the dependent variable, because assessments itself can alter the state participants are in (e.g., Kühnen, 2010).

emotions along with contempt were measured. Analysis revealed that the manipulation increased state compassion, $t(125) = 3.71$, $p < .001$; Cohen's $d = 0.70$, but did not significantly affect state moral outrage, $t(125) = 0.48$, $p = .64$; Cohen's $d = 0.09$, or any other emotion (all $ts < |1.20|$), except for sadness, $t(125) = 2.98$, $p < .01$; Cohen's $d = 0.53$. Overall, these results indicate a successful induction of compassion with this procedure.

Third-party punishment. After the manipulation of compassion, participants were told that Person A took $7 while giving Person B $3. Participants could then punish Person B, following the exact procedure of Studies 3a–3c.

Participants. Power analysis was equivalent to the previous studies, so we aimed to obtain at least $N = 128$ participants. We sampled $N = 140$ participants (55.7% women; $M_{age} = 21.5$, $SD_{age} = 2.9$).

Results and Discussion

It was tested whether compassion increased third-party punishment. This was the case; analysis revealed a significant difference, $t(138) = 2.56$, $p = .01$; Cohen's $d = 0.43$ between the low-compassion condition ($M = 2.46$, $SD = 1.55$) and the high-compassion condition ($M = 3.17$, $SD = 1.75$).

Study 4 mainly addressed an alternative explanation of the findings of Studies 3a to 3c, specifically, that we could have *directly* induced moral outrage regarding the unfair distribution of money, in addition to compassion. The procedure used in Study 4 manipulated compassion independently from the unfair distribution of money; as such, moral outrage was not directly induced, as indicated by the independent manipulation check study. We can thus rule out the mentioned alternative explanation.

Study 5

In the studies reported so far, suffering was caused by unjust behavior of the harmdoer(s). We have argued that compassion intensifies the perception of unjust behavior and consequently moral outrage and punishment inclinations. This reasoning would be supported if compassion intensified third-party punishment only in case harm was caused by unjust behavior, but not in case harm was *not* caused by unjust behavior, because then compassion is unlikely to intensify injustice perception as there is no injustice. Consequently, in this case, compassion should not lead to moral outrage, and consequently should not foster third-party punishment. Therefore, there should be no effect of compassion on third-party punishment given harm was caused by low unjust actions (as the cascade of mediation is interrupted; for the general logic of this approach, see Jacoby & Sassenberg, 2011; Pirlott & MacKinnon, 2016; Spencer et al., 2005). Accordingly, in this study, we manipulated the level of injustice of the harm-causing action, orthogonally to the manipulation of compassion. We expected that compassion intensifies third-party punishment only when harm was caused by unjust (vs. low unjust) behavior (i.e., an interaction effect involving the manipulation of compassion and the manipulation of injustice).

Method

Procedure. Study 5 was conducted online via Amazon Mechanical Turk (cf. Buhrmester et al., 2011). Compassion was manipulated first; the level of injustice (low vs. high) was manipulated orthogonally thereafter. We then assessed moral outrage, finally, third-party punishment was measured. Overall, a 2 (low vs. high compassion) by 2 (low vs. high injustice) between-factorial design was applied. The applied scenario was identical to Studies 2a and 2b (70-year-old George suffered a broken shoulder). Responses in this study were given on a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*.

Compassion. Compassion for George was manipulated as in Studies 2a and 2b. This manipulation *specifically* increased compassion and no other emotional states, as indicated by the successful manipulation checks in Studies 2a and 2b.

Injustice. The level of injustice involved in the harm that occurred was manipulated *after* the compassion manipulation (on a separate page). Injustice was manipulated by varying whether or not there was intent to cause suffering. Past research shows convincingly that individuals consider harm causing action to be far more unjust when the action was done intentionally, versus when the very same action was done unintentionally (e.g., Cushman, 2008; Hechler & Kessler, 2018; Malle & Knobe, 1997). Specifically, participants were given additional information. In the low-injustice condition, participants read that "Georg continues: Maybe you are interested in how I broke my shoulder. It was an accident. A young man unintentionally bumped into me. Then I fell. It wasn't the man's fault. This is how it happened." In the high-injustice condition, participants read that "Georg continues: Maybe you are interested how I broke my shoulder. It was an accident. A young man deliberately bumped into me. Then I fell. This is how it happened." The manipulation of injustice was pretested in an online study ($N = 51$).[8] Results revealed that participants consider the behavior of Person A to be more unjust in the high-injustice condition ($M = 6.03$, $SD = 1.55$) than in the low-injustice condition ($M = 2.57$, $SD = 1.96$). This difference was significant: $t(49) = 6.98$, $p < .001$, $d = 1.96$.

Moral outrage. We assessed moral outrage right before punishment using the same items as in Study 2b (e.g., "I am angry at the young man"). Moral outrage was reliably measured; Cronbach's alpha was .96.

Third-party punishment. The same items as in Studies 2a and 2b were used (e.g., "I want the young man to be severely punished"). Third-party punishment was reliably measured; Cronbach's alpha was .95.

Participants. In this study, we wanted to obtain very high power ($P = .99$) to detect at least an effect of small to medium size ($f = .20$). Using G*Power (Faul et al., 2009), the power analysis revealed a required sample size of $N = 462$ to detect a significant interaction effect (alpha level of .05) given there is a true effect. Financial opportunities when the study was run made it possible to overpower the study and therefore to reduce the likelihood of Type II error, and to detect even small effects. We obtained complete data from 973 U.S.-American individuals (62.0% women; $M_{age} = 36.7$, $SD_{age} = 12.1$).

[8] Injustice was measured using three items ($\alpha = .99$). The items read, "The behavior of the young man was immoral," "The behavior of the young man was unjust," and "The behavior of the young man was morally wrong."

Results and Discussion

Main results. Means of the conditions are presented in Figure 1. Analyses of variance yielded a significant main effect of compassion, $F(1, 969) = 7.16$, $p < .01$, $\eta_p^2 = .01$, and a significant main effect of injustice, $F(1, 969) = 452.63$, $p < .001$, $\eta_p^2 = .32$. The main effects were qualified by a significant two-way interaction, $F(1, 969) = 9.32$, $p < .01$, $\eta_p^2 = .01$. Decomposing the interaction revealed that when injustice was high, participants in the high-compassion condition reported significantly more third-party punishment ($M = 2.83$, $SD = 1.23$) compared with the low-compassion condition ($M = 2.48$, $SD = 1.16$), $F(1, 969) = 16.37$, $p < .001$, $\eta_p^2 = .02$. In contrast, such an effect of compassion on third-party punishment was not present when injustice was low: Participants in the high-compassion condition did not report significantly more third-party punishment inclinations ($M = 1.32$, $SD = 0.64$) compared with the low-compassion condition ($M = 1.35$, $SD = 0.70$), $F(1, 969) = 0.07$, $p = .79$, $\eta_p^2 = .00$).

Mediation analysis. For testing the mediation, we used the PROCESS macro provided by Hayes (2013) and we ran Model 8. This model tests the assumptions that compassion intensifies moral outrage, which in turn predicts third-party punishment, but that this cascade only takes place in the high-injustice condition, and not in the low-injustice condition. Accordingly, we tested for a conditional indirect effect (conditioned by the injustice manipulation).

Because of the complexity of the following statistical models, we predominantly report the relevant findings of the model in plain text. Yet, all statistical parameters are provided on the OSF (https://osf.io/fqnjm). It was first tested whether the significant positive relation of compassion and third-party punishment given high injustice is explained by moral outrage. This was indeed the case: Compassion intensified moral outrage, which in turn was positively related to third-party punishment. The indirect effect was also significant (the 95% bootstrapped CI excluded zero [0.16; 0.50]). This was already shown in Study 2b (using the same scenario), so Study 5 directly replicates this finding. In contrast, moral outrage did not function as significant mediator in the low-injustice condition (the 95% bootstrapped CI included zero [−0.09; 0.11]).

Next, the full conditional indirect effect model was tested. Analyses revealed a significant interaction of compassion and injustice when moral outrage was set as the dependent variable. Specifically, compassion significantly intensified moral outrage for individuals in the high-injustice condition but not in the low-injustice condition. Moral outrage, in turn, predicted third-party punishment when compassion, injustice, as well as their interaction, were controlled for. Overall, there was a significant index for a conditional indirect effect (the 95% bootstrapped CI excluded zero [0.13; 0.52]).

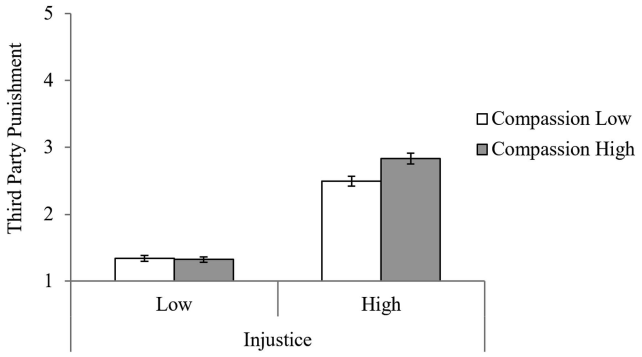

Figure 1. Third-party punishment tendencies as a function of injustice and compassion (Study 5).

Discussion. Overall, we found empirical support for our main assumption that compassion magnifies third-party punishment: The highly unjust action by the young man increased third-party punishment (vs. a low unjust action); this increase was magnified by compassion (see Figure 1). The latter increase could be explained by higher moral outrage in the high- compared to the low-compassion condition; moral outrage was, in turn, positively related to third-party punishment. These findings replicate Studies 2b and 3c. Overall, Study 5 showed that compassion *intensifies* moral outrage and thus serves as a magnifier of third-party punishment given high injustice. It should be noted that very low levels of third-party punishment emerged in the low injustice conditions. We argue that the low means had to emerge, because low (or no) injustice should lead to very low levels of third-party punishment inclinations. It is actually an interesting question for future research to examine at what level of injustice does compassion start to intensify moral outrage and third-party punishment.

Studies 6a and 6b

In all studies so far, we made use of a standard approach to manipulate compassion (i.e., the "Batson-manipulation"; e.g., Batson, 2009; Batson, Batson, et al., 1995; Batson, Polycarpou, et al., 1997). This manipulation includes the presentation of a sympathetic suffering target, and in the high-compassion condition participants are asked to imagine how the described person feels and are encouraged to feel for the suffering person; in the control condition, however, participants are asked to remain objective and detached. Recently, this manipulation was critically discussed: McAuliffe, Forster, Philippe, and McCullough (2018) showed that participants who received *no* instructions how they should approach a suffering target reported similar levels of compassion compared with a standard high-compassion condition (i.e., taking the perspective and feel for the suffering target), and higher levels of compassion than a standard low-compassion (i.e., instructions to stay objective and detached). McAuliffe and colleagues (2018) conclude that when confronted with a sympathetic suffering target, individuals respond spontaneously with compassion, and that encouraging to feel for the suffering target (using the "Batson-manipulation") adds only little additional compassion. Moreover, one can argue that by encouraging individuals to remain objective and detached, the low-compassion condition involves downregulation of compassion (and emotions in general); thus an effect might not be driven by compassion but instead by downregulation of emotions. To address these issues, in the next two studies, we manipulated compassion in a different way and used different control conditions.

Method

Procedure. The procedure and material followed Study 5. Only the manipulation of compassion differed, and only the high-injustice condition was applied in these studies. We had a com-

passion condition and a (two) control condition(s) in Study 6a (6b). In both studies, we first manipulated compassion and assessed the manipulation check. We then provided information that George suffered from unjust behavior by a young man, assessed moral outrage, and finally third-party punishment inclinations toward the young man who bumped into the old man (George).

As an additional measure, we exploratively assessed identification with the suffering target (using the item "I identified with the old man") as an alternative/additional explanation why compassion should foster third-party punishment. The idea is that compassion for a suffering target might increase identification with the harmed (Cassell, 2009), which in turn drives punishment. Because we did not find any significant effects involving identification (e.g., identification was not increased in the compassion conditions), we did not consider identification when further reporting the results.

Compassion manipulation. To manipulate compassion, we relied on work of Klimecki, Leiberg, Ricard, and Singer (2014), as well as from Stellar, Cohen, Oveis, and Keltner (2015) and Oveis, Horberg, and Keltner (2010). To induce compassion, these authors used sets of videos or pictures showing suffering people, but did not offer any explicit instructions of how to approach the stimuli (as in the "Batson-manipulation"). This way of inducing compassion is in line with McAuliffe et al. (2018), arguing that individuals respond spontaneously with compassion for a sympathetic suffering target. In our two studies, we made use of this idea.

In Study 6a, adapted from Klimecki et al. (2014) and Stellar et al. (2015), participants in the compassion condition read, "In the following, you will watch a video about George. George is 61 years old and homeless. The video takes two and a half minutes." Participants then watched a video of a suffering homeless old man (George) who talks about his difficult life on the street. Participants in the control group read, "Later in the study, you will get to know George. George is 61 years old and homeless. Before you judge a situation involving George, we ask you to watch a video about how to build a fence. The video takes two and a half minutes." Participants then watched a video about how to build a fence (cf. Stellar et al., 2015, Study 3). Videos in both conditions lasted 140 seconds.

In Study 6b, we adapted the procedure by Stellar et al. (2015, Study 2) and Oveis et al. (2010). Participants in the compassion condition read, "In the following, you will see 5 pictures of George. George is 61 years old. Every picture is presented for 5 seconds. The next picture will always appear automatically." Participants then saw five pictures of the suffering old homeless man of Study 6a. Participants in the first control group read the same text and then saw five different pictures of a nonsuffering man of roughly the same age sitting on a chair. Participants in the second control group read, "In the following, you will get to know George. George is 61 years old. Later in the study, you will judge a situation involving George." In this condition, participants received no pictures. As such, in both Studies 6a and 6b, we had *no* explicit instructions to downregulate emotions in the control group.

Manipulation check, moral outrage, and third-party punishment. The same items as in Study 2b and Study 5 were used. All scales showed satisfying reliability ($\alpha s > .85$). The manipulation check items were assessed right after watching the videos (Study 6a) or pictures (Study 6b). Then participants were given information that George suffered from unjust behavior. Participants read, "George reports about a recent accident: 'Last week I broke my shoulder. A young man deliberately bumped into me. Then I fell. This is how it happened.'" Finally, moral outrage and third-party punishment inclinations were assessed.

Participants. Power analysis was equivalent to previous studies (Studies 2–4), so we aimed to obtain at least $N = 64$ participants per cell. In Study 6a (two conditions) we sampled $N = 220$ participants (52.3% women; $M_{age} = 36.8$, $SD_{age} = 12.2$). In Study 6b (three conditions) we sampled $N = 332$ participants (53.6% women; $M_{age} = 36.5$, $SD_{age} = 12.6$).

Results and Discussion

Manipulation check. In Study 6a, analysis revealed a significant difference, $t(218) = 17.76$, $p < .001$; Cohen's $d = 2.39$ in state compassion between the compassion condition ($M = 5.61$, $SD = 1.19$) and control condition ($M = 2.45$, $SD = 1.44$). The same pattern emerged in Study 6b, $F(2, 329) = 84.60$, $p < .001$; state compassion was significantly higher, $t(218) = 10.90$, $p < .001$; Cohen's $d = 1.47$ in the compassion condition ($M = 5.19$, $SD = 1.29$) than in the first control condition ($M = 3.21$, $SD = 1.40$), and it was significantly higher, $t(219) = 11.96$, $p < .01$; Cohen's $d = 1.61$ than in the second control condition ($M = 2.90$, $SD = 1.54$). As such, in both studies, we successfully induced compassion using a different procedure than in previous studies.

Next, in Study 6a, although state outrage was on a low level (the means were far below the midpoint on the 1 to 7-point scale, $p < .001$), there was a significant difference in state outrage, $t(218) = 8.05$, $p < .001$; Cohen's $d = 1.08$ between the compassion condition ($M = 2.83$, $SD = 1.71$) and the control condition ($M = 1.33$, $SD = 0.94$). The same pattern emerged in Study 6b, $F(2, 329) = 41.00$, $p < .001$; state outrage was significantly higher, $t(218) = 7.70$, $p < .001$; Cohen's $d = 1.04$ in the compassion condition ($M = 2.68$, $SD = 1.49$) than in the first control condition ($M = 1.39$, $SD = 0.94$), and it was significantly higher, $t(219) = 6.94$, $p < .001$; Cohen's $d = 0.93$ than in the second control condition ($M = 1.48$, $SD = 1.04$).[9] In addition, happiness was reduced (Study 6a Cohen's $d = -1.26$, Study 6b $d = -1.11$) and disgust was increased (Study 6a Cohen's $d = 0.57$, Study 6b $d = 0.61$). We address below the problem that not only compassion was altered by running a multiple regression analysis including state compassion, state outrage, and the other emotions as predictor to see which emotion drives the effects (see section "Additional Analyses" below).

Third-party punishment. In Study 6a, analysis revealed a significant difference, $t(218) = 3.59$, $p < .001$; Cohen's $d = 0.48$ in third-party punishment between the compassion condition ($M = 5.09$, $SD = 1.45$) and the control condition ($M = 4.37$, $SD = 1.53$). This finding was replicated in Study 6b, $F(2, 329) = 8.43$, $p < .001$; third-party punishment was significantly higher, $t(218) = 2.93$, $p < .01$; Cohen's $d = 0.39$ in the compassion condition ($M =$

[9] We also tested in independent studies the material we relied on in our studies, i.e., Klimecki et al. (2014); Oveis et al. (2010), and Stellar et al. (2015), and found the same pattern; that is, compassion was increased, but so was state outrage on a low level. This is not surprising, because the stimuli include suffering people and the displayed suffering can be attributed to injustice and thus increase state outrage.

4.68, $SD = 1.63$) than in the first control condition ($M = 4.02$, $SD = 1.73$), and it was significantly higher, $t(219) = 3.99$, $p < .001$; Cohen's $d = 0.54$ than in the second control condition ($M = 3.75$, $SD = 1.82$).

Additional analyses. Because both state compassion and state outrage were increased by the experimental manipulation, we predicted third-party punishment in a multiple regression analysis to test which emotion is crucial and responsible for the effect of the experimental manipulation on third-party punishment. Accordingly, we simultaneously entered state compassion and state outrage and the other emotions in a multiple regression model. In both studies, whereas state compassion was a significant and relevant (in terms of effect size) predictor of third-party punishment (ΔR^2s $> .10$), state outrage and the other emotions were not a significant predictor and not relevant in terms of effect size (ΔR^2s $< .015$). We conclude that although the manipulation increased state outrage on a low level, in addition to compassion (as in other research, see footnote 9), our analyses show that state compassion but not state outrage is crucial, and it is responsible for the effect of the experimental manipulation on third-party punishment.

Mediation analysis. To reduce redundancy, the statistical parameters are reported in detail on the OSF. In line with the previous studies, the analyses showed that the experimental manipulation significantly increased moral outrage. Again, state compassion and not state outrage was responsible for the effect of the manipulation; only state compassion was a significant and relevant predictor of moral outrage in terms of effect size (ΔR^2s $> .10$), whereas state outrage was not a significant predictor and not relevant in terms of effect size (ΔR^2s $< .013$). Moral outrage in turn predicted third-party punishment. The bootstrapped 95% confidence intervals in both studies excluded zero (lower 95% CIs $> .16$).

Discussion. In Studies 6a and 6b we replicated the findings of the previous studies using a different manipulation of compassion. This manipulation builds on the idea that individuals respond spontaneously with compassion for a sympathetic suffering target (McAuliffe et al., 2018). Although the manipulation also increased state outrage, we showed statistically that only compassion was crucial and responsible for increased third-party punishment. Please note that in the previous studies reported in the contribution, we had a more specific manipulation of compassion that did not increase state outrage (the "Batson-manipulation"). This manipulation, however, included an explicit instruction to downregulate emotions in the control group. This shortcoming was addressed in Studies 6a and 6b. We conclude that there is overall strong evidence that compassion increases third-party punishment. This is also supported by the internal meta-analysis reported in the following.

Internal Meta-Analysis

We ran an internal meta-analysis including the experimental studies (Studies 2 to 6) and the three additional studies reported on the OSF (see also section "Additional Studies" in this paper). In total, this resulted in 13 main effects between an experimental (compassion) condition and a control condition included in the analysis. Again, we used fixed effects in which the mean effect size was weighted by sample size (Goh et al., 2016). The overall effect was significant, $M\,d = 0.38$, $SE = 0.04$, $Z = 9.17$, $p < .001$, two-tailed. The lower 95% CI of the effect was 0.30, and the upper 95% CI was 0.46. As such, the present set of studies suggests that compassion increases third-party punishment. Here, a small-to-medium effect was found.

General Discussion

People are frequently confronted with suffering, as is evident in the media's extensive coverage of terrorist attacks, the suffering of war, violent acts such as homicides, and brutal beatings of innocent persons. Given the omnipresent confrontation with suffering, we aimed to provide a better understanding of resulting interpersonal consequences. In Studies 1a–1f, we have shown that compassion for individuals suffering from unjust actions was correlated with punishment inclinations toward those responsible for the suffering. This relation could be documented across a variety of contexts of high societal relevance (i.e., terrorist attacks, sexual assaults, rape of children, and war). In the consecutive studies (Studies 2–6), we applied an experimental approach and showed that compassion magnifies punishment inclinations toward those responsible for unjust suffering. Additional analyses aiming to explain this relation revealed that compassion intensified moral outrage about the unjust suffering, moral outrage, in turn, was related to third-party punishment.

In sum, we have shown across six studies in the laboratory and online, as well as across different countries, that compassion for people suffering from unjust actions magnifies third-party punishment toward those who have caused unjust suffering. These results of the present set of studies have numerous implications for research on compassion and third-party punishment, as well as for society. These implications are discussed in the following sections.

Implications for Research on Compassion

Emotions prepare for action (Scherer, 2005). With respect to compassion, the last decades of research on this moral emotion have shown that compassion leads to a gamut of pro-social behaviors, ranging from caring for children and babies in need to helping people who contracted a disease; compassion even prepares individuals to engage in proenvironmental behavior (Goetz et al., 2010; Pfattheicher, Sassenrath, & Schindler, 2016). However, possible antisocial or harm-causing tendencies of compassion were almost completely overlooked. The few exceptions include research by Batson, Batson, et al. (1995), showing that compassion increases allocation of resources to a suffering individual, thus exploiting, however, the resources of the collective. Lupoli, Jampol, and Oveis (2017) document that compassion increases lying behavior to protect a suffering person. Buffone and Poulin (2014) show that compassion for a distressed person increases harming behavior toward a third individual who faces an upcoming competition with the distressed person. These findings indicate that compassion can have negative consequences for others. Still, they fit the overall picture that the main emerging tendency of compassion is to help a suffering other (Goetz et al., 2010); in each study, the person for whom compassion is felt benefits from compassionate individuals' actions. In the present work, we examined what happens when compassionate individuals cannot directly help a victim (i.e., improving the state of a suffering

individual by his or her own actions) but when the possibility to harm the perpetrator is given (i.e., impair the state of another individual). In such situations, compassion increases harm, that is, punishment. Of course, we must acknowledge that our findings remain restricted to situations in which it is difficult for individuals to help. However, direct helping is often difficult. Think, for example, about when terrorist attacks happen, when viewing media reports about the war in Syria, or when perpetrators commit sexual assaults; it is difficult to help the victims directly, but one can generate retaliatory tendencies and support for punishment of the suffering-causing third parties. Indeed, our findings apply to these specific situations, but this fact does not make the findings less impactful.

Our findings can additionally be discussed in reference to work by Batson, Klein, Highberger, and Shaw (1995). These authors tested consequences of compassion when altruism and justice are at odds, specifically whether compassion and resulting helping for one individual can increase injustice to other individuals. Empirically it is shown that with low feelings of compassion, individuals acted in line with the justice principle: The vast majority of participants did not reassign a terribly ill child on a waiting list to immediate help at the expense of others who were more needy. However, when inducing compassion for the child, a majority of participants violated the justice principle and helped the child at the expense of others. Batson, Klein, et al. (1995) created a situation in which altruism and justice are at odds and show that compassion motivates people to help—at the costs of violating the justice principle. Batson, Klein, et al. (1995, p. 1053) states that "if empathy can be evoked for the victims of injustice, then these two motives can be made to work together rather than at odds." This is exactly what we show in the present contribution: Compassion for a victim suffering from injustice intensifies the perception of injustice and its downstream consequences. Compassion thus reflects, as Batson, Klein, et al. (1995, p.1053) put it, the "emotional fire" that intensifies emotional reactions to cared-for suffering others.

We furthermore want to discuss the implications of the present work with respect to incidental and integral compassion. Condon and DeSteno (2011) have shown that incidental compassion, elicited by an uninvolved sad third person who suffered because she found out her brother had cancer, reduces punishment of a third uninvolved person who cheated on a task to win money. In the present work, we have shown that integral compassion can increase third-party punishment. Note, however, that our study and the study by Condon and DeSteno (2011) not only differed in the aspect of incidental versus integral compassion, but also regarding third-party punishment. Third-party punishment in our studies was about the perpetrator who hurt a victim the participant was feeling compassion toward, whereas the study by Condon and DeSteno (2011) was about third-party punishment toward an unrelated individual who has not hurt someone but has cheated on a task. Taken together, these findings point to the complexity of compassion in the context of third-party punishment, in that harm-reducing as well as harm-increasing tendencies can result from compassion.

A relevant question that needs further discussion is whether (a) compassionate individuals consider harming a perpetrator as an ultimate goal in itself, or (b) whether harm serves as an instrumental means to reach a higher goal, that is, to reestablish justice. The empirical findings of the present work speak in favor of the latter, because compassion-based third-party punishment is explained by moral outrage which represents a driving force for actions that reestablish justice (Carlsmith et al., 2002). Yet we acknowledge that we did not test these two explanatory accounts against each other. In discussing the specific motives of compassionate punishers, one can also consider the possibility that individuals punish to deter future suffering; in this regard, the seminal work by Carlsmith and colleagues (2002, 2008) shows that although people typically express strong preferences for deterrence, individuals' actual behavior is driven exclusively by just deserts concerns. Tan and Xiao's (in press) recent work supports this idea and shows that third-party punishment is driven by just deserts concerns rather than by deterrence motives. Of course, this does not conclusively answer the question about whether third-party punishment based on compassion may also be driven by just deserts concerns.

In addition we want to emphasize that additional processes can contribute to compassion increasing third-party punishment. That is to say, we only provide empirical support for compassion intensifying moral outrage, which in turn motivates third-party punishment. It could be that compassion is accompanied by being more identified with the victim (Cassell, 2009), which increases moral outrage about the victim being harmed (although we did not find initial support for this idea in Studies 6a and 6b), or that compassion is accompanied by a self-other overlap with the suffering person (Batson et al., 1997), motivating third-party punishment. It could also be that compassionate individuals who punish a perpetrator want to signal to others that they do not accept unjust behavior (Jordan, Hoffman, Bloom, & Rand, 2016). Basically, we do not argue for a single mediator perspective; it seems reasonable to assume that additional processes can contribute to compassion increasing third-party punishment (cf. Bullock et al., 2010; Fiedler, Schott, & Meiser, 2011, and Fiedler, Harris, & Schott, 2018, for this basic idea).

The overall picture in the literature on compassion suggests that compassion might function as a possibility to reduce escalation of violence (Batson & Ahmad, 2001; Condon & DeSteno, 2011; Singer & Steinbeis, 2009), but given the findings of the present research, compassion might also add fuel to the fire (Antonetti & Maklan, 2017; Condon & DeSteno, 2017). Specifically, it has been shown that peer punishment (such as third-party punishment in the present contribution) can create a downward cycle of retribution and counterpunishment (Nikiforakis, 2008; Pinker, 2011), leading to suboptimal intra- and interpersonal states (Dreber, Rand, Fudenberg, & Nowak, 2008; Pfattheicher, Böhm, & Kesberg, 2018). From this perspective, we cloud the rosy picture that is typically drawn of compassion as well as the notion that one can reach global peace by practicing compassion whenever suffering occurs (Dalai Lama & Ekman, 2008). These thoughts are in line with recent considerations about the limits of empathy and compassion (e.g., Bloom, 2016; Breithaupt, 2015; Decety & Cowell, 2014; Sassenrath, Hodges, & Pfattheicher, 2016).

Understanding the Emergence of Third-Party Punishment

The present research has implications for our understanding of how and why third-party punishment toward those who caused unjust suffering emerges. Past research has explained the emer-

gence of third-party punishment by arguing that unfair behavior reflects the violation of a fairness norm; this norm violation elicits anger, which in turn motivates punishment of the third party (Fehr & Fischbacher, 2004; Nelissen & Zeelenberg, 2009; Jordan, McAuliffe, & Rand, 2016). With our contribution, we point to a significant factor contributing to this cascade: compassion for a suffering victim. We show that compassion magnifies anger (moral outrage), which in turn fosters third-party punishment. As such, we contribute to a better understanding about the evolution of third-party punishment, and thus add one piece of significant knowledge to the existing literature.

We further emphasize that the present research has applied a broad conceptualization of third-party punishment, incorporating any behavior that includes negative sanctions and consequences or the support of negative sanctions and consequences targeted at a third party that has caused harm. This broad conceptualization made it possible to include a broad range of tendencies, including support for retribution of terror attacks, support of the death penalty for rapists, and third-party punishment of unfair individuals in economic games. As such, we did not restrict our study program to economic games (which one could argue is the current gold standard in the investigation of third-party punishment); rather, we aimed to test the idea of compassion motivating third-party punishment across a vast variety of social situations and across different moral domains (e.g., harm, fairness; Graham et al., 2011), which reflects a strength of the present contribution.

Finally, we argue that the idea of compassion fostering third-party punishment can also be applied to intergroup hate and aggression, for instance in the Middle East region (Böhm, Thielmann, & Hilbig, 2018; Bruneau, Cikara, & Saxe, 2017; Vanman, 2016). Today, political leaders from Israel and Palestine stress past suffering caused by members of the other group. When people feel compassion for suffering individuals of their own group, they might judge this suffering as particularly unjust, eliciting moral outrage and, consequently, increase the chances for intergroup aggressive acts. Basically, we argue that compassion for victims of the in-group might stimulate punishment inclinations and aggressive tendencies toward out-group members who caused the suffering (cf. Gordijn, Wigboldus, & Yzerbyt, 2001). We do not, however, want to demonize compassion as it is an important emotion that has the potential to motivate pro-social behavior (Haidt, 2003; Goetz et al., 2010). What seems important is for whom compassion is felt. In this regard, Mazziotta, Feuchte, Gausel, and Nadler (2014) have shown that thinking about harm during war that was inflicted by one's own group (i.e., a perpetrator-focus) elicits compassionate feelings for victims of an outgroup and increased willingness to engage in cross-group contact, in contrast to thinking about harm that was inflicted on the own group (i.e., a victim-focus; see also Roth, Shane, & Kanat-Maymon, 2017). By combining our research and the work of Mazziotta et al. (2014), important policy implications can be made. Compassion for victims' suffering, caused by their own harmful actions, or by the in-group, seems beneficial for peace (Mazziotta et al., 2014); in contrast, as shown in the present research, compassion for victims suffering from harmful actions caused by third persons can lead to punishment inclinations, potentially stimulating escalation of violence rather than preventing it (Condon & DeSteno, 2017). Basically, we argue that one should be cautious about claims that promote the practice of unconditional compassion whenever suffering occurs (Dalai Lama & Ekman, 2008).

Limitations and Outlook

At this point we acknowledge limitations of the present work and point to potential future research. First, we want to discuss three methodological points. The first methodological point is the discussion of possible demand characteristics. We think that two arguments speak against the possibility that demand characteristics are driving the effects of the present contribution. First, one could assume that it is socially demanded that one shows helping behavior toward a victim, especially when one is asked to feel compassion (as the work by Batson and colleagues suggest). However, we think that social demands are less likely present when it comes to punishment (and thus harm) of a perpetrator. We acknowledge, however, that we have no empirical data for this claim. Second, in our four studies that made use of the third-party punishment paradigm (Studies 3a–3c, 4), punishment was incentivized and thus was associated with financial costs. Accordingly, adhering to demand characteristics (resulting in increased punishment) is costly, which likely reduces its occurrence. Overall, although we cannot fully rule out that demand characteristics might have played a role, we argue that they are too weak to explain away the basic effect of compassion magnifying third-party punishment.

The second methodological point is that we have not, in the present set of studies, examined compassion-based third-party punishment in face-to-face situations. We have deliberately avoided face-to-face interactions and used scenario versions and a game theoretical paradigm (i.e., the third-party punishment paradigm) because these were relatively neutral. Hence, these situations are more parsimonious regarding the information that is salient in the situation. Accordingly, possible confounding factors that may come along with real face-to-face interactions (e.g., attractiveness, gender) can be reduced. This leaves room for future research to study the idea of compassion-based third-party punishment in face-to-face interactions, for instance in front of a court where judges might feel compassion for raped children, fostering their inclination to impose the death penalty on the rapist.

The third methodological point that we want to discuss is that in Studies 3a–3c (and two studies reported above in the section "Additional Studies") we have manipulated compassion for another individual who suffered from injustice by a third person. As such, this situation confounded compassion (elicited by suffering) and moral outrage (elicited by injustice). On the one hand, this confound is, of course, a shortcoming. On the other hand, it is a strength in the sense that the present research program has included situations in which suffering and injustice *naturally* occur together, which is often the case. Terrorist attacks provide a useful example; these create substantial suffering that is highly unjust. The same applies to the rapist that creates highly unjust suffering in children, or people suffering from unjust actions in the Syrian war. The confound is addressed in Studies 2 and 4–6. Here, compassion was manipulated prior to injustice so that compassion and injustice were independent. These situations might less likely occur in real life but are, from a methodological perspective, more valid as the confounding condition is avoided. In sum, it is a strength of the present contribution that situations are included in

which both suffering and injustice naturally occur together as well as situations in which compassion precedes injustice.

We further note that we have examined effects of compassion on third-party punishment on a short-time scale. Thus, we cannot draw inferences about effects of constant and ongoing suffering and whether retaliatory tendencies on the basis of compassion become chronic (e.g., chronic hostile tendencies against those responsible for terrorist attacks). Past research has shown that overwhelming levels of suffering can lead to a collapse of compassion, especially for individuals who are skilled at emotion regulation (Cameron & Payne, 2011). Thus, in cases where individuals constantly self-regulate compassion, we expect that retaliatory tendencies are less likely to become chronic. However, if individuals do not self-regulate recurring compassion toward individuals suffering from injustice, hostile tendencies on the basis of compassion might become chronic. Indeed, past research has revealed positive relations between chronic compassion and chronic hostility toward other people in general (Keller & Pfattheicher, 2013), yet long-term effects of the exposure to suffering still need to be investigated.

Moreover, our research remains silent regarding the default behavioral tendency that results from compassion. This thought refers to the question of whether the first impulse of compassionate individuals is to help a person suffering from injustice, or whether the first impulse is to harm those responsible for the suffering. Because compassion is elicited by the attention to a suffering person, we assume that the first impulse is to help the person and to only later turn attention to those who caused the suffering. These considerations are in line with research of Weng, Fox, Hessenthaler, Stodola, and Davidson (2015; see also Jordan, Hoffman, et al., 2016) who offered their participants the possibility of helping a victim *and* punishing the responsible harmdoer. The authors only found an effect of compassion training on helping behavior, not on punishment. These findings suggest that the default tendency of compassionate individuals is to help, yet with no possibility of helping, as in our research, compassion can also foster third-party punishment.

We emphasize that the present research incorporates a variety of situations in which compassion-based third-party punishment may emerge, yet the applicability of the basic idea to additional situations of high societal relevance seems possible. One can wonder whether compassion for suffering victims causes judges to inflict stronger punishment on a perpetrator. It even seems possible that compassion for a victim might increase the chance that an accused perpetrator receives a death sentence. This idea was put to a first correlational test in Study 1f but has not been experimentally investigated; thus, causal conclusions cannot be drawn. Moreover, the idea of compassion-based third-party punishment can be applied to intergroup contexts, for instance, to the conflict between Israel and Palestine where substantial suffering, compassion, and retaliation occur. Basically, we argue that the present research can reflect a solid basis for studies on compassion and third-party punishment in a variety of social contexts.

Conclusions

In this contribution, we have shown that punishment inclinations can be magnified by an emotion that is considered to be a prosocial tendency par excellence, that is, compassion (Dalai Lama & Ekman, 2008; Goetz et al., 2010). Specifically, by showing that compassion relates to harmful tendencies when unjustified suffering has occurred, we applied a differentiated perspective on compassion and its interpersonal consequences. As such, the current research opens a new avenue of research for studying compassion, hopefully inspiring both basic research as well as research from an applied, societal perspective.

References

Antonetti, P., & Maklan, S. (2017). Concerned protesters: From compassion to retaliation. *European Journal of Marketing, 51,* 983–1010. http://dx.doi.org/10.1108/EJM-03-2016-0145

Ashar, Y. K., Andrews-Hanna, J. R., Yarkoni, T., Sills, J., Halifax, J., Dimidjian, S., & Wager, T. D. (2016). Effects of compassion meditation on a psychological model of charitable donation. *Emotion, 16,* 691–705. http://dx.doi.org/10.1037/emo0000119

Bardslay, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental economics. Rethinking the rules*. New Jersey: Princeton University Press.

Barton, A. (2004). Just deserts theory. In M. Bosworth (Ed.), *Encyclopedia of prisons and correctional facilities* (pp. 504–507). Thousand Oaks, CA: Sage.

Batson, C. D. (2009). These things called empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–16). Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/9780262012973.003.0002

Batson, C. D. (2011). *Altruism in humans*. United Kingdom: Oxford University Press.

Batson, C. D., & Ahmad, N. (2001). Empathy-induced altruism in a prisoner's dilemma II: What if the target of empathy has defected? *European Journal of Social Psychology, 31,* 25–36. http://dx.doi.org/10.1002/ejsp.26

Batson, C. D., Batson, J. G., Griffitt, C. A., Barrientos, S., Brandt, J. R., Sprengelmeyer, P., & Bayly, M. J. (1989). Negative-state relief and the empathy–altruism hypothesis. *Journal of Personality and Social Psychology, 56,* 922–933. http://dx.doi.org/10.1037/0022-3514.56.6.922

Batson, C. D., Batson, J. G., Todd, R. M., Brummett, B. H., Shaw, L. L., & Aldeguer, C. M. (1995). Empathy and the collective good: Caring for one of the others in a social dilemma. *Journal of Personality and Social Psychology, 68,* 619–631. http://dx.doi.org/10.1037/0022-3514.68.4.619

Batson, C. D., Chang, J., Orr, R., & Rowland, J. (2002). Empathy, attitudes, and action: Can feeling for a member of a stigmatized group motivate one to help the group? *Personality and Social Psychology Bulletin, 28,* 1656–1666. http://dx.doi.org/10.1177/014616702237647

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

Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology, 68,* 1042–1054. http://dx.doi.org/10.1037/0022-3514.68.6.1042

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., . . . Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72,* 105–118. http://dx.doi.org/10.1037/0022-3514.72.1.105

Bekkers, R., & Ottoni-Wilhelm, M. (2016). Principle of care and giving to help people in need. *European Journal of Personality, 30,* 240–257. http://dx.doi.org/10.1002/per.2057

Bloom, P. (2016). *Against empathy: The case for rational compassion*. New York, NY: HarperCollins.

Böhm, R., Thielmann, I., & Hilbig, B. E. (2018). The brighter the light, the deeper the shadow: Morality also fuels aggression, conflict, and violence. *Behavioral and Brain Sciences, 41,* e98.

Breithaupt, F. (2015). Why empathy is not the best basis for humanitarianism. *Global Cooperation Research Papers, 9,* 5–16.

Bruneau, E. G., Cikara, M., & Saxe, R. (2017). Parochial empathy predicts reduced altruism and the endorsement of passive harm. *Social Psychological and Personality Science, 8,* 934–942. http://dx.doi.org/10.1177/1948550617693064

Buffone, A. E., & Poulin, M. J. (2014). Empathy, target distress, and neurohormone genes interact to predict aggression for others-even without provocation. *Personality and Social Psychology Bulletin, 40,* 1406–1422. http://dx.doi.org/10.1177/0146167214549320

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology, 98,* 550–558. http://dx.doi.org/10.1037/a0018933

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology, 100,* 1–15. http://dx.doi.org/10.1037/a0021643

Cameron, C. D., & Payne, B. K. (2012). The cost of callousness: Regulating compassion influences the moral self-concept. *Psychological Science, 23,* 225–229. http://dx.doi.org/10.1177/0956797611430334

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in Experimental Social Psychology, 40,* 193–236. http://dx.doi.org/10.1016/S0065-2601(07)00004-4

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. http://dx.doi.org/10.1037/0022-3514.83.2.284

Cassell, E. J. (2009). Compassion. In S. J. Lopez & C. R. Snyder (Eds.), *Oxford library of psychology. Oxford handbook of positive psychology* (pp. 393–403). New York, NY: Oxford University Press.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology, 73,* 481–494. http://dx.doi.org/10.1037/0022-3514.73.3.481

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Condon, P., & DeSteno, D. (2011). Compassion for one reduces punishment for another. *Journal of Experimental Social Psychology, 47,* 698–701. http://dx.doi.org/10.1016/j.jesp.2010.11.016

Condon, P., & DeSteno, D. (2017). Enhancing compassion: Social psychological perspectives. In E. Seppala, J. Doty, M. Worline, D. Cameron, S. Brown, & E. Simon-Thomas (Eds.), *Oxford handbook of compassion science* (pp. 287–298). United Kingdom: Oxford University Press.

Crocker, J., Canevello, A., & Lewis, K. A. (2017). Romantic relationships in the ecosystem: Compassionate goals, nonzero-sum beliefs, and change in relationship quality. *Journal of Personality and Social Psychology, 112,* 58–75. http://dx.doi.org/10.1037/pspi0000076

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

Dalai Lama, & Ekman, P. (2008). *Emotional awareness: Overcoming the obstacles to psychological balance and compassion*. New York, NY: Henry Holt.

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. http://dx.doi.org/10.1023/A:1005552203727

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44,* 113–126. http://dx.doi.org/10.1037/0022-3514.44.1.113

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science, 9,* 525–537. http://dx.doi.org/10.1177/1745691614545130

DeSteno, D. (2015). Compassion and altruism: How our minds determine who is worthy of help. *Current Opinion in Behavioral Sciences, 3,* 80–83. http://dx.doi.org/10.1016/j.cobeha.2015.02.002

Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don't punish. *Nature, 452,* 348–351. http://dx.doi.org/10.1038/nature06723

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., Smith, M., . . . Suh, K. (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *Journal of Personality and Social Psychology, 66,* 776–797. http://dx.doi.org/10.1037/0022-3514.66.4.776

Eisenberg, N., McCreath, H., & Ahn, R. (1988). Vicarious emotional responsiveness and prosocial behavior: Their interrelations in young children. *Personality and Social Psychology Bulletin, 14,* 298–311. http://dx.doi.org/10.1177/0146167288142008

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200. http://dx.doi.org/10.1080/02699939208411068

Eriksson, K., Strimling, P., Andersson, P. A., & Lindholm, T. (2017). Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff reduction. *Journal of Experimental Social Psychology, 69,* 59–64. http://dx.doi.org/10.1016/j.jesp.2016.09.004

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry, 41,* 20–26. http://dx.doi.org/10.1093/ei/41.1.20

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

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

Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences, 36,* 1015–1029. http://dx.doi.org/10.1016/S0191-8869(03)00197-1

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests—An analysis of articles published in 2015. *Journal of Experimental Social Psychology, 75,* 95–102. http://dx.doi.org/10.1016/j.jesp.2017.11.008

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, 47,* 1231–1236. http://dx.doi.org/10.1016/j.jesp.2011.05.007

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology, 18,* 255–297. http://dx.doi.org/10.1111/0162-895X.00058

Gervais, M. M., & Fessler, D. M. T. (2017). On the deep structure of social affect: Attitudes, emotions, sentiments, and the case of "contempt." *Behavioral and Brain Sciences, 40,* e225. http://dx.doi.org/10.1017/S0140525X16000352

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin, 136,* 351–374. http://dx.doi.org/10.1037/a0018807

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10,* 535–549. http://dx.doi.org/10.1111/spc3.12267

Gordijn, E. H., Wigboldus, D., & Yzerbyt, V. (2001). Emotional consequences of categorizing victims of negative outgroup behavior as in-

group or outgroup. *Group Processes & Intergroup Relations, 4,* 317–326. http://dx.doi.org/10.1177/1368430201004004002

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101,* 366–385. http://dx.doi.org/10.1037/a0021847

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

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hechler, S., & Kessler, T. (2018). On the difference between moral outrage and empathic anger: Anger about wrongful deeds or harmful consequences. *Journal of Experimental Social Psychology, 76,* 270–282. http://dx.doi.org/10.1016/j.jesp.2018.03.005

Hoffman, M. L. (1990). Empathy and justice motivation. *Motivation and Emotion, 14,* 151–172. http://dx.doi.org/10.1007/BF00991641

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. United Kingdom: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511805851

Horberg, E. J., Oveis, C., & Keltner, D. (2011). Emotions as moral amplifiers: An appraisal tendency approach to the influences of distinct emotions upon moral judgment. *Emotion Review, 3,* 237–244. http://dx.doi.org/10.1177/1754073911402384

Jacoby, J., & Sassenberg, K. (2011). Interactions do not only tell us when, but can also tell us how: Testing process hypotheses by interaction. *European Journal of Social Psychology, 41,* 180–190. http://dx.doi.org/10.1002/ejsp.762

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530,* 473–476. http://dx.doi.org/10.1038/nature16981

Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics, 19,* 741–763. http://dx.doi.org/10.1007/s10683-015-9466-8

Kant, I. (1998). *Groundwork for the metaphysic of morals*. United Kingdom: Cambridge University Press. (Original work published 1785) http://dx.doi.org/10.1017/CBO9780511809590.004

Keller, J., & Pfattheicher, S. (2013). The compassion-hostility paradox: The interplay of vigilant, prevention-focused self-regulation, compassion, and hostility. *Personality and Social Psychology Bulletin, 39,* 1518–1529. http://dx.doi.org/10.1177/0146167213499024

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social Cognitive and Affective Neuroscience, 9,* 873–879. http://dx.doi.org/10.1093/scan/nst060

Kühnen, U. (2010). Manipulation checks as manipulation: Another look at the ease-of-retrieval heuristic. *Personality and Social Psychology Bulletin, 36,* 47–58. http://dx.doi.org/10.1177/0146167209346746

Lazarus, R. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.

Lerner, M. J. (1980). *Belief in a just world. A fundamental delusion*. New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-0448-5

Lupoli, M. J., Jampol, L., & Oveis, C. (2017). Lying because we care: Compassion increases prosocial lying. *Journal of Experimental Psychology: General, 146,* 1026–1042. http://dx.doi.org/10.1037/xge0000315

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

Mazziotta, A., Feuchte, F., Gausel, N., & Nadler, A. (2014). Does remembering past ingroup harmdoing promote postwar cross-group contact? Insights from a field-experiment in Liberia. *European Journal of Social Psychology, 44,* 43–52. http://dx.doi.org/10.1002/ejsp.1986

McAuliffe, W. H. B., Forster, D. E., Philippe, J., & McCullough, M. E. (2018). Digital altruists: Resolving key questions about the empathy-altruism hypothesis in an Internet sample. *Emotion, 18,* 493–506. http://dx.doi.org/10.1037/emo0000375

Mikula, G. (1993). On the experience of injustice. *European Review of Social Psychology, 4,* 223–244. http://dx.doi.org/10.1080/14792779343000077

Ministero, L. M., Poulin, M. J., Buffone, A. E. K., & DeLury, S. (2018). Empathic concern and the desire to help as separable components of compassionate responding. *Personality and Social Psychology Bulletin, 44,* 475–491. http://dx.doi.org/10.1177/0146167217741345

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

Nelissen, R. M., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby Effect. *Emotion, 9,* 118–122. http://dx.doi.org/10.1037/a0014540

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92,* 91–112. http://dx.doi.org/10.1016/j.jpubeco.2007.04.008

Nussbaum, M. (1996). Compassion: The basic social emotion. *Social Philosophy & Policy, 13,* 27–58. http://dx.doi.org/10.1017/S0265052500001515

Olderbak, S., Sassenrath, C., Keller, J., & Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Frontiers in Psychology, 5,* 653. http://dx.doi.org/10.3389/fpsyg.2014.00653

Oveis, C., Horberg, E. J., & Keltner, D. (2010). Compassion, pride, and social intuitions of self-other similarity. *Journal of Personality and Social Psychology, 98,* 618–630. http://dx.doi.org/10.1037/a0017628

Pfattheicher, S., Böhm, R., & Kesberg, R. (2018). The advantage of democratic peer punishment in sustaining cooperation within groups. *Journal of Behavioral Decision Making, 31,* 562–571. http://dx.doi.org/10.1002/bdm.2050

Pfattheicher, S., Sassenrath, C., & Schindler, S. (2016). Feelings for the suffering of others and the environment: Compassion fosters proenvironmental tendencies. *Environment and Behavior, 48,* 929–945. http://dx.doi.org/10.1177/0013916515574549

Pinker, S. (2011). *The better angels of our nature: The decline of violence in history and its causes*. London, UK: Penguin Books.

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology, 66,* 29–38. http://dx.doi.org/10.1016/j.jesp.2015.09.012

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences, 17,* 413–425. http://dx.doi.org/10.1016/j.tics.2013.06.003

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

Roth, G., Shane, N., & Kanat-Maymon, Y. (2017). Empathising with the enemy: Emotion regulation and support for humanitarian aid in violent conflicts. *Cognition and Emotion, 31,* 1511–1524. http://dx.doi.org/10.1080/02699931.2016.1237348

Rothschild, Z. K., & Keefer, L. A. (2018). Righteous or self-righteous anger? Justice sensitivity moderates defensive outrage at a third-party harm-doer. *European Journal of Social Psychology, 48,* 507–522. http://dx.doi.org/10.1002/ejsp.2349

Sassenrath, C., Hodges, S. D., & Pfattheicher, S. (2016). It's all about the self: When perspective taking backfires. *Current Directions in Psychological Science, 25,* 405–410. http://dx.doi.org/10.1177/0963721416659253

Sassenrath, C., Pfattheicher, S., & Keller, J. (2017). I might ease your pain, but only if you're sad: The impact of the role of empathized emotion in the empathy-helping association. *Motivation and Emotion, 41,* 96–106. http://dx.doi.org/10.1007/s11031-016-9586-2

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review, 22,* 32–70. http://dx.doi.org/10.1177/1088868317698288

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Sciences Information/Information Sur les Sciences Sociales, 44,* 695–729. http://dx.doi.org/10.1177/0539018405058216

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26,* 559–569. http://dx.doi.org/10.1177/0956797614567341

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology, 24,* R875–R878. http://dx.doi.org/10.1016/j.cub.2014.06.054

Singer, T., & Steinbeis, N. (2009). Differential roles of fairness- and compassion-based motivations for cooperation, defection, and punishment. *Annals of the New York Academy of Sciences, 1167,* 41–50. http://dx.doi.org/10.1111/j.1749-6632.2009.04733.x

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89,* 845–851. http://dx.doi.org/10.1037/0022-3514.89.6.845

Stellar, J. E., Cohen, A., Oveis, C., & Keltner, D. (2015). Affective and physiological responses to the suffering of others: Compassion and vagal activity. *Journal of Personality and Social Psychology, 108,* 572–585. http://dx.doi.org/10.1037/pspi0000010

Stellar, J. E., & Keltner, D. (2014). Compassion. In M. Tugade, L. Shiota, & L. Kirby (Eds.), *Handbook of positive emotion* (pp. 329–341). New York, NY: Guilford Press.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109,* 451–471. http://dx.doi.org/10.1037/0033-295X.109.3.451

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78,* 853–870. http://dx.doi.org/10.1037/0022-3514.78.5.853

Thulin, E. W., & Bicchieri, C. (2016). I'm so angry I could help you: Moral outrage as a driver of victim compensation. *Social Philosophy & Policy, 32,* 146–160. http://dx.doi.org/10.1017/S0265052516000145

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology, 11,* 59–63. http://dx.doi.org/10.1016/j.copsyc.2016.06.007

Vitaglione, G. D., & Barnett, M. A. (2003). Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. *Motivation and Emotion, 27,* 301–325. http://dx.doi.org/10.1023/A:1026231622102

Von Hirsch, A. (1976). *Doing justice: The choice of punishments*. New York, NY: Hill and Wang.

Weng, H. Y., Fox, A. S., Hessenthaler, H. C., Stodola, D. E., & Davidson, R. J. (2015). The role of compassion in altruistic helping and punishment behavior. *PLoS ONE, 10,* e0143794. http://dx.doi.org/10.1371/journal.pone.0143794

Zaki, J. (2018). Empathy is a moral force. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 49–58). New York, NY: Guilford Press.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111,* 493–504. http://dx.doi.org/10.1037/pspa0000056

Received January 29, 2018
Revision received August 27, 2018
Accepted August 27, 2018 ■