

Moral barrier to compassion: How perceived badness of sufferers dampens observers' compassionate responses

Hongbo Yu ^{a,*}, Jie Chen ^a, Bernadette Dardaine ^a, Fan Yang ^{b,*}

^a Department of Psychological & Brain Sciences, University of California Santa Barbara, Santa Barbara, CA 93106, USA

^b Department of Psychology, The University of Chicago, Chicago, IL 60637, USA



ARTICLE INFO

Keywords:
Compassion
Willingness to help
Deservedness
Morality
Judgment
Suffering

ABSTRACT

Compassion has been theorized as a desirable prosocial emotion due to its potential to transcend arbitrary boundaries (e.g., race, physical distance) and motivate us to alleviate the suffering of all human beings. Our paper nevertheless examines a potential moral barrier to compassion—whether and how moral evaluations of the suffering and the sufferer hinder our compassion and prosocial motivation. In four pre-registered studies (total $N = 421$, within-participant design), we demonstrated that adult U.S. participants withheld their compassion and willingness to help when they perceived moral badness of the sufferer, even when the perceived moral badness did not directly cause the suffering. The effects were found in terms of diverse types of moral judgments, including the sufferers' immoral intention (e.g., harming another; Study 1), bad moral character (e.g., being a dishonest person; Study 2), and even mere associations with groups perceived as deserving of suffering based on moral status (Studies 3–4). Deservedness judgment—how much the sufferer was viewed as deserving the suffering—mediated the effect between moral judgment and compassionate responses. Importantly, participants judged withholding compassion based on moral deservedness as what should be done and what morally good people would do, suggesting that people hold a normative view of the tendency that might make it difficult to overcome. Our findings thus reveal moral judgment as a barrier that prevents us from alleviating the suffering of all human beings.

1. Introduction

As human beings, feeling compassion¹ for other people's suffering is one of our deeply cherished capacities and virtues (Goetz, Keltner, & Simon-Thomas, 2010; Mascaro et al., 2020; T. Singer & Klimecki, 2014). But do we feel compassion equally for all human beings, including morally bad people who are suffering? Or should we? For example, do we feel more compassion toward seniors who live in nursing homes than senior prisoners, when they are equally vulnerable to suffering from a deadly infectious disease? This is not merely a hypothetical scenario. At the early stage of COVID vaccination, Colorado Governor Jared Polis issued a statement reflecting his lack of compassion toward prisoners,

"There's no way it [COVID vaccine] is going to go to prisoners before it goes to people who haven't committed any crime. That's obvious."² Where might the Governor's intuition (and that of those who shared his view) come from? Do people withhold compassion to someone based on their moral character? That is, people's moral judgment could be a potential barrier to compassion, a possibility and psychological mechanism that this paper examines.

Universal compassion to all human beings, and even to all living things who are capable of experiencing pain and suffering, is often valued as one of our moral ideals and have been advocated in different philosophical and religious traditions (e.g., care ethics, Buddhism, utilitarianism; see Bentham, 1780/1988; Caviola, Everett, & Faber,

* Corresponding authors.

E-mail addresses: hongbo.yu@psych.ucsb.edu (H. Yu), fan.yang@uchicago.edu (F. Yang).

¹ Here, following recent theoretical and synthetic work, we adopted a definition of compassion as a "benevolent emotional response toward another who is suffering, coupled with the motivation to alleviate their suffering and promote their well-being" (Mascaro et al., 2020; see also, Goetz et al., 2010; Singer & Klimecki, 2014; Strauss et al., 2016; Scheffer et al., 2021). We therefore treat compassion and sympathy as synonyms, but distinguish compassion from empathy and its subordinate concepts (e.g., emotional contagion, theory-of-mind).

² John Ingold "Colorado's governor says prisoners won't be prioritized for a coronavirus vaccine. A state plan outlines otherwise.", *The Colorado Sun*, December 2, 2020.

2019; Lampert, 2005; Gilbert, 2020). Albert Schweitzer, for example, has stated that “unless we extend the circle of compassion to all living things, humanity will not find peace”. Compared to other prosocial emotions such as empathy³ (i.e., the ability to understand and share the feelings of others), compassion (i.e., concerns for suffering or misfortunes of others) is theorized to be unique in terms of its higher potential to extend to all human beings. For example, compassion has been argued (e.g., Bloom, 2017a, 2017b) to rely less on physical or social similarities between the observers and the sufferers, factors irrelevant to care and help, as empathy does (e.g., Cikara, Bruneau, & Saxe, 2011). Moreover, compassion appears to be less susceptible to the “identifiable victim” effect than empathy is—the tendency to care about the suffering of a concrete victim as opposed to a larger collective of individuals (Small, Loewenstein, & Slovic, 2007). It has been found that people more positively evaluated those who withheld empathy for negative targets who suffer as a result of their immorality (Wang & Todd, 2021). Therefore, it has been argued that empathy is not an optimal driver for alleviating as much suffering as possible (P. Singer, 2016). Compassion, in contrast, relies less on automatic physiological responses and more on rational, principle-based cognitive processes (Batson, Klein, Highberger, & Shaw, 1995; Bloom, 2017a, 2017b; Decety & Cowell, 2014; T. Singer & Klimecki, 2014). Based on these considerations, compassion has recently been proposed as an alternative to empathy as a driver of altruism and prosociality (e.g., Bloom, 2017a, 2017b), especially in health care context (Sinclair et al., 2016, 2018; Fernando III and Condeline, 2014).

In contrast to the view of compassion as a wide-reaching affective capacity and driver behind prosociality, some evolutionary theories posit the opposite, that compassion by design should be sensitive to the deservedness of the sufferer. These theories argue that compassion has evolved to conditionally help reciprocal cooperation partners, which requires assessing their (moral) deservedness, such as trustworthiness, reputation, and cooperative history (Axelrod & Hamilton, 1981; Frank, 1988; Goetz et al., 2010; Raihani, 2021; McCullough, 2020; Sznycer et al., 2017; Trivers, 1971). As Goetz et al. (2010) nicely put, the implication of these evolutionary theories is that “compassion should be sensitive to appraisals of deservingness and to whether or not the person suffering is altruistic, cooperative, and of good character”. This contrasts with the adaptive problem that empathy is evolved to solve, namely, unconditionally caring and alleviating suffering for kins, offspring, and those close to oneself, regardless of or even against certain moral concerns (e.g., fairness) (Fitouchi, André and Baumard, 2023). Indeed, psychological research has suggested that people sometimes judge their own empathy-driven decisions as unfair, such as when empathizing with someone leads them to unfairly favor one person over equally deserving individuals (Batson et al., 1995). In this view, both empathy and compassion are constrained by the adaptive goals or functions they should serve. While the former should track genetic relatedness and closeness (not moral deservedness), the latter should be sensitive to reciprocity (and moral deservedness), especially among strangers. From a cognitive and computational perspective, the mechanisms of compassion seem computationally connected to the mechanisms of moral evaluation while the mechanisms of empathy may operate independently of moral evaluation (Cosmides, Guzmán, & Tooby, 2018; Sznycer et al., 2017).

One way to tease apart these two contrasting views is to test empirically whether and how moral considerations hinder compassion for all human beings. While little empirical work has systematically examined the moral barriers to compassion, there is evidence suggesting that compassion may indeed have its limitations and costs. For example,

people report that experiencing compassion in response to other's suffering can be unpleasant (Condon & Feldman Barrett, 2013), cognitively demanding (Cameron et al., 2019; Scheffer, Cameron, & Inzlicht, 2021), and require effortful self-regulation to avoid distress (Eisenberg & Eggum, 2009). More directly relevant to the boundaries of compassion, people express less compassion and more anger or schadenfreude to victims who are causally responsible for their own suffering (Van Dijk, Ouwerkerk, Goslinga, & Nieweg, 2005). Participants also expressed less compassion to a victim who accidentally harmed themselves (high causality in one's own suffering) compared to a victim who was intentionally harmed by someone else (low causality in one's own suffering) (Pfattheicher, Sassenrath, & Keller, 2019). These findings are consistent with the appraisal theory of compassion (Goetz et al., 2010; Rudolph, Roesch, Greitemeyer, & Weiner, 2004; Weiner, 1985), that for a sufferer to elicit strong compassion, they must not bring about the suffering through their own fault.

These findings provide evidence that judgment of causality of suffering is relevant to the appraisal of compassion, but it leaves unexamined one key characteristics of the sufferers that might influence compassion—the moral nature of the sufferers' characters and actions. Admittedly, morally bad people may often directly or indirectly cause their own suffering due to their moral badness, such as being punished, ostracized, and revenged for their bad actions and intentions, or suffer from immoral actions that they should not be doing (e.g., having a car accident from drunk driving). But just like morally good people, morally bad people may also suffer from factors clearly unrelated to their moral character and completely out of their own control (e.g., being hit by a falling tree while walking, or receiving disrespectful treatments from racists). Do we give compassion to people we perceive as morally bad who suffer from these non-self-caused, random misfortunes? Morally bad people who suffer from these misfortunes are clearly still a member of humanity, and we should feel compassion for them given the universal ideal of compassion (Gilbert, 2020; Lampert, 2005). Indeed, oftentimes to elicit the best from morally bad people and to help them become better, compassion is greatly needed (Zaki, 2019). But it is an empirical question whether that is people's actual tendency or not. Is having a morally bad character and status, though irrelevant and non-causal to suffering, sufficient to trigger the morality-based withholding of compassion? If there is such a tendency, then how generalized is it? For instance, does the person have to commit a moral transgression, or is simply associating with a morally bad person or group sufficient? And, importantly, is this tendency viewed as normative—do people believe that withholding compassion on the moral basis is itself morally good and required? Our paper examines moral judgment as an important and independent mechanism that influences compassion beyond causality judgment.

In psychology, it is known that morally bad characters and actions elicit a range of negative responses from observers (Brambilla, Sacchi, Rusconi, & Goodwin, 2021): individuals assign harsher blame and punishment (J. Jordan, Amir, & Bloom, 2016; Siegel, Crockett, & Dolan, 2017), attribute less happiness (Phillips, De Freitas, Mott, Gruber, & Knobe, 2017; Yang, Knobe, & Dunham, 2021), and grant lower moral status (Crimston, Hornsey, Bain, & Bastian, 2018) to morally bad characters and actions, even if moral character was not directly related to each of these cases at the first sight. The same wrongdoing would elicit harsher punishment and blame if the agent was described as a morally bad person compared to when the agent was described as a morally neutral person, both in the controlled laboratory setting (Siegel et al., 2017) and in the court (Roberts, 1997). Given the far-reaching and surprising impact of moral judgment, it is conceivable that people may have the natural tendency (although not necessarily the normative endorsement) to withhold compassion if they perceive moral badness in the sufferer, as a type of negative reaction to perceived moral badness.

Alternatively, even though people's moral judgment has pervasive consequences, it does not follow that people would necessarily withhold compassion toward the suffering of morally bad people. As we discussed

³ In some literature (e.g., Batson et al., 1987), researchers used “empathy” in a similar sense as our definition for “compassion”. For conceptual clarity, here we focused on the definition of empathy that is clearly distinct from compassion.

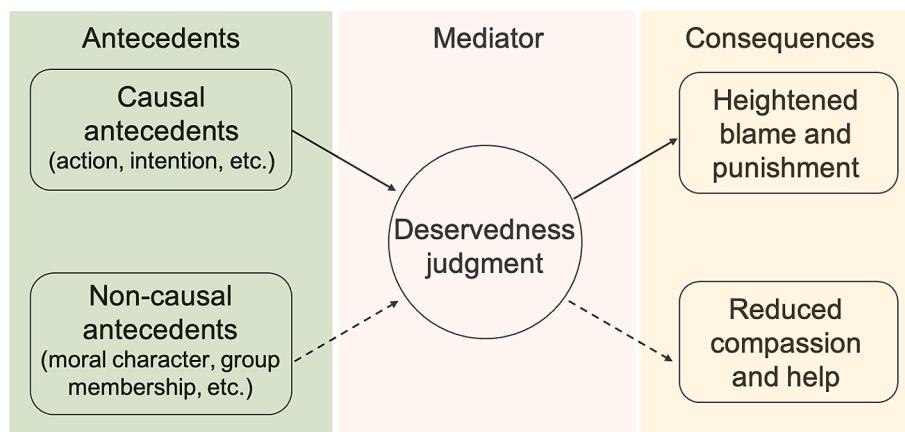


Fig. 1. A conceptual model of the antecedents and consequences of deservedness judgment.

at the beginning, given that compassion is a valued virtuous tendency and has been theorized as having the potential to be applied to all humanity in general (Bloom, 2017a), it is possible that people may not have a strong tendency to withhold compassion even to morally bad people, at least it should not be the ideal thing to do. But if, as evolutionary accounts of compassion would predict, moral evaluations of the suffering or the sufferer indeed hinder compassion, then it would reveal the constraints for compassion as a driving force for altruism and prosociality. Therefore, it is important to examine whether and how morality-based considerations pose a barrier to compassion.

More broadly, the role of moral judgment in compassion has interested scholars outside of psychology. For example, the potential moral barriers to compassion have been the focus of fruitful debates and research in sociology (Clark, 2007; Maestri & Monforte, 2020; Williams, 2008) and philosophy (Crisp, 2008; Kristjánsson, 2018; Nussbaum, 1996, 2003). For example, Clark (2007) provides sociological support for the presence of moral barriers to compassion. Clark observed that contemporary American adults feel compassion only for sufferings caused by “bad luck” or external problems, not for sufferings due to moral reasons within the person’s control (e.g., work ethics and efforts), in which case the suffering would be seen as “well-deserved”. Relatedly, research in political sciences has suggested that deservingness cues play a key role in determining people’s support for redistribution toward economically disadvantaged individuals (Aarøe & Petersen, 2014). The tendency might be due to the moralization of hard work influenced by a meritocracy ideology (Celniker et al., 2023; Sandel, 2020; Van Kleef et al., 2008). As the philosopher Michael Sandel argues, the logic of meritocracy “bases our obligation to help those in need not on compassion or solidarity but on how they came to be needy in the first place.” (Sandel, 2020).

Relatedly, a large body of contemporary philosophical discussions about compassion have focused on Nussbaum (1996, 2003) exposition of the Aristotelian concept of *eleos*—the modern equivalent of compassion. One element of Aristotle’s compassion has clear moral implications: the suffering should not be brought about by the sufferer’s own moral fault; if it were, then the sufferer *deserves* the suffering and should be more appropriately met with anger and blame than compassion (Nussbaum, 1996). In this philosophical framework, it is not only that people do not naturally feel compassion if the sufferer brings about the suffering through their own fault, but they *should* not feel compassion in that circumstance. This view and Clark’s observations (2017) also suggest that deservedness judgment—the extent to which the sufferer deserves the suffering—could be a potential mechanism mediating the effect of moral judgment on compassion.

In social psychology, deservedness judgment has been centered

around perception of causality and responsibility (Fig. 1, solid arrows). For example, the research along the line of attributional analysis has consistently shown that responsibility for one’s own plight (e.g., physical harm, poverty, etc.) elicits judgment of deservedness and, consequently, moral anger, heightened tendency to blame and punish in the observers (Dyer, Pizarro, & Ariely, 2022; Weiner, 1985, 1986, 1995; Weiner, 2010). This reflects an underlying assumption that an individual’s intention and action, typically seen as capable of *causing* a state of affairs (Robb & Heil, 2021), plays a critical, if not the only, role in assessing the moral worth of the individual and the situation they are in (Engelmann & Waldmann, 2022). In both the legal systems and everyday moral discourse in the Western, educated, industrialized, rich and democratic (WEIRD) societies, assigning blame and punishment typically involves fine-grained distinctions about mental states, as illustrated by the concept of *mens rea* (Young, Cushman, Hauser, & Saxe, 2007; Barrett & Saxe, 2021). Essentially, according to this view, only when an individual’s mental states and/or actions in some way physically cause a morally problematic outcome does this individual deserve blame and punishment. However, this assumption does not necessarily hold everywhere. Anthropological research has shown that the degree to which an individual’s intentions influence moral judgments varies across cultures, with intentions in some small-scale societies playing little to no role (Barrett et al., 2016; McNamara, Willard, Norenzayan, & Henrich, 2019). Classics also provides anecdotal examples that in Homeric Greek society, judgments of an individual’s deservedness (of punishment or suffering) have nothing to do with the intention or action of the individual themselves, but solely based on the family lineage of the individual (Dodds, 1951). In social psychology, the non-causal antecedent of deservedness judgment has largely been overlooked. In the present research, we systematically examined the role of perceived moral badness as a distinct non-causal antecedent of deservedness judgment and compared its effects with those of the causal antecedents (Fig. 1, dotted arrows).

We addressed these and related questions in four experimental studies. To measure participants’ compassionate responses, we included two main dependent variables: 1) feelings of compassion, which was measured by combining several emotion terms adopted from previous work on compassion (Goetz et al., 2010), 2) prosocial motivation or willingness to help the sufferer (Stellar, Anderson, & Gatchpazian, 2020). In addition, based on the philosophical conjecture and sociological observations, we hypothesized that deservedness of suffering might be a psychological mechanism through which the morality-based withholding of compassion takes effect. We therefore also measured participants’ judgments of the deservedness of the suffering as the main explanatory variable. We ran several mediation models to test the role of

deservedness judgment in mediating the effect of moral evaluation of the sufferer and withholding of compassionate responses. Finally, we are interested in participants' normative judgments of withholding compassion on the basis of moral evaluation (i.e., their belief that withholding compassion on the basis of moral evaluation is morally good and required).

In the first two studies, we manipulated the moral status of the sufferer's intention (Study 1) and moral character (Study 2) and the causal role of the intention or the character in the suffering. Study 3 examined whether people would withhold compassion to the sufferers who share morally charged group membership, as an exploration of the boundary of people's tendencies to withhold compassion. Study 4 explored whether the effect of deservedness judgment on compassion might extend from a single individual to other group members to whom the moral evaluations may not even apply. As a synthesis, we carried out an internal meta-analysis to formally compare the effects of moral status of the sufferer (i.e., the Morality Effect) and its causal role in the suffering (i.e., Causality Effect) on compassion and willingness to help.

All studies reported in this paper were approved by the Human Subjects Committee of the first author's institution (Protocol number: 2-20-0206) and were pre-registered at [AsPredicted.org](https://aspredicted.org) (see links below). All the data and analysis reported in this paper can be accessed at the Open Science Framework account:

2. Study 1

Do observers withhold their compassion and willingness to help immoral sufferers, even when the immoral sufferers do not directly bring the suffering onto themselves? Existing studies (e.g., Pfattheicher et al., 2019) were not conclusive in this regard because the causal agency (caused by victim themselves vs. someone else) and the moral status of the agent's intention (harmful vs. neutral) were not teased apart. In a pilot study not reported here (see *Supplementary Materials: Supplementary Analyses and Results* for details), we explored this question by examining the situation in which suffering is caused by someone's intention, but orthogonalizing causal agency and the moral status of intention behind the harm inducing behaviors. The pilot study showed that when the sufferer's immoral intention causes their own suffering, observers would withhold compassion and willingness to help. However, in the pilot study, we only focused on cases in which someone's intention is causally related to suffering. Do people withhold compassion for a sufferer who has an immoral intention, even if the immoral intention does not directly cause the suffering? In Study 1, we created vignettes that orthogonalize the moral status of the victim's intention (immoral vs. neutral) and the causal role of that intention in suffering (causal vs. non-causal) to better understand their respective contribution to the observed effects on compassion and willingness to help.

2.1. Method

2.1.1. Participants

An a priori power analysis indicated that a sample of $N = 54$ would be needed to detect a medium effect size ($d_z = 0.5$) in a paired sample *t*-test with the power of 95%. We acknowledged that this and the following prior power analysis were conducted to obtain a rough estimate of required sample size. We preregistered a sample size of 65 participants to account for potential participant attrition due to failure in attention checks (https://aspredicted.org/CTD_IWM). Sixty-five participants self-reported as American residents and currently lived in the United States were recruited from Prolific. Among them, three were excluded due to failure in attention checks, leaving 62 participants in the final dataset (35 female; $M_{age} = 26.1$ years, $s.d. = 8.8$; Caucasian/White: 31%, Black/African: 34%, Hispanic/Latinx: 18%, Asian: 8%, Native American: 10%, Pacific Islander: 0%, Prefer not to answer: 0%, Others: 0%; 53% with a bachelor's or higher degree).

2.1.2. Experimental design and procedure

This study had a 2 (moral status of the sufferer's intention: immoral vs. neutral) \times 2 (causality of the sufferer's intention: causal vs. non-causal) within-participant design.⁴ Participants read four vignettes where the sufferer either has or does not have the intention to harm another person in the vignette. This factor was crossed with another factor indicating whether the sufferer is accidentally harmed by the action stemming from their prior intention, or by an event that is completely irrelevant to the sufferer's prior intention. Four different vignettes were thus created, each having four variations corresponding to each condition. Each participant saw each vignette exactly once, but in different conditions. The mapping between vignettes and conditions, and the presentation order of conditions were randomized across participants (please see the online *Supplemental Materials* for the materials). After reading each vignette, participants completed a number of self-reported questions assessing their emotional responses to and moral evaluations of the sufferer's suffering. After they completed all four conditions, the participants then completed individual differences measures, including personality questionnaires and demographic information. To make sure participants paid attention to the survey, we also included attention check questions, the answer of which were obvious (e.g., "We would like to make sure you are paying attention. Please select 2"). Participants would be excluded from data analysis if they failed any of these attention checks. As an example, one vignette starts with the following background information,

Asher and her supervisor are taking a tour of a chemical plant. When Asher goes over to the coffee machine to pour some coffee for the two of them, Asher's supervisor asks for some sugar in hers. There is white powder in a container by the coffee. The white powder is a very toxic substance left behind by a scientist and can cause serious nausea when ingested in any form.

Then in the Neutral Non-causal condition, the vignette continues,

The container is labeled "sugar", so Asher believes that the white powder by the coffee is sugar left out by the kitchen staff. Recalling her supervisor's requested sugar, Asher puts the substance in her supervisor's coffee to make it less bitter. However, on her way back Asher accidentally steps on a nail and falls on the floor feeling awful.

In the Immoral Non-causal condition, the vignette continues,

The container is labeled "toxic", so Asher believes that the white powder is a toxic substance left behind by a scientist. Hating her supervisor, Asher puts the substance in her supervisor's coffee to cause her pain. However, on her way back Asher accidentally steps on a nail and falls on the floor feeling awful.

In the Neutral Causal condition, the vignette continues,

The container is labeled "sugar", so Asher believes that the white powder by the coffee is sugar left out by the kitchen staff. Recalling her supervisor's requested sugar, Asher puts the substance in her supervisor's coffee to make it less bitter. However, when bringing the coffee back Asher accidentally takes her supervisor's cup, drinks the coffee and feels awful.

In the Immoral Causal condition, the vignette continues,

⁴ This and the following studies adopted within-participants design to maximize the sensitivity in detecting a true positive effect. To mitigate the potential issues with this design (e.g., carryover effect), for all our studies, we included multiple vignettes to fully counterbalance the mapping between experimental conditions and vignettes. As such, any one participant never saw the same vignettes more than once. In our regression analysis, we also included vignette version that the participants experienced as a random effect to further control its potential influence on our interested effects.

The container is labeled “toxic”, so Asher believes that the white powder is toxic substance left behind by a scientist. Hating her supervisor, Asher puts the substance in her supervisor’s coffee to cause her pain. However, when bringing the coffee back Asher accidentally takes her supervisor’s cup, drinks the coffee and feels awful.

2.1.3. Primary dependent variables and explanatory variables

Compassion and prosocial tendency. As our first primary dependent variable, self-reported compassion was measured by six related emotion terms (“compassion”, “sympathy”, “tenderness”, “moved”, “soft-hearted”, “pity”) adopted from previous work on compassion (e.g., Goetz et al., 2010; Goetz & Peng, 2019). For these measures, participants read the question “To what extent do you feel the following for [sufferer’s name] suffering?”, and then evaluated each of these emotion terms by moving a slider bar on a continuous scale ranging from 0 = *Not at all* to 100 = *Extremely*). The order of the emotion terms was randomized across participants.

Our second primary dependent variable, participants’ tendency to react prosocially to the sufferer’s suffering (referred to as “willingness to help” hereafter), was measured by two items: how likely the participants would be to offer sympathies to the sufferer, and to offer help as much as they can to the sufferer, on 7-point Likert scales (1 = *Not at all*, 7 = *Very likely*) (cf. Stellar et al., 2020). These measures were included in all the four studies.

Deservedness judgment. As our key explanatory variable, we measured participants’ judgment of how much the sufferer deserves the suffering. We measured this by using a single item, “To what extent do you think [sufferer’s name] deserves the suffering?” (1 = *not at all*, 7 = *very much*).

2.1.4. Supplemental variables

We also measured other emotion and control variables for exploratory purposes. The results of the analysis with these supplemental variables, for this and the following studies, are reported in the *Supplementary Materials* (Supplementary Tables 1–5).

Other emotion variables. We measured anxiety using the following items (“alarmed”, “upset”, “disturbed”, “distressed”, “worried”, “perturbed”) (Goetz et al., 2010). A few other emotion terms were also included as filler questions (e.g., “angry”, “disgust”, “afraid/scared”). These were also measured using a continuous scale as the primary emotion variables.

Moral evaluations of the suffering and the sufferer. As moral deservedness judgment has been theorized to be influenced by judgments of an agent’s behaviors and moral character (also referred to as “desert basis”; (Dennett and Caruso, 2021; Kagan, 2014; Feldman & Skow, 2020), we included additional measures to gauge these evaluations. Specifically, for the moral judgment of the sufferer’s behaviors, participants evaluated “Overall, how morally right or wrong was [sufferer’s name]’s decision?” on a 7-point Liker scale (−3 = *totally wrong*, 0 = *neither right nor wrong*, 3 = *totally right*). For the judgment of the sufferer’s moral character, participants answered “What is your general impression of [sufferer’s name]?” on a 7-point Liker scale (−3 = *very nasty*, 0 = *neither nasty nor nice*, 3 = *very nice*). The labels “nasty” and “nice” were adopted based on previous works on moral character inference (Siegel et al., 2017; Siegel, Mathys, Rutledge, & Crockett, 2018; Yu, Siegel, Clithero, & Crockett, 2021). We also asked the participants to evaluate some aspects of the suffering event, such as its likelihood to occur in reality and to happen to the participants themselves, and the severity of the suffering.

Manipulation check variables. To check whether our manipulation of causality was successful, we included a measure probing participants’ evaluation of the objective causality of the suffering (“To what extent do you think [sufferer’s name] caused her own suffering?”). We used this

objective causality measure as the check of the manipulation of causality and the moral judgment of the sufferer’s behaviors as the check of the manipulation of the moral status of the sufferer’s intention.

2.2. Results and discussion

2.2.1. Manipulation checks

In a linear mixed effect model, we included the self-reported objective causality as the dependent variable, causality (casual vs. non-causal), intention (immoral vs. neutral), and their interaction as key independent variables, and participant ID and vignette version as random intercept. Validating our manipulation of causality, we found that the main effect of causality was significant ($B = 2.79 \pm 0.31$, $t = 8.94$, $p < 0.001$). Interestingly, the main effect of intention was also significant ($B = 3.52 \pm 0.31$, $t = 11.26$, $p < 0.001$), consistent with the findings that moral evaluations of an agent tainted the perception of the agent’s intention and causation in an immoral behavior (Alicke, 1992, 2000; Alicke, Mandel, Hilton, Gerstenberg, & Lagnado, 2015). The interaction between causality and moral status of intention was also significant ($B = 1.68 \pm 0.44$, $t = 3.80$, $p < 0.001$), such that the causality manipulation induced a larger effect when the agent’s intention was immoral ($B = 2.79 \pm 0.33$, $t = 8.51$, $p < 0.001$) than when it was neutral ($B = 1.11 \pm 0.28$, $t = 4.03$, $p < 0.001$).

We then examined the validity of the manipulation of moral status of intention. Specifically, we looked at participants’ moral judgment of the agent’s behavior. In a linear mixed effect model similar to the above ones, we included the moral judgment score as the dependent variable. Validating our manipulation of the moral status of the sufferer’s intention, the main effect of moral status of intention on moral judgment score was significant ($B = 3.36 \pm 0.21$, $t = 15.96$, $p < 0.001$). Neither the main effect of causality, nor the interaction between causality and moral status of intention, was significant.

2.2.2. Compassion and willingness to help

To examine how causality and moral status of intention influence compassion and willingness to help, we estimated two linear mixed effects models, with the compassion score and willingness to help as dependent variables, respectively. We included causality (casual vs. non-causal), intention (immoral vs. neutral), and their interaction as key independent variables, and participant ID and vignette version as random intercept (Table 1; Supplementary Figure 1).

Both of the key independent variables had significant effects on compassion scores, such that participants exhibited less compassion when the sufferer had an immoral intention ($B = -48.79 \pm 2.85$, $t = -17.10$, $p < 0.001$), and when the sufferer themselves caused the suffering ($B = -7.28 \pm 2.85$, $t = -2.55$, $p = 0.012$). Crucially, the interaction between the key independent variables was significant ($B = -9.18 \pm 4.04$, $t = -2.27$, $p = 0.024$), such that the effect of causality was only significant when the sufferer had an immoral intention ($B = -7.28 \pm 1.73$, $t = -4.20$, $p < 0.001$), but not when the sufferer had a neutral intention ($B = 1.90 \pm 2.40$, $t = 0.79$, $p = 0.433$). In contrast, willingness to help was only sensitive to the moral status of intention ($B = -2.65 \pm 0.21$, $t = -12.44$, $p < 0.001$), such that participants indicated lower willingness to help when the sufferer had an immoral, relative to neutral, intention. Neither the main effect of causality nor the interaction was significant.

Given that there was no significant difference between the two Neutral intention conditions, we collapsed these two conditions so that the data structure would be more comparable to the following studies. We ran new regression models for compassion score and willingness to help, where we included the collapsed condition as the fixed effect predictor (with three levels: Immoral causal, Immoral non-causal, Morally neutral). Random intercepts were the same as the above

Table 1
Results of hypothesis-driven regression analysis (Study 1).

Predictors	Compassion				Willingness to help				Deservedness						
	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p
Intercept	32.70 ± 1.83					4.77 ± 0.14					3.77 ± 0.14				
Neutral > Immoral Noncausal	40.56 ± 2.47	1.36	35.70–45.42	16.44	<0.001	2.37 ± 0.18	1.22	2.01–2.73	12.86	<0.001	-3.40 ± 0.21	-1.40	-3.82 to -2.99	-16.03	<0.001
Immoral Noncausal > Immoral Causal	7.28 ± 2.85	0.24	1.66–12.89	2.55	0.011	0.24 ± 0.21	0.12	-0.18–0.66	1.14	0.257	-0.58 ± 0.25	-0.24	-1.06 to -0.10	-2.37	0.019
Random Effects						1.40									
σ^2							Participant:0.82								
τ^2							Vignette:0.00								
N							Participant:62								
							Vignette:23								
								Participant:64							
								Vignette:0.00							
								Participant:62							
								Vignette:23							

models. This regression model was adopted for this and the subsequent studies for better comparison across studies.⁵ Compassion score was the lowest in the Immoral causal condition (14.33 ± 2.13), which was significantly lower than that in the Immoral non-causal condition (21.61 ± 2.62 ; $B = 7.28 \pm 2.85$, $t = 2.55$, $p = 0.011$), which in turn was significantly lower than that in the Morally neutral condition (62.17 ± 1.85 ; $B = 40.56 \pm 2.47$, $t = 16.44$, $p < 0.001$) (Fig. 2A). Willingness to help was similarly low in the two Immoral conditions (3.82 ± 0.24 for the Immoral causal condition, 4.06 ± 0.23 for the Immoral non-causal condition; $B = 0.24 \pm 0.21$, $t = 1.14$, $p = 0.257$), but was significantly higher in the Morally neutral condition (6.44 ± 0.09 ; $B = 2.37 \pm 0.18$, $t = 12.86$, $p < 0.001$) (Fig. 2B).

2.2.3. Deservedness of suffering

We used a similar single predictor model as the one described in the above section to examine how participants' deservedness judgment of the sufferer's suffering was modulated by the sufferer's intention and the causal role of the intention in the suffering (Fig. 2C). The participants judged the sufferer as more deserving of the suffering when the sufferer's immoral intention caused the suffering (5.29 ± 0.24) than when the sufferer's immoral intention was irrelevant to their suffering (4.71 ± 0.28 ; $B = 0.58$, $t = 2.37$, $p = 0.019$). The participants judged the sufferer as the least deserving of the suffering when the sufferer's intention is morally neutral (1.31 ± 0.08 ; $B = -3.40 \pm 0.21$, $t = -16.03$, $p < 0.001$, relative to the Immoral non-causal condition).

These results suggest that the moral status of the sufferer's intention matters in observers' compassion and willingness to help. By orthogonalizing the moral status of the sufferer's intention and the causal role of the moral status in the suffering, these results further demonstrate that observers consider how the suffering is brought about in their evaluation of the suffering. Specifically, when the suffering is a direct consequence of the sufferer's immoral intention, the observers express the least compassion. But importantly, when the sufferer has an immoral intention, even if the intention was not causally related to the suffering, observers still withhold their compassion. The results thus go beyond existing findings in the literature (e.g., Pfattheicher et al., 2019) by suggesting that sufferer's moral badness per se, independent from its causal role in the suffering, leads observers to believe that the sufferer deserves the suffering and to reduce their compassionate responses. These results lead to the question of how generalized the effect of moral deservedness is. The key moral appraisal in Study 1 is what the sufferer *does* (or intends to do). In the next three studies, we extended that to what the sufferer *is*, namely, the sufferer's morally relevant character traits and group identities.

3. Study 2

Study 1 showed that both the moral status and causal role of the sufferer's prior intention in their suffering mattered for observers' expressed compassion. In Study 2, we examined whether this effect would expand to the more stable moral character of the sufferer. To this end, we created vignettes that orthogonalize the sufferer's moral character and its causal relevance to their later suffering or bad experiences. As a secondary goal, we were interested in the individual differences in the effect of moral character on compassion. Specifically, we hypothesized that individuals who tend to think of moral traits as biological essence would be more likely to base their compassion on the sufferer's past moral character, and therefore would show a larger effect of moral character on compassion.

⁵ We admitted that this analysis might deviate from some of our pre-registered analyses. For transparency, we reported the results based on the pre-registered models in the *Supplementary Materials* (Supplementary Tables 1–5).

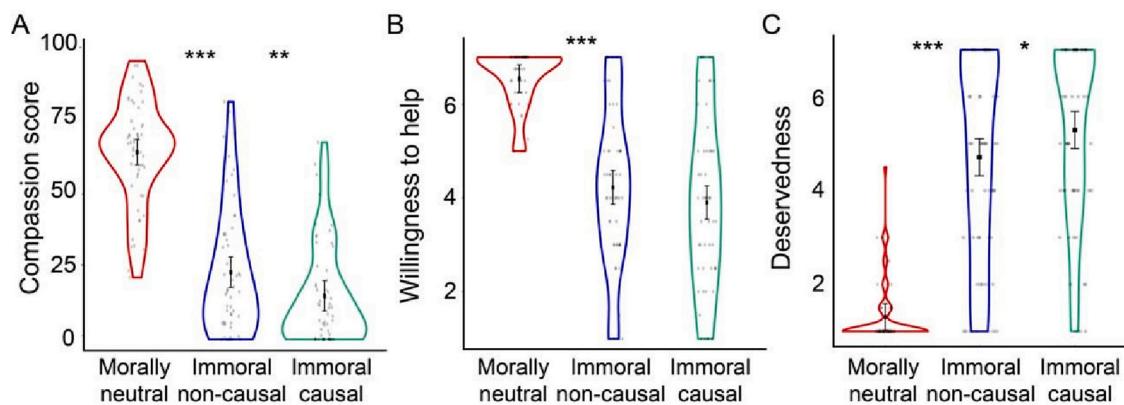


Fig. 2. Effects of condition on compassion (A), willingness to help (B), and deservedness (C) in Study 1.

3.1. Method

3.1.1. Participants

We preregistered to detect a small-to-medium effect size ($f^2 = 0.1$) for the coefficient of moral essentialism score (see below for detail) in the regression model that predicts the effect of moral character on compassion (<https://aspredicted.org/HHU.SJZ>). To detect such an effect at $p < 0.05$ level with 90% power, we need a total sample of at least 108 participants. One hundred and twenty U.S. participants were recruited from Prolific. Among them, ten were excluded due to failure in attention checks, leaving 110 participants in the final dataset (62 female; $M_{age} = 32.4$ years, $s.d. = 11.5$; Caucasian/White: 69%, Black/African: 5%, Hispanic/Latinx: 12%, Asian: 12%, Native American: 0%, Pacific Islander: 0%, Prefer not to answer: 2%, Others: 1%, 53% with a bachelor's or higher degree).

3.1.2. Experimental design and procedure

This study had a 3-level within-participant design. Participants read three vignettes depicting a sufferer and their suffering. The sufferer either has a morally bad character (e.g., being a racist) or a morally neutral trait (e.g., being a football fan). The immoral case was further divided into two conditions. In the Immoral causal condition, the sufferer's morally bad character directly causes their suffering or bad experiences (e.g., being fired due to racist behaviors); in the Immoral non-causal condition, the sufferer's morally bad character is irrelevant to their suffering (e.g., being laid off due to economic recession). In the Morally neutral condition, the sufferer's suffering event is the same as in the Immoral non-causal condition. Thus, the first two conditions only differ in terms of the causal role of the sufferer's morally bad character in their suffering, whereas the latter two conditions only differ in terms of the sufferer's moral character.

Three different vignettes were created, each having three variations corresponding to each condition (please see the online *Supplemental Materials* for the materials). Each participant saw each vignette exactly once, but in different conditions. The mapping between vignettes and conditions, and the presentation order of conditions were randomized across participants. The procedure was the same as that of Study 1. As an example, the Immoral Non-causal version of a vignette reads,

Tyler used to work for an accounting firm. At that time, he was a racist person and constantly ridiculed and threatened his co-workers who have a minority background. Due to economic hardship, the firm Tyler had been working for went bankrupt and Tyler has become unemployed since then.

The Immoral Casual version of the vignette reads,

Tyler used to work for an accounting firm. At that time, he was a racist person and constantly ridiculed and threatened his co-workers who have a minority background. Due to his racist behaviors and

attitudes, Tyler was fired by the firm and has become unemployed since then.

The Morally neutral vignette reads,

Adan owned a small barber shop and had served the local residents for years. After work, he liked hiking and surfing, and is a big fan of the local football team. Due to the COVID-19 pandemic, Adan had to close his barber shop for a long time, and even when it was allowed to reopen, few customers came. Adan had no choice but to close his barber shop permanently.

3.1.3. Primary dependent variables, explanatory variables, and control variables

Measures of compassion, willingness to help, deservedness of the suffering, moral evaluations of the suffering and the sufferer, and other control variables were the same as in Study 1.

Manipulation check variables. As a check for the manipulation of moral character, we examined whether the manipulation had a significant effect on the perceived moral character of the sufferer (see *Moral evaluations of the Suffering and the Sufferer* in Study 1). Similar to Study 1, we included a manipulation check item regarding the manipulation of causality, “To what extent do you think [sufferer’s name]’s moral character led to [sufferer’s name]’s suffering?” (1 = *Not at all*, 4 = *Moderately*, 7 = *Very much*).

Perceived normativity of compassion. To dissociate the effect of our manipulation on participants' actual feelings of compassion versus their ‘ideal affect’, namely, what they believe *ought to be felt* in a given situation, we assessed participants' perceived normativity regarding compassion in each condition. We included two items to measure this construct, “To what extent do you think a morally good person will feel compassion and sympathy to [sufferer’s name]?” and “To what extent do you think people should be compassionate and sympathetic to [sufferer’s name]?”

3.1.4. Individual difference measures

Moral Essentialism was measured using a questionnaire adapted from Gelman and colleagues (Gelman, Heyman, & Legare, 2007). Specifically, we asked the participants to imagine that a person named Asher is very morally bad. Then we asked the participants to evaluate, on 7-point Likert scales, their attitudes toward the following statements, “To what extent do you agree that Asher was born morally bad?”, “To what extent do you agree that Asher’s brain is different from someone who is not morally bad?”, “To what extent do you agree that Asher can change whether or not he is morally bad, if he wants to?” (reverse coding), and “To what extent do you agree that Asher’s moral badness is due to things that people around him did?” (reverse coding). The scores on these items were averaged to form a composite score of moral essentialism.

3.2. Results and discussion

3.2.1. Manipulation checks

The impression of the sufferer's moral character was significantly nicer in the Morally neutral condition (5.92 ± 0.11) than in the two Immoral conditions (Immoral non-causal: 1.83 ± 0.09 , $B = 4.09 \pm 0.13$, $t = 31.93$, $p < 0.001$; Immoral causal: 1.53 ± 0.08 , $B = 4.39 \pm 0.13$, $t = 34.27$, $p < 0.001$). As for causality, participants agreed more in the Immoral causal condition that the sufferer's suffering was caused by their moral character (6.70 ± 0.07) than in the Immoral non-causal condition (3.36 ± 0.20 , $B = 3.34 \pm 0.18$, $t = 18.62$, $p < 0.001$). These results confirmed that our manipulations of casualty and moral character were successful.

3.2.2. Compassion and willingness to help

As in Study 1, we estimated two linear mixed effect models to examine how self-reported compassion and willingness to help were modulated by the sufferer's moral character and the causal relevance of the moral character in the sufferer's suffering (Table 2). The participants reported highest compassion to the Morally neutral sufferer (72.7 ± 2.0), less so when the sufferer was described as a morally bad person who suffered from a cause that was unrelated to their moral badness (14.9 ± 1.8 ; $B = -57.75 \pm 2.11$, $t = -27.40$, $p < 0.001$, relative to the Morally neutral condition). Interestingly, participants expressed even less compassion to a morally bad sufferer if the sufferer's suffering was caused by their moral badness (8.3 ± 1.2 ; $B = -6.60 \pm 2.11$, $t = -3.13$, $p = 0.002$, relative to the Immoral non-causal condition) (Fig. 3A). Willingness to help exhibited a similar pattern (Fig. 3B). Specifically, the participants were most willing to help the sufferer in the Morally neutral condition (5.98 ± 0.11), less so in the Immoral non-causal condition (2.01 ± 0.11 ; $B = -3.97 \pm 0.14$, $t = -28.62$, $p < 0.001$, relative to the Morally neutral condition). Participants were even less willing to help the sufferer when the sufferer's suffering was caused by their moral badness (1.57 ± 0.10 ; $B = -0.44 \pm 0.14$, $t = -3.18$, $p = 0.002$, relative to the Immoral non-causal condition).

3.2.3. Deservedness of suffering

We used a similar linear mixed effect model as the ones described above to examine how participants' judgment of the deservedness of the sufferer's suffering was modulated by the sufferer's moral character and the causal relevance of the moral character in the suffering (Fig. 3C). The participants judged the sufferer as more deserving of the suffering when the sufferer's moral badness caused the suffering (6.53 ± 0.09) than when the sufferer's moral badness was irrelevant to their suffering (5.10 ± 0.17 ; $B = -1.43 \pm 0.16$, $t = -8.92$, $p < 0.001$). The participants judged the sufferer as least deserving of the suffering in the Morally neutral condition (1.18 ± 0.06 ; $B = 3.92 \pm 0.16$, $t = 24.50$, $p < 0.001$, relative to the Immoral non-causal condition).

3.2.4. Normativity of feeling compassion

Internal reliability of the normativity items ranged from 0.76 to 0.92. We therefore combined the scores into a composite score of normativity. Participants judged that it was least normatively required to feel compassion and sympathy in the Immoral causal condition (2.00 ± 0.11). Normativity judgment in the Immoral causal condition was significantly lower than that in the Immoral non-causal condition (2.61 ± 0.12 ; $B = -0.61 \pm 0.15$, $t = -4.23$, $p < 0.001$), which was in turn significantly lower than that in the Morally neutral condition (6.35 ± 0.10 ; $B = -3.74 \pm 0.15$, $t = 25.79$, $p < 0.001$).

3.2.5. Individual differences analyses

Not supporting our hypothesis, Moral Essentialism was not significantly associated with the effect of moral character on either the compassion score ($B = 1.66 \pm 2.85$, $t = 0.58$, $p = 0.561$) or the willingness to help score ($B = 0.19 \pm 0.18$, $t = 1.08$, $p = 0.282$).

The findings of Study 2 demonstrated that a sufferer's morally bad

Preditors	Compassion						Willingness to help						Deservedness					
	$B \pm SE$	b	$95\% CI of B$	t	p		$B \pm SE$	b	$95\% CI of B$	t	p		$B \pm SE$	b	$95\% CI of B$	t	p	
Intercept	31.97 ± 1.18					3.19 ± 0.10							4.27 ± 0.07					
Neutral > Immoral Noncausal	57.75 ± 2.11	1.70	53.60–61.89	27.40	<0.001	3.97 ± 0.14	1.74	3.70–4.24	28.62	<0.001	-3.92 ± 0.16	-1.53	-4.23 to -3.60	-24.50	<0.001	1.41	Participant:0.04	
Immoral Noncausal > Immoral Causal	6.60 ± 2.11	0.19	2.45–10.74	3.13	0.002	0.44 ± 0.14	0.19	0.17–0.71	3.18	0.002	-1.43 ± 0.16	-0.56	-1.74 to -1.11	-8.92	<0.001	Vignette:2	Participant:110	
Random Effects	244.23					1.06												
σ^2																		
t_{00}																		
N																		

Table 2
Results of hypothesis-driven regression analysis (Study 2).

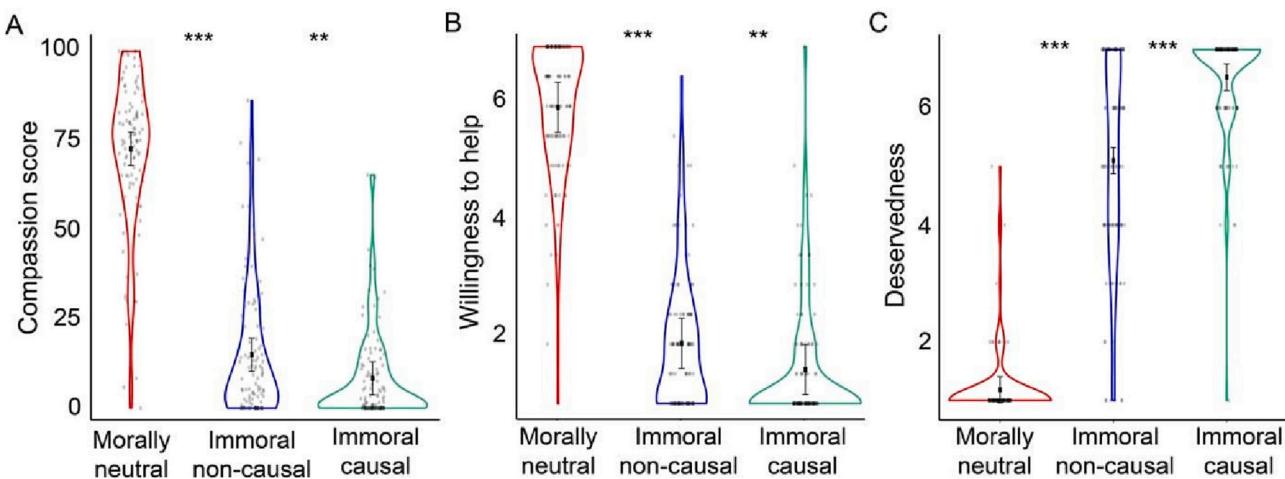


Fig. 3. Effects of condition on compassion (A), willingness to help (B), and deservedness judgment (C) in Study 2.

character, even when it did not cause the suffering, rendered the suffering as more deserved in the eyes of observers and reduced observers' compassion and willingness to help. This extends the attribution theory's action-based appraisal (e.g., whether the sufferer is responsible for their suffering; (Rudolph et al., 2004)) to person-centered appraisal (e.g., whether the sufferer's global character traits justify their suffering; (Knobe, 2010; Uhlmann, Pizarro, & Diermeier, 2015)). The participants not only exhibited such emotion and motivation patterns themselves, but they also believed that such responses were morally good. In other words, the participants did not believe that withholding compassion and help on the basis of the sufferer's moral badness was a moral weakness or failure. This may pose challenges for interventions aiming to expand compassion to those who bear the label of "morally bad people", regardless of whether such labeling is appropriate (see General Discussion).

4. Study 3a and Study 3b

Studies 1 and 2 demonstrated that moral intentions and characters influence perceived deservedness of suffering and withholding of compassion. In Study 3, we investigated whether and how a sufferer's morally relevant *identity* would function as a moral attribute that affects observers' compassion and willingness to help. To that end, we focused on a morally relevant group membership, namely political (i.e., party) affiliation in the United States. Outgroup membership in general, and political divide specifically, has been shown to be a "dampening" factor of empathy (Bruneau, Cikara, & Saxe, 2017; Cikara et al., 2011; Hudson, Cikara, & Sidanius, 2019; O'Brien & Ellsworth, 2012) and its underlying neurocognitive processes (Avenanti, Sirigu, & Aglioti, 2010; Azevedo et al., 2013; Xu, Zuo, Wang, & Han, 2009). Some researchers reject empathy as a proper moral emotion and driver of prosocial behaviors on these bases and propose rational compassion instead (Bloom, 2017a, 2017b; P. Singer, 2015). Here, we explored whether compassion is constrained by another attribute of the suffering, namely, deservedness judgment of the suffering based on sufferer's group membership or identity. Specifically, in Study 3a, we recruited politically left participants (self-identified Democrats from Prolific) while in Study 3b, we recruited politically right participants (self-identified Republicans from Prolific).

4.1. Method

4.1.1. Participants

For Study 3a, we preregistered a small-to-medium effect size ($\eta^2 = 0.2$) for the comparison of the compassion scores between the two key conditions based on the results of Study 2 (Outgroup relevant vs.

Outgroup irrelevant; see below for detail; https://aspredicted.org/RKX_XRM). Power analysis indicated that a sample of $N = 80$ would achieve 95% power to detect this effect at $p < 0.05$ level. Eighty-five participants who self-reported as American residents, currently living in the United States, and affiliated with the Democratic Party were recruited from Prolific. Three of them were excluded due to failure in attention checks. Another one was excluded because their self-reported political affiliation was inconsistent with Prolific's demographic filter, leaving 81 participants in the final dataset (50 female; $M_{age} = 31.4$ years, $s.d. = 11.7$; Caucasian/White: 66%, Black/African: 7%, Hispanic/Latinx: 6%, Asian: 19%, Native American: 0%, Pacific Islander: 0%, Prefer not to answer: 0%, Others: 1%; 52% with a bachelor's or higher degree). Data collection took place on December the 3rd, 2020, when it was certain that the Democratic candidate, Mr. Joseph. R. Biden won the 59th U.S. presidential election.

For Study 3b, we preregistered a sample size of $N = 80$ to be consistent with Study 3a (https://aspredicted.org/9LR_84F). Ninety-eight participants who self-reported as American residents, currently living in the United States, and affiliated with the Republican Party were recruited from Prolific. Four of them were excluded due to failure in attention checks. Another twelve were excluded because their self-reported political affiliation was inconsistent with Prolific's demographic filter, leaving 82 participants in the final dataset (46 female; $M_{age} = 39.4$ years, $s.d. = 12.3$; Caucasian/White: 87%, Black/African: 0%, Hispanic/Latinx: 6%, Asian: 4%, Native American: 1%, Pacific Islander: 0%, Prefer not to answer: 0%, Others: 1%; 61% with a bachelor's or higher degree). Data collection took place on the January the 21st, 2022.

4.1.2. Experimental design and procedure

This study had a 3-level within-participant design. Participants read three vignettes depicting a sufferer and their suffering. The sufferer is either a member of the political rivalry group ("outgroup" hereafter) relative to the participants (i.e., Trump supporter for Study 4a, Biden supporter for Study 4b) or a member of a political neutral group (i.e., supporter of the local baseball team; "neutral group" hereafter). The outgroup member case was further divided into two conditions. In the Outgroup relevant condition, the sufferer's political affiliation is causally relevant to their suffering (e.g., contracting COVID from a non-socially distanced gathering in support of Trump); in the Outgroup irrelevant condition, the sufferer's political affiliation is irrelevant to their suffering (e.g., contracting COVID when doing grocery shopping). In the Neutral group condition, the sufferer's suffering event is the same as in the Outgroup irrelevant condition. Therefore, the first two conditions only differed in terms of the causal role of the sufferer's political affiliation in their suffering, whereas the latter two conditions only differed in terms

of the sufferer's political affiliation relative to the participants.

Three different vignettes were thus created, each had three variations corresponding to each condition (please see the online *Supplemental Materials* for the materials). Each participant saw each vignette exactly once, but in different conditions. The mapping between vignettes and conditions and the presentation order of conditions were randomized across participants. The procedure was the same as that of Study 2. As an example, the Outgroup Relevant version of a vignette in Study 3a reads,

Tyler supported Donald Trump in the recent presidential election. He recently voted at a local poll station for Trump. The poll station was packed and Tyler got COVID from someone else at the poll station. Though not life threatening, the COVID-related symptoms have made Tyler feel pretty awful.

The Outgroup Irrelevant version of the vignette reads,

Tyler supported Donald Trump in the recent presidential election. He recently visited a local grocery store to buy food. The grocery store was packed and Tyler got COVID from someone else at the grocery store. Though not life threatening, the COVID-related symptoms have made Tyler feel pretty awful.

The Neutral Group version of the vignette reads,

Tyler is a supporter of the local baseball team. He recently visited a local grocery store to buy food. The grocery store was packed and Tyler got COVID from someone else at the grocery store. Though not life threatening, the COVID-related symptoms have made Tyler feel pretty awful.

For Study 3b, the Outgroup Relevant version of a vignette reads,

Adan supported Joe Biden in the recent presidential election. Earlier that year, he joined a large in-person Black Lives Matter demonstration, condemning the murder of George Floyd and police brutality. The event went violent and Adan got physically injured. Though not life threatening, the injury has made Adan feel pretty awful.

The Outgroup Irrelevant version of the vignette reads,

Adan supported Joe Biden in the recent presidential election. Earlier that year, he watched a football game between the local team and a rivalry team. After the game, Adan was caught in a violent fight between fans of the two teams and got physically injured. Though not life threatening, the injury has made Adan feel pretty awful.

The Neutral Group version of the vignette reads,

Adan is a supporter of the local baseball team. Recently, he watched a baseball game between the local team and a rivalry team. After the game, Adan was caught in a violent fight between fans of the two teams and got physically injured. Though not life threatening, the injury has made Adan feel pretty awful.

4.1.3. Primary dependent variables, explanatory variables, and control variables

Measures of compassion, willingness to help, deservedness of suffering, and other exploratory and control variables were the same as in Study 2.

Perceived normativity of compassion. Because the scenarios used in this study had an intergroup aspect, we slightly modified the items for the perceived normativity of compassion. Specifically, we used the following items “To what extent do you think a supporter for Joe Biden/Donald Trump will feel compassion and sympathy toward [sufferer's name]?” (1 = *Definitely will not*, 4 = *Maybe*, 7 = *Definitely will*), “To what extent do you think supporters for Joe Biden/Donald Trump should be compassionate and sympathetic to [sufferer's name]?” (1 = *Definitely should not*, 4 = *Maybe*, 7 = *Definitely should*), and “To what extent do you

think a morally good person will feel compassion and sympathy to [sufferer's name]?” (1 = *Definitely will not*, 4 = *Maybe*, 7 = *Definitely will*).

Manipulation check variables. As a check for the manipulation of group membership, we asked participants two questions, “Overall, how would you describe [sufferer's name]'s political ideology?” (1 = *Extremely liberal*, 4 = *Middle of the road*, 7 = *Extremely conservative*) and “To what extent [sufferer's name]'s political ideology is similar to yours?” (1 = *Not at all similar*, 4 = *Somewhat similar*, 7 = *Very much similar*). We also included two questions about the participants' own political ideology in the demographic questionnaire, “How would you describe your political ideology?” (1 = *Extremely liberal*, 4 = *Middle of the road*, 7 = *Extremely conservative*) and “Generally speaking, which party do you support?”. Perceived closeness with the sufferer was measured using the Inclusion of Other in the Self (IOS) Scale (1 = *minimally overlapping*, 5 = *maximally overlapping*). As a check for the manipulation of causality, we asked the participants “To what extent do you think [sufferer's name]'s political ideology led to his illness?” (1 = *Not at all*, 4 = *Moderately*, 7 = *Very much*).

4.2. Results and discussion

4.2.1. Manipulation checks

We first examined whether our group manipulation was successful. For Study 3a, we recruited U.S. participants who self-identified as Democrats, while for Study 3b, we recruited U.S. participants who self-identified as Republicans. On average, the participants in Study 4a reported that they were quite liberal in terms of their political ideology (1.84 ± 0.11 , on a scale ranging from 1 = *Extremely liberal* to 7 = *Extremely conservative*) and that their political conviction was strong (5.68 ± 0.16 , on a scale ranging from 1 = *Rather weak conviction* to 7 = *Very strong conviction*). In contrast, the participants in Study 4b reported that they were quite conservative in terms of their political ideology (5.73 ± 0.09) and that their political conviction was strong (5.28 ± 0.15). These indicated that for both studies the participants' political affiliation was as we expected.

We next examined whether the participants perceived the political identity of the sufferers correctly (Table 3). For Study 3a, when evaluating the political ideology of the sufferers in the vignettes, the participants judged the sufferer in the “Outgroup relevant” condition (i.e., Trump-supporter who engaged in activities aiming to support Trump's reelection) to be most conservative (6.54 ± 0.09), followed by the “Outgroup irrelevant” condition (i.e., Trump-supporter who engaged in activities irrelevant to politics; 5.99 ± 0.13 , $B = -0.53 \pm 0.12$, $t = -4.55$, $p < 0.001$, relative to the Outgroup relevant condition), and the Neutral group condition (4.01 ± 0.08 , $B = -2.02 \pm 0.12$, $t = -17.37$, $p < 0.001$, relative to the Out-group irrelevant condition). The participants judged the sufferers in the Outgroup relevant condition and in the Outgroup irrelevant condition to be not at all similar to themselves in terms of political ideology (1.20 ± 0.06 for the Out-group relevant condition; 1.30 ± 0.08 for the Outgroup irrelevant condition; the difference between the two conditions was not significant, $B = 0.14 \pm 0.12$, $t = 1.09$, $p = 0.277$). The participants judged the sufferer in the Neutral group condition to be “somewhat similar” to them in terms of political ideology (3.40 ± 0.13 , $B = 2.10 \pm 0.12$, $t = 16.86$, $p < 0.001$, relative to the Outgroup irrelevant condition). The sufferer in the Outgroup relevant condition was least close to the participants (1.37 ± 0.07), followed by the Outgroup irrelevant condition (1.62 ± 0.09 , $B = 0.23 \pm 0.09$, $t = 2.51$, $p = 0.013$, relative to the Outgroup relevant condition) and the Neutral group condition (2.38 ± 0.11 , $B = 0.79 \pm 0.09$, $t = 8.45$, $p < 0.001$, relative to the Outgroup irrelevant condition). These results indicated that the participants perceived the sufferers in the outgroup conditions (i.e., Trump supporters) as more conservative, less similar to themselves, and less close.

For Study 3b, when evaluating the political ideology of the sufferers in the vignettes, the participants judged the sufferer in the “Outgroup relevant” condition (i.e., Biden-supporter who engaged in activities

Table 3

Perception of the sufferer's political ideology as a function of experimental condition.

Sufferer political ideology (1 = Extremely liberal, 7 = Extremely conservative)			Similarity to the self (1 = Not at all similar, 7 = Extremely similar)			Perceived closeness (1 = least close, 5 = very close)		
Neutral group	Outgroup irrelevant	Outgroup relevant	Neutral group	Outgroup irrelevant	Outgroup relevant	Neutral group	Outgroup irrelevant	Outgroup relevant
<i>Study 3a</i>								
4.01 ± 0.08	5.99 ± 0.13	6.54 ± 0.09	3.40 ± 0.13	1.30 ± 0.08	1.20 ± 0.06	2.38 ± 0.11	1.62 ± 0.09	1.37 ± 0.07
<i>Study 3b</i>								
4.11 ± 0.06	2.28 ± 0.11	1.77 ± 0.09	3.68 ± 0.12	1.85 ± 0.13	1.60 ± 0.10	2.11 ± 0.11	1.77 ± 0.11	1.62 ± 0.10

aiming to support Biden's campaign) to be most liberal (1.77 ± 0.09), followed by the "Outgroup irrelevant" condition (i.e., Biden-supporter who engaged in activities irrelevant to politics; 2.28 ± 0.11 , $B = 0.51 \pm 0.11$, $t = 4.50$, $p < 0.001$, relative to the Outgroup relevant condition), and the Neutral group condition (4.11 ± 0.06 , $B = 1.83 \pm 0.11$, $t = 16.06$, $p < 0.001$, relative to the Out-group irrelevant condition). The participants judged the sufferers in the Outgroup relevant condition and in the Outgroup irrelevant condition to be not at all similar to themselves in terms of political ideology (1.60 ± 0.10 for the Out-group relevant condition; 1.85 ± 0.13 for the Outgroup irrelevant condition; difference between the two conditions was marginally significant, $B = 0.26 \pm 0.14$, $t = 1.84$, $p = 0.068$). The participants judged the sufferer in the Neutral group condition to be "somewhat similar" to them in terms of political ideology (3.68 ± 0.12 , $B = 1.83 \pm 0.14$, $t = 13.12$, $p < 0.001$, relative to the Outgroup irrelevant condition). The sufferer in the Outgroup relevant condition was perceived as least close to the participants themselves (1.62 ± 0.10), followed by the Outgroup irrelevant condition (1.77 ± 0.11 ; the difference between the two conditions was not significant, $B = 0.15 \pm 0.09$, $t = 1.53$, $p = 0.129$) and the Neutral group condition (2.11 ± 0.11 , $B = 0.34 \pm 0.10$, $t = 3.56$, $p < 0.001$, relative to the Outgroup irrelevant condition). These results indicated that the participants perceived the sufferers in the outgroup conditions (i.e., Biden-supporters) as more liberal, less similar to themselves, and less close.

In Studies 3a and 3b, we manipulated the causal relevance of the sufferer's group membership or identity to their suffering. While in both of the Outgroup conditions, the sufferer's behavior leads to their suffering, they differ with regard to how diagnostic of the outgroup membership or identity the behavior is. In other words, the two conditions differ in terms of the causal relevance of their group identity in their suffering. Specifically, we expected that in the Outgroup relevant condition participants would perceive the sufferer's outgroup membership or identity to be more responsible for the sufferer's suffering. This is exactly what we found: the participants more strongly agreed that the sufferer's political ideology led to their suffering in the Outgroup relevant condition (Study 3a: 4.53 ± 0.27 ; Study 3b: 2.88 ± 0.24) than in the Outgroup irrelevant condition (Study 3a: 2.47 ± 0.22 , $B = 2.06 \pm 0.27$, $t = 7.68$, $p < 0.001$; Study 3b: 1.34 ± 0.11 , $B = 1.54 \pm 0.22$, $t = 7.09$, $p < 0.001$).

4.2.2. Compassion and willingness to help

To examine how group identity and its causal role in the suffering influence compassion and willingness to help, we estimated two linear mixed effects models, with the composite scores of compassion and willingness to help as dependent variables, respectively. We included condition (Outgroup relevant, Outgroup irrelevant, Neutral group) as the key independent variable, and participant ID and vignette version as random intercepts (for details, see Tables 4 and 5).

As Fig. 4A shows, in Study 3a, participants reported the highest compassion to the sufferer in the Neutral group condition (56.45 ± 2.53), followed by the Outgroup irrelevant condition (40.57 ± 2.83 , $B = -15.89 \pm 2.74$, $t = -5.80$, $p < 0.001$, relative to the Neutral group condition). Participants reported the lowest compassion to the sufferer

in the Outgroup relevant condition (31.90 ± 2.87 , $B = -24.55 \pm 2.74$, $t = -8.96$, $p < 0.001$, relative to the Outgroup irrelevant condition). Willingness to help exhibited a similar pattern (Fig. 4B). Specifically, participants were most willing to help the sufferer in the Neutral group condition (5.33 ± 0.17), followed by the Outgroup irrelevant condition (4.05 ± 0.20 , $B = -1.28 \pm 0.20$, $t = -6.56$, $p < 0.001$, relative to the Neutral group condition). Participants were least willing to help the sufferer in the Outgroup relevant condition (3.33 ± 0.22 , $B = -0.72 \pm 0.20$, $t = -3.69$, $p < 0.001$, relative to the Outgroup irrelevant condition).

In Study 3b, the patterns were slightly different. Participants reported similar levels of compassion to the sufferer in the Neutral group condition (49.98 ± 2.65) and the Outgroup irrelevant condition (49.50 ± 2.94 , $B = 0.48 \pm 2.86$, $t = 0.17$, $p = 0.866$, relative to the Neutral group condition; Fig. 4D). Participants reported the lowest compassion to the sufferer in the Outgroup relevant condition (40.60 ± 3.31 , $B = -8.90 \pm 2.86$, $t = -3.11$, $p = 0.002$, relative to the Outgroup irrelevant condition). Willingness to help exhibited a similar pattern (Fig. 4E). Specifically, participants were equally willing to help the sufferer in the Neutral group condition (4.78 ± 0.18) and the Outgroup irrelevant condition (4.66 ± 0.20 , $B = 0.12 \pm 0.19$, $t = 0.64$, $p = 0.524$, relative to the Neutral group condition). Participants were least willing to help the sufferer in the Outgroup relevant condition (4.04 ± 0.21 , $B = -0.62 \pm 0.19$, $t = -3.25$, $p = 0.001$, relative to the Outgroup irrelevant condition).

As an exploratory analysis, we combined the data from the two studies and examined if there was: 1) a main effect of Morality (Outgroup irrelevant vs. Neutral group), 2) a main effect of Relevance (Outgroup relevant vs. Outgroup irrelevant), and 3) modulation of participant group (Democrats vs. Republicans) on the Morality and Relevance effects. To this end, we ran a new set of regression models that included data from both studies. We included the main effect of condition, the main effect of participant groups, and the condition by group interaction as fixed effect predictors, and participant ID, vignette version, and participant group as random intercepts. For compassion, the main effect of experimental condition was significant ($F = 36.74$, $p < 0.001$; Outgroup irrelevant vs. Neutral group: $B = -16.17 \pm 2.84$, $t = -5.70$, $p < 0.001$; Outgroup relevant vs. Outgroup irrelevant: $B = -8.71 \pm 2.84$, $t = -3.07$, $p = 0.002$). Critically, the condition by group interaction was also significant ($F = 10.15$, $p < 0.001$). Specifically, the participants affiliated with the Republican party exhibited a significantly smaller Morality effect than the participants affiliated with the Democratic party ($B = -15.69 \pm 4.00$, $t = -3.92$, $p < 0.001$), but showed a similar Causality effect compared with the participants affiliated with the Democratic party ($B = 0.19 \pm 4.00$, $t = 0.05$, $p = 0.963$).

4.2.3. Deservedness of suffering

For Study 3a (Fig. 4C; Table 4), the participants judged the sufferer as deserving of the suffering more when the sufferer's outgroup identity was causally relevant to their suffering (3.06 ± 0.24) than when the sufferer's outgroup identity was irrelevant to their suffering (2.08 ± 0.19 , $B = 0.98 \pm 0.22$, $t = 4.37$, $p < 0.001$, relative to the Outgroup relevant condition). The participants judged the sufferer as least

Table 4
Results of hypothesis-driven regression analysis (Study 3a).

Predictors	Compassion				Willingness to help				Deservedness						
	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p
Intercept	43.25 ± 3.24					4.24 ± 0.16					2.14 ± 0.15				
Neutral group > Outgroup irrelevant	15.89 ± 2.74	0.59	10.49–21.29	5.80	<.0001	1.28 ± 0.20	0.65	0.90–1.67	6.56	<.0001	-0.83 ± 0.22	-0.46	-1.27 to -0.39	-3.73	<.0001
Outgroup irrelevant > Outgroup relevant	24.55 ± 2.74	0.32	19.16–29.95	8.96	<.0001	0.72 ± 0.20	0.36	0.34–1.11	3.69	<.0001	-0.98 ± 0.22	-0.54	-1.42 to -0.54	-4.37	<.0001
Random Effects															
σ^2	311.63					1.59									
τ^2	Participant:292.54					Participant:1.67									
N	Vignette:33.50					Vignette:0.00									
	Participant:83					Participant:83									
	Vignette:6					Vignette:6									

Table 5
Results of hypothesis-driven regression analysis (Study 3b).

Predictors	Compassion				Willingness to help				Deservedness						
	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p
Intercept	46.77 ± 3.40					4.49 ± 0.18					1.52 ± 0.08				
Neutral group > Outgroup irrelevant	0.48 ± 2.86	0.02	-5.15–6.11	0.17	0.866	0.12 ± 0.19	0.07	-0.25–0.50	0.64	0.524	-0.06 ± 0.15	-0.06	-0.35–0.23	-0.42	0.678
Outgroup irrelevant > Outgroup relevant	8.90 ± 2.86	0.33	3.27–14.53	3.11	0.002	0.62 ± 0.19	0.34	0.25–1.00	3.25	0.001	-0.32 ± 0.15	-0.29	-0.61 to -0.03	-2.16	0.032
Random Effects															
σ^2	335.04					1.50									
τ^2	Participant:368.26					Participant:1.69									
N	Vignette:32.48					Vignette:0.03									
	Participant:82					Participant:82									
	Vignette:6					Vignette:6									

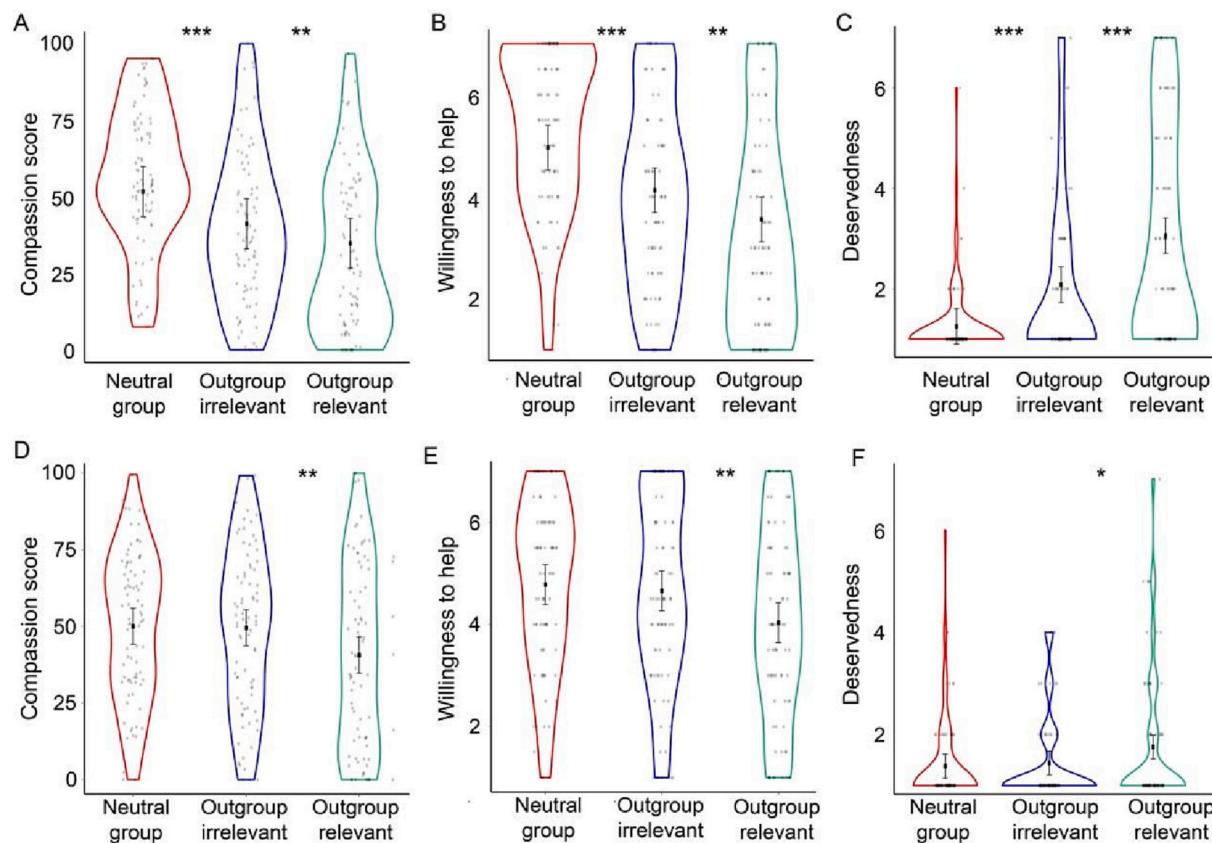


Fig. 4. Effects of condition on compassion (Study 3a: A, Study 3b: D), willingness to help (Study 3a: B, Study 3b: E), and deservedness judgment (Study 3a: C, Study 3b: F) in Study 3.

deserving of the suffering in the Neutral group condition (1.25 ± 0.08 ; $B = -0.83 \pm 0.22$, $t = -3.73$, $p < 0.001$, relative to the Outgroup irrelevant condition).

For Study 3b (Fig. 4F; Table 5), the participants judged the sufferer in the Neutral group condition (1.38 ± 0.09) and the Outgroup irrelevant condition (1.44 ± 0.10) as similarly not deserving of the suffering ($B = -0.06 \pm 0.15$, $t = -0.42$, $p = 0.678$). The participants judged the sufferer as deserving of the suffering more when the sufferer's outgroup identity was causally relevant to their suffering (1.76 ± 0.15) than when the sufferer's outgroup identity was irrelevant to their suffering ($B = 0.32 \pm 0.15$, $t = 2.16$, $p = 0.032$, relative to the Outgroup relevant condition).

4.2.4. Normativity of feeling compassion

For study 3a, internal reliability of the normativity items ranged from 0.79 to 0.88. We therefore combined them into a composite score of normativity. Participants judged that it was least normatively required to (i.e., should or ought to) feel compassion for the political outgroup member whose political identity was relevant to or played a role in their suffering (4.26 ± 0.17 ; no difference relative to the midpoint of the scale, $t = 1.37$, $p = 0.17$). Normativity judgment was significantly higher in the Outgroup irrelevant condition (5.18 ± 0.13 ; $B = 0.77 \pm 0.14$, $t = 5.54$, $p < 0.001$, relative to the Outgroup relevant condition) and the Neutral group condition (5.95 ± 0.12 ; $B = 0.92 \pm 0.14$, $t = 6.58$, $p < 0.001$, relative to the Outgroup irrelevant condition). For study 3b, participants judged that it was least normatively required to (i.e., should or ought to) feel compassion for the political outgroup member whose political identity was relevant to or played a role in their suffering (4.51 ± 0.18). Normativity judgment was significant higher in the Outgroup irrelevant condition (5.11 ± 0.17 ; $B = 0.60 \pm 0.18$, $t = 3.35$, $p = 0.001$) relative to the Outgroup relevant condition, but there was no significant difference between the Outgroup irrelevant condition

and the Neutral group condition (5.37 ± 0.15 ; $B = 0.26 \pm 0.18$, $t = 1.46$, $p = 0.146$).

These results suggested that for the same inculpable behavior that causes harm, sufferers affiliated with a political outgroup elicited less compassion and less willingness to help than sufferers from a politically neutral group. In the contemporary political atmosphere in the United States, political outgroups have constantly been the target of moral outrage in public discourses (Brady, Crockett, & Van Bavel, 2020; Brady, Wills, Jost, Tucker, & Van Bavel, 2017; Tappin & McKay, 2019), and are regarded as less human (Haslam, 2006; Cassese, 2021; Pacilli, Roccato, Pagliaro, & Russo, 2016). Therefore, compassion may also have the problem of parochialism as empathy does (Västfjäll, Erlandsson, Slovic, & Tinghög, 2017; Västfjäll, Slovic, Mayorga, & Peters, 2014), thereby potentially calling into question the more diffused and encompassing nature of compassion, as has been argued (Bloom, 2017a, 2017b). Although our findings seem to suggest that participants affiliated with the Democratic party exhibited stronger Morality Effect on compassionate responses than participants affiliated with the Republican party, future research is needed to replicate this finding.

5. Study 4

In Studies 3a and 3b, we showed that the moral status of group membership had an influence on observers' judgment of whether a sufferer deserves the suffering and their compassionate responses to the sufferer. We varied the moral status of group by manipulating whether the sufferer identified with the participants' rivalry political party in the contemporary U.S. political system. In Study 4, we aimed to conceptually replicate the findings using a different manipulation of the group's moral status, namely citizenship of a (hypothetical) place. It has been shown that nationality is viewed as a meaningful source of social identity (Hussak & Cimpian, 2019; Kinzler, 2020; Liberman, Gerdin,

Table 6
Results of hypothesis-driven regression analysis (Study 4).

Predictors	Compassion				Willingness to help				Deservedness						
	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p	B ± SE	b	95% CI of B	t	p
Intercept	55.25 ± 3.68	0.54	8.82–18.28	5.64	<0.001	5.06 ± 0.13	0.44	0.31–1.08	3.53	0.001	2.98 ± 0.16	-1.12 ± 0.21	-0.57	-1.53 to -0.70	-5.31 <0.001
Neutral group > Outgroup irrelevant	13.55 ± 2.40	0.23	1.02–10.47	2.39	0.017	0.69 ± 0.20	0.27	0.05–0.82	2.21	0.028	-0.32 ± 0.21	-0.16	-0.73–0.10	-1.51	0.133
Outgroup irrelevant > Outgroup relevant	5.75 ± 2.40	0.23	1.02–10.47	2.39	0.017	0.44 ± 0.20	0.27	0.05–0.82	2.21	0.028	-0.32 ± 0.21	-0.16	-0.73–0.10	-1.51	0.133
Random Effects						1.65					1.89				
σ^2	244.94		Participant:277.66			Participant:0.68		Participant:1.59			Vignette:0.00		Participant:0.00		
t_{00}			Vignette:55.26			Vignette:0.02		Vignette:0.02			Vignette:6		Participant:85		
N			Participant:85			Vignette:6		Vignette:6							

Kinzler, & Shaw, 2020), and compared to political groups, people may have less choice for the place they are born into. By manipulating the moral status of the sufferer's citizenship in a hypothetical place, this manipulation also avoids the potential confound of participants' own relationship to the group. Specifically, we tested the hypothesis that observers would reduce their expressed compassion and willingness to help toward the individuals who may not themselves engage in any morally problematic activities or possess any morally bad character but happen to belong to a group in which some members behave immorally (and therefore perceived as deserving of suffering). Moreover, along the lines of the previous studies, we hypothesized that self-reported compassion and willingness to help would diminish even more when the group members' immoral activities are causally relevant to the suffering.

5.1. Method

5.1.1. Participants

Based on the results of Study 3, we preregistered a small-to-medium effect size ($f^2 = 0.169$) for the comparison of the compassion scores between the two key conditions (see below for details; https://aspredicted.org/BXS_WZW). Power analysis indicates that a sample of $N = 79$ is needed to detect this effect at $p < 0.05$ level with 95% power. Ninety participants who self-reported as American residents and currently lived in the United States were recruited from Prolific. Four of them were excluded due to failure in attention checks, leaving 86 participants in the final dataset (46 female; $M_{age} = 36.2$ years, s.d. = 10.3; Caucasian/White: 65%, Black/African: 8%, Hispanic/Latinx: 7%, Asian: 13%, Native American: 2%, Pacific Islander: 0%, Prefer not to answer: 1%, Others: 4%; 76% with a bachelor's or higher degree).

5.1.2. Experimental design and procedure

This study had a 3-level within-participant design. Participants read three vignettes depicting a fictitious city and the suffering of some of its residents. In the two Immoral conditions (i.e., Relevant and Irrelevant), many residents of the city engage in a morally problematic activity (e.g., ignoring the directives from the scientists during a pandemic). In the Relevant condition, some residents of the city, regardless of whether they themselves engage in the morally questionable activity, suffer from the bad consequences of the morally problematic activity (e.g., getting sick due to the infectious disease). In the Irrelevant condition, some residents of the city, regardless of whether they themselves engage in the morally problematic activity, suffer from some bad consequences irrelevant to the morally problematic activity (e.g., getting sick due to extremely cold weather). In the third, Neutral condition, the residents of the city are described to engage in a morally neutral activity (e.g., decorating their home with indoor plants). Then, some residents, regardless of whether they engage in the morally neutral activity, suffer from the same bad consequences as in the Irrelevant condition. Thus, the first two conditions only differed in terms of the causal role of the morally problematic activity of a subgroup of the residents in the suffering residents' bad consequences, whereas the latter two conditions only differed in terms of the presence of a morally problematic activity in a subgroup of the entire resident population (i.e., the described activity was equally irrelevant to the bad consequences).

As in Study 3a, three vignettes were created, each had three variations corresponding to each condition (please see the online *Supplemental Materials* for the materials). Each participant read each vignette exactly once, but in different conditions. The mapping between vignettes and conditions and the presentation order of conditions were randomized across participants. The procedure was the same as that of Study 3a. As an example, the Relevant version of a vignette reads,

In a pandemic, many Naland citizens deliberately ignore the directives from the scientists and continue social gathering as usual. As

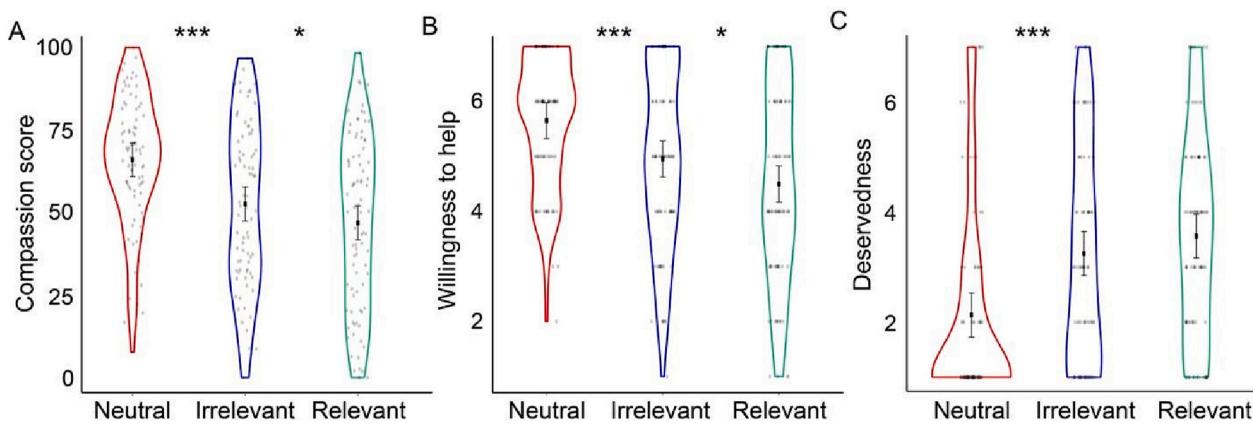


Fig. 5. Effects of condition on compassion (A), willingness to help (B), and deservedness judgment (C) in Study 4.

Table 7
Definition of Morality Effect and Causality Effect.

	Morality Effect	Causality Effect
Study 1	Morally neutral > Immoral non-causal	Immoral non-causal > Immoral causal
Study 2	Morally neutral > Immoral non-causal	Immoral non-causal > Immoral causal
Study 3a/ Study 3b	Neutral group > Outgroup irrelevant	Outgroup irrelevant > Outgroup relevant
Study 4	Neutral > Irrelevant	Irrelevant > Relevant

a result, some Nalando citizens, regardless of their attitudes toward the scientists' advice, get sick and suffer from health problems.

The Irrelevant version of the vignette reads,

In a pandemic, many Nalando citizens deliberately ignore the directives from the scientists and continue social gathering as usual. This year, a terrible snowstorm hits the area. Some Nalando citizens, regardless of their attitudes toward the scientists' advice, get sick and suffer from the extreme weather.

The Neutral version of the vignette reads,

In a pandemic, many Nalando citizens develop the habit of decorating their home with indoor plants. Later, a terrible snowstorm hits the area. Some Nalando citizens, regardless of whether they develop the habit of decorating their home with indoor plants, get sick and suffer from the extreme weather.

5.1.3. Primary dependent variables, explanatory variables, and control variables

Measures of compassion, deservedness, moral evaluations of the suffering and the sufferers, perceived normativity of compassion, and other supplemental variables were the same as in Study 3a. The only difference was that in Study 4, we used "citizens of [city's name] who are

Table 9
Internal meta-analysis of the Causality Effect.

Study	Compassion			Willingness to help		
	b ± SE	95% CI	Weight (%)	b ± SE	95% CI	Weight (%)
1	0.24 ± 0.42	[0.06, 0.42]	17.9	0.12 ± 0.11	[-0.09, 0.33]	13.8
2	0.19 ± 0.31	[0.07, 0.31]	40.2	0.19 ± 0.06	[0.07, 0.31]	43.8
3a	0.33 ± 0.55	[0.11, 0.55]	12.0	0.36 ± 0.10	[0.16, 0.56]	15.8
3b	0.33 ± 0.43	[0.13, 0.43]	14.5	0.34 ± 0.10	[0.14, 0.54]	15.8
4	0.23 ± 0.42	[0.04, 0.42]	15.4	0.27 ± 0.12	[0.03, 0.50]	10.9
Common effect model	0.24 [0.17, 0.32]			0.24 [0.16, 0.32]		
Heterogeneity (I^2)	0% [0.0%, 79.2%]			10.2% [0.0%, 81.3%]		

suffering from the dire situation" to refer to the sufferers, instead of a specific name.

Measuring willingness to help. Unlike in the previous four studies, where the sufferer is a single person, here the sufferers are a group of people. To adjust to this large-scale nature of the suffering, we modified the measure of willingness to help. Specifically, we asked the participants to imagine themselves as the leader of a charitable organization that has the ability to provide humanitarian aid to the affected city. The participants read the following short scenario:

An international charity organization has some limited resources of humanitarian aid. Because the resources are limited, using it to help people in one area means not being able to help people elsewhere who equally need the help. Moreover, each individual will receive help proportional to their need,

Table 8
Internal meta-analysis of the Morality Effect.

Study	Compassion			Willingness to help		
	b ± SE	95% CI	Weight (%)	b ± SE	95% CI	Weight (%)
1	1.36 ± 0.08	[1.20, 1.52]	17.1	1.22 ± 0.10	[1.03, 1.41]	16.5
2	1.70 ± 0.06	[1.58, 1.82]	30.5	1.74 ± 0.06	[1.62, 1.86]	43.1
3a	0.61 ± 0.10	[0.41, 0.81]	11.0	0.67 ± 0.10	[0.47, 0.87]	15.5
3b	0.02 ± 0.10	[-0.18, 0.22]	11.0	0.07 ± 0.10	[-0.13, 0.27]	15.5
4	0.54 ± 0.06	[0.42, 0.66]	30.5	0.44 ± 0.13	[0.19, 0.69]	9.5
Common effect model	0.98	[0.92, 1.05]		1.11	[1.03, 1.18]	
Heterogeneity (I^2)	98.8%	[98.2%, 99.2%]		98.5%	[97.8%, 99.0%]	

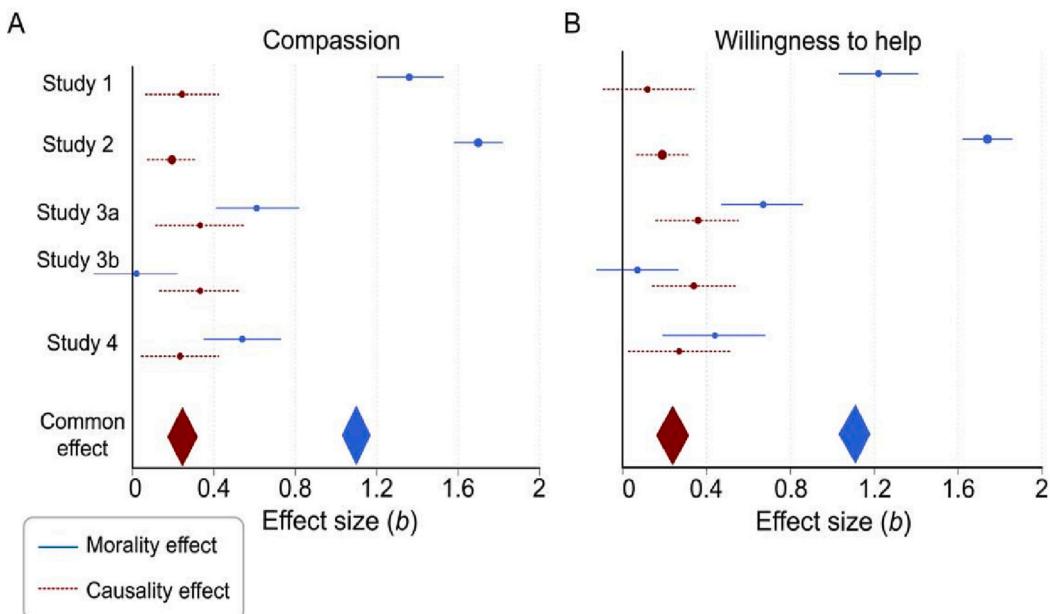


Fig. 6. Results of the internal meta-analysis. The effect size (standardized regression coefficient) of moral status of the sufferer (“Morality effect”) and its causal role in the suffering (“Causality effect”) on observers’ compassion (A) and willingness to help (B).

regardless of their attitudes or behaviors in the past. If you were the leader of this charity and responsible for how its resources are used, how likely would you be to use the resources to help the citizens of [city’s name] who are experiencing the dire situation?

Then, the participants indicated their willingness to offer help on a 7-point Likert scale (1 = Not at all likely, 4 = Moderately likely, 7 = Extremely likely).

5.2. Results and discussion

5.2.1. Compassion and willingness to help

We estimated two linear mixed effects models, with the compassion and willingness to help as dependent variables, respectively. We included condition (Relevant, Irrelevant, Neutral) as the key independent variable, and participants’ age, gender, and socioeconomic status as covariates of no interest, and participant ID and vignette version as random intercepts (Table 6).

As Fig. 5A shows, participants reported the highest compassion to the sufferers in the Neutral condition (66.49 ± 2.12), followed by the Irrelevant condition (52.94 ± 2.80 ; $B = -13.55 \pm 2.40$, $t = -5.64$, $p < 0.001$, relative to the Neutral condition). Participants reported the lowest compassion to the sufferers in the Relevant condition (47.19 ± 2.81 ; $B = -5.75 \pm 2.40$, $t = -2.39$, $p = 0.017$, relative to the Irrelevant condition). Willingness to help exhibited a similar pattern (Fig. 5B). Specifically, participants were most willing to help the sufferer in the Neutral condition (5.67 ± 0.13), followed by the Irrelevant condition (4.98 ± 0.17 ; $B = -0.69 \pm 0.20$, $t = -3.53$, $p = 0.001$, relative to the Neutral condition). Participants were least willing to help the sufferer in the Relevant condition (4.54 ± 0.19 ; $B = -0.44 \pm 0.20$, $t = -2.21$, $p = 0.028$, relative to the Irrelevant condition).

5.2.2. Deservedness of suffering

We next examined whether participants’ deservedness judgment was modulated by a subset of group members’ morally problematic behaviors/attitudes, and the causal relevance of such behaviors/attitudes to the suffering. The participants judged the sufferers as least deserving of the suffering in the Neutral condition (2.13 ± 0.19). In comparison, the deservedness judgment was significantly higher when a subset of group members exhibited morally problematic behaviors/attitudes, which

were irrelevant to the suffering (3.25 ± 0.22 ; $B = 1.12 \pm 0.21$, $t = 5.31$, $p < 0.001$, relative to the Neutral condition). The deservedness judgment was slightly higher in the Relevant condition (3.56 ± 0.20) relative to the Irrelevant condition, but the difference was not significant ($B = 0.32 \pm 0.21$, $t = 1.51$, $p = 0.133$). As in the previous studies, deservedness judgment was significantly and negatively associated with both expressed compassion ($B = -2.03 \pm 1.13$, $t = -2.03$, $p = 0.044$; Fig. 5C) and willingness to help ($B = -0.29 \pm 0.09$, $t = -3.36$, $p < 0.001$; Fig. 5D), controlling for the main effect of condition and the interaction between condition and moral desert judgment.

5.2.3. Normativity of feeling compassion

Internal reliability of the normativity items ranged from 0.88 to 0.91. We therefore combine the items into a composite measure of normativity. Participants judged that it was least normatively required to feel compassion and sympathy in the Relevant condition (4.89 ± 0.14), which was significantly lower than that in the Irrelevant condition (5.32 ± 0.15 ; $B = -0.43 \pm 0.15$, $t = -2.96$, $p = 0.003$), which was in turn significantly lower than that in the Neutral condition (5.86 ± 0.11 ; $B = -0.55 \pm 0.15$, $t = -3.77$, $p < 0.001$).

Results of this study show that the dampening effect of moral desert judgment on compassion spillover to individuals who have done nothing wrong but merely share group membership with those who are responsible for the suffering. However, compared with the previous studies (Study 1 through Study 4), the effect sizes of both moral status and causal relevance were smaller in the current study, indicating the existence of boundary conditions of the effect of morality-based withholding of compassion and willingness to help. The relative reluctance to generalize the judgment of a subset of a group to other group members may reflect the individualist cultural value of respecting individual autonomy, of which the American culture is representative (Markus & Kitayama, 1991; Nisbett, 2004). Future research is needed to examine how cultural value may modulate the tendency to generalize the moral evaluation of a subset of a group to the entire group.

6. Internal meta-analysis

To more formally compare the effects of moral status of the sufferer (“Morality effect”) and its causal role in the suffering (“Causality effect”)

Table 10
Results of the mediation analysis for compassion.

	Path a (95% CI)	Path b (95% CI)	Path a*b (95% CI)	c (95% CI)	c' (95% CI)
<i>Morality effect</i>					
Study 1	-0.83 ± 0.16 [-1.15, -0.51]	-4.00 ± 1.37 [-6.74, -1.26]	3.31 ± 1.31 [0.86, 6.14]	5.90 ± 1.81 [2.28, 9.51]	2.59 ± 2.04 [-1.51, 6.68]
Study 2	-0.41 ± 0.11 [-0.63, -0.20]	-4.66 ± 1.11 [-6.86, -2.46]	1.93 ± 0.69 [0.71, 3.43]	9.73 ± 1.33 [7.10, 12.36]	7.80 ± 1.32 [5.19, 10.42]
Study 3a	-0.64 ± 0.15 [-0.93, -0.35]	-3.32 ± 1.63 [-6.55, -0.08]	2.12 ± 1.15 [0.22, 4.72]	9.13 ± 2.17 [4.82, 13.44]	7.01 ± 2.37 [2.30, 11.72]
Study 3b	-0.39 ± 0.08 [-0.55, -0.24]	-4.11 ± 2.26 [-8.62, 0.40]	1.62 ± 1.07 [-0.33, 3.96]	5.86 ± 1.60 [0.62, 9.05]	4.24 ± 1.82 [7.85]
Study 4	-0.46 ± 0.11 [-0.69, -0.24]	-1.40 ± 1.04 [-3.47, 0.67]	0.65 ± 0.60 [-0.40, 2.02]	6.79 ± 1.08 [4.65, 8.94]	6.15 ± 1.18 [3.81, 8.49]
<i>Causality effect</i>					
Study 1	-0.67 ± 0.27 [-1.21, -0.13]	-3.57 ± 0.95 [-5.46, -1.68]	2.39 ± 0.98 [0.63, 4.50]	3.58 ± 2.19 [-0.81, 7.96]	1.18 ± 2.08 [-2.99, 5.35]
Study 2	-0.81 ± 0.17 [-1.14, -0.48]	-2.68 ± 0.71 [-4.09, -1.26]	2.18 ± 0.98 [0.62, 4.42]	5.91 ± 1.31 [3.32, 8.50]	3.73 ± 1.37 [1.02, 6.44]
Study 3a	-0.84 ± 0.21 [-1.26, -0.42]	-1.98 ± 1.15 [-4.27, 0.32]	1.66 ± 1.01 [0.16, 4.04]	9.06 ± 2.22 [4.64, 13.48]	7.40 ± 2.40 [2.63, 12.18]
Study 3b	-0.55 ± 0.12 [-0.79, -0.32]	-1.27 ± 2.26 [-5.77, 3.22]	0.71 ± 0.85 [-1.06, 2.38]	8.57 ± 2.37 [3.85, 13.28]	7.86 ± 2.69 [2.51, 13.21]
Study 4	-0.32 ± 0.15 [-0.61, -0.03]	-3.13 ± 1.10 [-5.31, -0.94]	0.99 ± 0.57 [0.10, 2.35]	8.46 ± 1.52 [5.43, 11.49]	7.47 ± 1.50 [4.48, 10.46]

CI: confidence interval; c: total effect; c': direct effect.

on observers' compassion and willingness to help, we carried out an internal meta-analysis. For Study 1, we collapsed the two Morally neutral conditions, and the Morality Effect was defined as the contrast between the Morally neutral and the Immoral non-causal condition, while the Causality Effect was defined the contrast between the Immoral non-causal condition and the Immoral causal condition (Table 7). For all the studies, the definition of the Morality effect and the Causality effect was shown in Table 7. Effect size was calculated as the standardized coefficient corresponding to the respective contrast reported in Study 1 through Study 4. We calculated, separately for compassion and willingness to help, the common effect and heterogeneity of both the Morality Effect and the Causality Effect using the 'meta' package in R (Schwarzer, 2007). The comparison indicated that, across all the four studies, the moral status of the sufferer had a stronger effect on observers' compassion and willingness to help than the causality of the moral badness in the suffering (Tables 8–9; Fig. 6).

7. Mediation analysis

To test our hypothesis that people perceive morally bad people as more deserving of suffering and therefore express less compassion and

Table 11
Results of the mediation analysis for willingness to help.

	Path a (95% CI)	Path b (95% CI)	Path a*b (95% CI)	c (95% CI)	c' (95% CI)
<i>Morality effect</i>					
Study 1	-0.83 ± 0.16 [-1.15, -0.51]	-0.33 ± 0.10 [-0.52, -0.14]	0.28 ± 0.10 [0.07, 0.49]	0.33 ± 0.13 [0.07, 0.58]	0.05 ± 0.14 [-0.23, 0.33]
Study 2	-0.41 ± 0.11 [-0.63, -0.20]	-0.27 ± 0.07 [-0.40, -0.14]	0.11 ± 0.04 [0.04, 0.21]	0.57 ± 0.08 [0.42, 0.73]	0.46 ± 0.08 [0.31, 0.61]
Study 3a	-0.64 ± 0.15 [-0.93, -0.35]	-0.13 ± 0.10 [-0.34, 0.08]	0.08 ± 0.07 [-0.04, 0.25]	0.71 ± 0.14 [0.44, 0.98]	0.63 ± 0.15 [0.33, 0.93]
Study 3b	-0.39 ± 0.08 [-0.55, -0.24]	-0.45 ± 0.16 [-0.76, -0.14]	0.18 ± 0.09 [-0.00, 0.37]	0.61 ± 0.11 [0.38, 0.83]	0.43 ± 0.12 [0.18, 0.68]
Study 4	-0.46 ± 0.11 [-0.69, -0.24]	-0.18 ± 0.10 [-0.38, 0.01]	0.09 ± 0.07 [-0.03, 0.23]	0.54 ± 0.10 [0.33, 0.74]	0.45 ± 0.11 [0.23, 0.68]
<i>Causality effect</i>					
Study 1	-0.67 ± 0.27 [-1.21, -0.13]	-0.50 ± 0.11 [-0.72, -0.28]	0.34 ± 0.15 [0.09, 0.69]	0.01 ± 0.27 [-0.52, 0.54]	-0.33 ± 0.24 [-0.81, 0.15]
Study 2	-0.81 ± 0.17 [-1.14, -0.48]	-0.24 ± 0.06 [-0.35, -0.12]	0.19 ± 0.08 [0.07, 0.39]	0.54 ± 0.11 [0.33, 0.75]	0.35 ± 0.11 [0.13, 0.56]
Study 3a	-0.84 ± 0.21 [-1.26, -0.42]	-0.22 ± 0.09 [-1.26, -0.42]	0.19 ± 0.08 [0.06, 0.37]	0.55 ± 0.17 [0.36, 0.88]	0.36 ± 0.18 [0.01, 0.71]
Study 3b	-0.55 ± 0.12 [-0.79, -0.32]	-0.11 ± 0.13 [-0.38, 0.15]	0.06 ± 0.07 [-0.09, 0.20]	0.61 ± 0.14 [0.22, 0.89]	0.55 ± 0.16 [0.24, 0.86]
Study 4	-0.32 ± 0.15 [-0.61, -0.03]	-0.15 ± 0.08 [-0.32, -0.03]	0.05 ± 0.04 [-0.00, 0.15]	0.55 ± 0.11 [0.32, 0.77]	0.50 ± 0.12 [0.27, 0.73]

CI: confidence interval; c: total effect; c': direct effect.

willingness to help, we estimated two sets of mediation models, one corresponding to the Morality Effect, the other corresponding to the Causality Effect, as defined in Table 7. In these mediation models, we first calculated for each variable the difference scores corresponding to the Morality Effect and the Causality Effect, respectively. We then entered the moral inference of the sufferer's character as the independent variable (i.e., perceived moral badness), the deservedness judgment as the mediator, and the compassion score or the willingness to help score as the dependent variable, respectively. The PROCESS macro for SPSS was used for these analyses (Hayes, 2012).

Results of the mediation analyses for compassion and willingness to help can be found in Table 10 and Table 11. Both the causality-based deservedness (i.e., the Causality effect) and the non-causality-based deservedness (i.e., the Morality effect) mediated the relationship between the moral impression of the sufferer and the withholding of compassionate responses (i.e., compassion and willingness to help) for the first two studies. The mediating effect of deservedness judgment was weaker or less consistent for studies 3 and 4, when deservedness was more based on what the sufferer is (e.g., group identity) than what the sufferer does (e.g., intention, behaviors). These results both lent support to the hypothesis that moral deservedness of the sufferer acts as a barrier

to compassion and suggested the existence of boundary conditions of the effect of desert-based barrier to compassion. We note that, as a limitation of the mediation analysis, the mediator was measured at the same time as outcome variables. It should be noted that these results are correlational and should be interpreted with caution.

8. General discussion

In four studies, we demonstrated that adult American participants withheld their compassion and willingness to help when the sufferers possess of an immoral intention (e.g., harming others) prior to their own suffering (Study 1), have a bad moral character in the past (e.g., being a dishonest person; Study 2), or merely sharing the identity of a morally laden group (Studies 3 and 4). These effects held even when the perceived moral badness of the sufferers did not objectively cause their own suffering. Moral deservedness—the extent to which the sufferer was viewed as deserving the suffering—mediated the effects of moral judgment on compassion. Surprisingly, withholding compassion based on moral judgment was viewed as normative, such that participants consistently judged that extending compassion to those who deserve the suffering was not morally required and not something they would even expect morally good people to do.

Our study contributes to a deeper understanding of the nature of and potential barriers to compassion. Compassion to all human beings is valued as a moral ideal in different philosophical and religious traditions (e.g., Gilbert, 2020; Lampert, 2005). While compassion may be less biased by the sufferers' race, physical distance, and similarities compared with other prosocial tendencies such as empathy (Cikara et al., 2011; Singer, 2016; Small et al., 2007), our studies nevertheless demonstrate that compassion is constrained by how observers perceived the sufferers' moral status. Across studies, supporting the evolutionary and functionalist theories (Goetz et al., 2010; Szycer et al., 2017; Fig. 1), we consistently found that more negative moral judgments give rise to perceptions of deservedness of suffering, which in turn leads to reduced compassionate responses. Importantly, our studies demonstrate that moral deservedness functions as a key proximate and constraining factor in the appraisal of compassion. We also went one step further by systematically distinguishing the effects of causal and non-causal antecedents of moral deservedness judgment on the appraisal of compassion. Thus, our studies can serve as bridges connecting evolutionary and functionalist theories, moral philosophy, and empirical psychological investigations of compassion.

The fact that we found similar effects in terms of people's normative expectations further suggests that this is not only a natural tendency but is indeed a barrier deeply rooted in people's moral beliefs. This is in line with a recent work examining how third-party observers' evaluation of empathizer varies as a function of the moral status of the target of empathy (Wang & Todd, 2021). This study shows that third-party observers' evaluation of the empathizer is more negative when the target of the empathy is morally problematic, suggesting a normative evaluation of the deployment of empathy toward morally problematic sufferers. Of note, across our studies, perceived normativity of compassion was not extremely highly correlated with compassion itself (correlation coefficients ranging from 0.44 to 0.68; for details, see Supplementary Table 6), indicating that these two measures might reflect distinct psychological constructs. These findings provide empirical evidence for philosophical theories about the deservedness-based qualifications of compassion and sociological observations of compassion both in contemporary American society and in ancient civilizations (Clark, 2007; Nussbaum, 2003).

We acknowledge that the impacts of moral judgments on social and moral cognition are broader than influencing compassion and willingness to help. Indeed, as previous work has shown, moral evaluation of a person has profound and broad influences on many aspects of person perception and social decision-making, such as warmth, competence, and subjective well-being (Brambilla et al., 2021; Goodwin, Piazza, &

Rozin, 2014). Specifically in the fairness and justice literature, it has been shown that a morally bad person will be assigned harsher punishment and blame than a morally neutral or good person, even if their behaviors are equally blameworthy (Siegel et al., 2017). Given this prior literature, our goal here was not to demonstrate that the effects of moral judgment are specific to compassion and willingness to help, but instead to empirically demonstrate that the influence of moral judgment extends to compassionate responses, a psychological faculty that some argue to be inclusive and transcend arbitrary social boundaries.

By focusing on participants' compassion and willingness to help as third-party bystanders (rather than as second-party targets of the sufferer's actions), our findings cannot be explained by self-interests or retaliation, thereby going beyond previous research involving second-person interactions (e.g., Singer et al., 2006) and providing strong evidence that desert-based moral consideration per se does function as a mediator between moral judgment and compassion. Although in the present research we were primarily interested in how moral perceptions may elicit deservedness judgment, theoretically deservedness may also be triggered by other non-moral characteristics of the sufferers (e.g., pure behavioral causality without blameworthy intention). Therefore, deservedness may have the potential to integrate diverse findings in explaining and conceptualizing the scope of compassion, a fruitful direction that future theoretical and empirical research could focus on.

Our findings are consistent with and contribute to the growing research on the limitations of compassion. Existing research has revealed that experiencing compassion can be costly and effortful (Condon & Barrett, 2013; Eisenberg & Eggum, 2009; Scheffer et al., 2021), and people withhold compassion to sufferers who cause their own suffering (Rudolph et al., 2004; Weiner, 1985). Our studies went a step further by revealing how moral judgment influences compassion independent from the effect of causality judgment. Even when the perceived immoral behavior, character, or group identity did not objectively cause the suffering, participants still expressed significantly lower levels of compassion, by comparison to the same suffering of a morally neutral individual. Although we experimentally separated the causal and the non-causal contribution to the sufferer's suffering, we cannot rule out the possibility that *subjectively*, some participants may still imagine and attribute some causal role to the sufferer perceived to be morally bad even in the non-causal situations (due to just-world or related system justification beliefs; Jost, Banaji, & Nosek, 2004; Jost, 2020; Lerner, 1980). Indeed, across studies, the participants assigned some causality to the sufferers even when the sufferers do not objectively cause their suffering (for details about the Morality Effect and Causality Effect on self-reported objective causality, see Supplementary Table 7). But across the studies, the Morality Effect was stronger than the Causality Effect (Tables 8-9, Fig. 6), suggesting causality attributions alone may not account for the moral barrier effect. The existence of such a psychological tendency may reveal the powerful role of moral evaluations in attributing causality (even in the absence of direct causal information), and it is an empirical question to what extent such attributions could also mediate the effect between moral judgment and compassion. While this important question calls for a new set of studies to be adequately addressed, the present studies do provide some preliminary evidence showing that perceived causality does not mediate the effect between moral judgment and compassion, as deservedness does. Specifically, we estimated another set of mediation models where we replaced the self-reported deservedness with self-reported causality as the mediator between moral judgment difference and compassion difference. In four out of the five studies, self-reported causality was not a significant mediator (Supplementary Table 8), suggesting that subjective causal attribution might not function in the same way as deservedness judgment in mediating moral judgment and compassion.

Our findings have also revealed that people's judgment of deservedness is influenced by perceived moral status of the sufferer's group membership. Although the effect is smaller compared to the effect of intention and moral character, it clearly exists and reveals a boundary

of our key effect. History abounds with moral denigration of the oppressed and marginalized groups (e.g., minorities and the financially deprived), with the powerful ruling class adopting narratives to accuse them of bringing the “deserved” suffering to themselves and rationalizing their suffering, for the sake of justifying large-scale purge and cleansing (Godfrey & Wolf, 2016; Sainz, Martínez, Sutton, Rodríguez-Bailón, & Moya, 2020; Sandel, 2020). Our results showed that American adults may indeed have the psychological underpinnings to be susceptible to such desert-based narratives: they withheld compassion and help when the sufferer was a member of a political rivalry group, especially when the group membership was relevant to the suffering (Study 3). Such morality-based withholding of compassion and help even extended to morally innocent members of a “deserved” group (Study 4).

This desert-based moral barrier to compassion could be challenging to overcome because observers not only withhold their compassion and help, but also believe that it is morally good and required to do so. In other words, withholding compassion based on moral deservedness could be part of people’s moral narratives, instead of being simply a natural tendency (Atkins, 2008). Because morally charged narratives are easy to adhere to emotionally and associated with true self (Brady et al., 2020; Strohminger & Nichols, 2014), interventions aiming to challenge such narratives and to expand compassion to the morally condemned (under certain narratives) would be difficult (Muddiman & Stroud, 2017; Nithyanand, Schaffner, & Gill, 2017).

One potential approach is to adopt counter narratives that emphasize the common humanity shared by all human beings, and to highlight the fact that regardless of how the suffering is brought about, suffering itself is the same across human race (Gaita, 2013). Humanizing processes through physical connections and feelings, though criticized as sometimes leading to the narrow focus of empathy (Bloom, 2017b; Jordan et al., 2016), may actually have the potential to counteract desert-based moral barriers to compassion (when applied to people who are otherwise viewed as deserving of the suffering). As the historian Katharina von Kellenbach reported in her book *The Mark of Cain*, a brief eye contact between a Nazi perpetrator and a dying Jewish victim “humanized him [the victim] and made this scene unbearable to Zakis [the Nazi perpetrator]” (Von Kellenbach, 2013). Empirical studies have shown that paying attention to the feelings of stigmatized victims (e.g., individuals living with HIV) reduces desert-based prejudice (Batson et al., 1997; Zaki, 2019), and narratives on common humanity could promote compassion in healthcare workers (Baguley, Dev, Fernando, & Considine, 2020; Ling, Olver, & Petrakis, 2020; Ling, Petrakis, & Olver, 2021; Sinclair et al., 2018). It would be important to examine if similar approaches would be effective for removing the moral barriers of compassion.

Several limitations of the current research are worth noting. First, the vignettes were hypothetical, despite that the contents were plausible and relatable in everyday life. Future studies adopting more real-life experimental setup, either by including real human partners or by using virtual reality techniques (Hortensius, Neyret, Slater, & de Gelder, 2018; Zanon, Novembre, Zangrandino, Chittaro, & Silani, 2014), are needed to replicate the findings in a more ecologically valid manner. Second, the dependent variables were all self-reports. In the Western cultures, empathy and compassion are socially desirable and praiseworthy, as demonstrated both in empirical research (Fowler, Law, & Gaesser, 2021; Nook, Ong, Morelli, Mitchell, & Zaki, 2016) and public discourse (e.g., Armbruster, 2019; Obama, 2006). The effects of moral deservedness on self-reported compassion thus go against social desirability tendency. Notwithstanding, it would still be valuable to provide converging evidence using more implicit measures (e.g., physiological and neural correlates of compassion) and behavioral measures of helping. Third, our participants were all American adults from Prolific. It will be theoretically interesting to elucidate whether certain cultural values and characteristics (e.g., collectivism, interdependence, essentialism, meritocracy, etc.) might exacerbate or reduce the effect of moral desert.

To conclude, the present studies addressed a series of questions

regarding how perceived moral badness of the sufferer influences observers’ compassion and willingness to help. The results consistently demonstrate that sufferers who are perceived as more morally bad based on various characteristics (e.g., intentions, character traits, shared group memberships) are met with reduced compassion and willingness to help, a relationship mediated by the perceived deservingness of the suffering. Across studies, the effect of moral desert judgment was independent from and even stronger than the effect of causality judgment. This pattern, largely in line with meritocracy, system justification, and other deservedness-based ideologies, points to a potential moral barrier to compassion, a blind spot hitherto largely overlooked by proponents (both theorists and psychologists) of compassion and a mechanism that prevents us from alleviating the greatest suffering possible.

CRediT authorship contribution statement

Hongbo Yu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Supervision, Writing - original draft, Writing - review & editing. **Jie Chen:** Data curation, Formal analysis, Methodology, Validation, Writing – review & editing. **Bernadette Dardaine:** Data curation, Formal analysis, Methodology, Validation, Writing – review & editing. **Fan Yang:** Conceptualization, Investigation, Validation, Supervision, Writing – review & editing.

Data availability

Data and analysis codes have been shared openly on OSF

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2023.105476>.

References

- Aarøe, L., & Petersen, M. B. (2014). Crowding out culture: Scandinavians and Americans agree on social welfare in the face of deservingness cues. *The Journal of Politics*, 76(3), 684–697.
- Aliche, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63(3), 368.
- Aliche, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126(4), 556.
- Aliche, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science*, 10(6), 790–812.
- Armbruster, H. (2019). “It was the photograph of the little boy”: reflections on the Syrian Vulnerable Persons Resettlement Programme in the UK. *Ethnic and Racial Studies*, 42(15), 2680–2699.
- Atkins, K. (2008). *Narrative identity and moral identity: A practical perspective*. Routledge.
- Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology*, 20(11), 1018–1022.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *science*, 211(4489), 1390–1396.
- Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2013). Their pain is not our pain: brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping*, 34(12), 3168–3181.
- Baguley, S. I., Dev, V., Fernando, A. T., & Considine, N. S. (2020). How do health professionals maintain compassion over time? Insights from a study of compassion in health. *Frontiers in Psychology*, 11, 3327.
- Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., ... Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 113(17), 4688–4693.
- Barrett, H. C., & Saxe, R. R. (2021). Are some cultures more mind-minded in their moral judgements than others? *Philosophical Transactions of the Royal Society B*, 376(1838), 20200288.
- 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(6), 1042–1054.
- 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(1), 105.

- Bentham, J. (1780/1988). *The principles of morals and legislation*. Buffalo: New York: Prometheus Books.
- Bloom, P. (2017a). *Against empathy: The case for rational compassion*. Random House.
- Bloom, P. (2017b). Empathy and its discontents. *Trends in Cognitive Sciences*, 21(1), 24–31.
- Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: the role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, 15(4), 978–1010.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318.
- Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In, Vol. 64. *Advances in experimental social psychology* (pp. 187–262). Academic Press.
- 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(8), 934–942.
- Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiafreou, E., & Izquierdo, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, 148(6), 962–976. <https://doi.org/10.1037/xge0000595>
- Cassese, E. C. (2021). Partisan dehumanization in American politics. *Political Behavior*, 43 (1).
- Caviola, L., Everett, J. A., & Faber, N. S. (2019). The moral standing of animals: Towards a psychology of speciesism. *Journal of personality and social psychology*, 116(6), 1011.
- Celniker, J., Gregory, A., Koo, H., Piff, P. K., Ditto, P. H., & Shariff, A. (2023). The moralization of unproductive effort. *Journal of Experimental Psychology*, 152(1), 60–79.
- Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them: intergroup failures of empathy. *Current Directions in Psychological Science*, 20(3), 149–153.
- Clark, C. (2007). *Misery and company*. University of Chicago Press.
- Condon, P., & Feldman Barrett, L. (2013). Conceptualizing and experiencing compassion. *Emotion*, 13(5), 817.
- Cosmides, L., Guzmán, R. A., & Tooby, J. (2018). The evolution of moral cognition. In *The Routledge handbook of moral epistemology* (pp. 174–228). Routledge.
- Crimson, C. R., Hornsey, M. J., Bain, P. G., & Bastian, B. (2018). Toward a psychology of moral expansiveness. *Current Directions in Psychological Science*, 27(1), 14–19.
- Crisp, R. (2008). Compassion and beyond. *Ethical Theory and Moral Practice*, 11(3), 233–246.
- Decety, J., & Cowell, J. M. (2014). Friends or foes: is empathy necessary for moral behavior? *Perspectives on Psychological Science*, 9(5), 525–537.
- Dennett, D. C., & Caruso, G. D. (2021). *Just deserts: Debating free will*. Polity Press.
- Dodds, E. R. (1951). *The Greeks and the irrational*. University of California Press.
- Dyer, R. L., Pizarro, D. A., & Ariely, D. (2022). They had it coming: the relationship between perpetrator-blame and victim-blame. *Social Cognition*, 40(6), 503–527.
- Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety, & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71–83). Cambridge: MIT press.
- Engelmann, N., & Waldmann, M. R. (2022). How causal structure, causal strength, and foreseeability affect moral judgments. *Cognition*, 226, Article 105167.
- Feldman, F., & Skow, B. (2020). In E. N. Zalta (Ed.), “Desert”, *The Stanford encyclopedia of philosophy* (Winter 2020 edition). <https://plato.stanford.edu/archives/win2020/entries/desert/>.
- Fernando, A. T., III, & Consedine, N. S. (2014). Beyond compassion fatigue: the transactional model of physician compassion. *Journal of Pain and Symptom Management*, 48(2), 289–298.
- Fitouchi, L., André, J.-B., & Baumard, N. (2023). Are there really so many moral emotions? Carving morality at its functional joints. In L. Al-Shawaf, & T. K. Shackelford (Eds.), *The Oxford handbook of evolution and the emotions*. New York: Oxford University Press (in press).
- Fowler, Z., Law, K. F., & Gaesser, B. (2021). Against empathy bias: The moral value of equitable empathy. *Psychological*, 32(5), 766–779.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
- Gaita, R. (2013). *A common humanity: Thinking about love and truth and justice*. Routledge.
- Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, 78(3), 757–774.
- Gilbert, P. (2020). Compassion: from its evolution to a psychotherapy. *Frontiers in Psychology*, 11, 3123.
- Godfrey, E. B., & Wolf, S. (2016). Developing critical consciousness or justifying the system? A qualitative analysis of attributions for poverty and wealth among low-income racial/ethnic minority and immigrant women. *Cultural Diversity and Ethnic Minority Psychology*, 22(1), 93.
- Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: an evolutionary analysis and empirical review. *Psychological Bulletin*, 136(3), 351.
- Goetz, J. L., & Peng, K. (2019). Sympathy and responses to suffering: Similarity and variation in China and the United States. *Emotion*, 19(2), 320.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148.
- Haslam, N. (2006). Dehumanization: an integrative review. *Personality and Social Psychology Review*, 10(3), 252–264.
- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. Retrieved from <http://www.aphayes.com/public/process2012.pdf>.
- Hortensius, R., Neyret, S., Slater, M., & de Gelder, B. (2018). The relation between bystanders' behavioral reactivity to distress and later helping behavior during a violent conflict in virtual reality. *PLoS One*, 13(4), Article e0196074.
- Hudson, S. T. J., Cikara, M., & Sidanius, J. (2019). Preference for hierarchy is associated with reduced empathy and increased counter-empathy towards others, especially out-group targets. *Journal of Experimental Social Psychology*, 85, Article 103871.
- Hussak, L. J., & Cimpian, A. (2019). “It feels like it's in your body”: how children in the United States think about nationality. *Journal of Experimental Psychology: General*, 148(7), 1153.
- Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct? *Emotion*, 16(8), 1107–1116. <https://doi.org/10.1037/emo0000228>
- Jost, J. T. (2020). *A theory of system justification*. Harvard University Press.
- Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25(6), 881–919.
- Kagan, S. (2014). *The geometry of desert*. Oxford University Press.
- Kinzler, K. D. (2020). *How You Say It: Why You Talk the Way You Do? And What It Says About You*. Houghton Mifflin.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33 (4), 315–329.
- Kristjánsdóttir, K. (2018). *Virtuous emotions*. Oxford University Press.
- Lampert, K. (2005). *Traditions of compassion: From religious duty to social activism*. New York: Palgrave Macmillan.
- Lerner, M. J. (1980). *The belief in a just world*. New York: Plenum.
- Liberman, Z., Gerdin, E., Kinzler, K. D., & Shaw, A. (2020). (Un) common knowledge: Children use social relationships to determine who knows what. *Developmental Science*, 23(6), Article e12962.
- Ling, D., Olver, J., & Petrakis, M. (2020). Investigating how viewing common humanity scenarios impacts compassion: a novel approach. *The British Journal of Social Work*, 50(6), 1724–1742.
- Ling, D., Petrakis, M., & Olver, J. H. (2021). The use of common humanity scenarios to promote compassion in healthcare workers. *Australian Social Work*, 74(1), 110–121.
- Maestri, G., & Monforte, P. (2020). Who deserves compassion? The moral and emotional dilemmas of volunteering in the ‘refugee crisis’. *Sociology*, 54(5), 920–935.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224.
- Mascaro, J. S., Florian, M. P., Ash, M. J., Palmer, P. K., Frazier, T., Condon, P., & Raison, C. (2020). Ways of knowing compassion: how do we come to know, understand, and measure compassion when we see it? *Frontiers in Psychology*, 11, 2467.
- McCullough, M. E. (2020). *The kindness of strangers: How a selfish ape invented a new moral code*. Simon and Schuster.
- McNamara, R. A., Willard, A. K., Norenzayan, A., & Henrich, J. (2019). Weighing outcome vs. intent across societies: how cultural models of mind shape moral reasoning. *Cognition*, 182, 95–108.
- Muddiman, A., & Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, 67(4), 586–609.
- Nisbett, R. (2004). *The geography of thought: How Asians and westerners think differently... and why*. Simon and Schuster.
- Nithyanand, R., Schaffner, B., & Gill, P. (2017). *Online political discourse in the Trump era*. ArXiv Preprint ArXiv:1711.05303.
- Nook, E. C., Ong, D. C., Morelli, S. A., Mitchell, J. P., & Zaki, J. (2016). Prosocial conformity: prosocial norms generalize across behavior and empathy. *Personality and Social Psychology Bulletin*, 42(8), 1045–1062.
- Nussbaum, M. (1996). Compassion: the basic social emotion. *Social Philosophy and Policy*, 13(1), 27–58.
- Nussbaum, M. (2003). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
- Obama, B. (2006, June 19). Commencement address at Northwestern University. <https://www.northwestern.edu/newscenter/stories/2006/06/barack.html>.
- O'Brien, E., & Ellsworth, P. (2012). Polar opposites: empathy does not extend across the political aisle. *Jury Expert*, 24, 25.
- Pacilli, M. G., Roccato, M., Pagliaro, S., & Russo, S. (2016). From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup. *Group Processes & Intergroup Relations*, 19(3), 360–373.
- Pfattheicher, S., Sassenrath, C., & Keller, J. (2019). Compassion magnifies third-party punishment. *Journal of Personality and Social Psychology*, 117(1), 124.
- Phillips, J., De Freitas, J., Mott, C., Gruber, J., & Knobe, J. (2017). True happiness: the role of morality in the folk concept of happiness. *Journal of Experimental Psychology: General*, 146(2), 165.
- Raihani, N. (2021). *The social instinct: How cooperation shaped the world*. Random House.
- Robb, D., & Heil, J. (2021). In E. N. Zalta (Ed.), “Mental causation”, *The Stanford encyclopedia of philosophy* (Spring 2021 edition). <https://plato.stanford.edu/archives/spr2021/entries/mental-causation/>.
- Roberts, J. V. (1997). *Role of criminal record in the sentencing process (from crime and justice: A review of research)* (vol. 22, pp. 303–362). Michael Tonry. ed.–See NCJ-166203.
- Rudolph, U., Roesch, S., Greitemeyer, T., & Weiner, B. (2004). A meta-analytic review of help giving and aggression from an attributional perspective: contributions to a general theory of motivation. *Cognition and Emotion*, 18(6), 815–848.
- Sainz, M., Martínez, R., Sutton, R. M., Rodríguez-Bailón, R., & Moya, M. (2020). Less human, more to blame: animalizing poor people increases blame and decreases support for wealth redistribution. *Group Processes & Intergroup Relations*, 23(4), 546–559.
- Sandel, M. J. (2020). *The tyranny of merit: What's become of the common good?* Penguin UK.

- Scheffer, J. A., Cameron, D., & Inzlicht, M. (2021). Caring is costly: people avoid the cognitive work of compassion. *PsyArxiv*. <https://doi.org/10.1037/xge0001073>
- Schwarzer, G. (2007). meta: an R package for meta-analysis. *R News*, 7(3), 40–45.
- Siegel, J. Z., Crockett, M. J., & Dolan, R. J. (2017). Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition*, 167, 201–211. <https://doi.org/10.1016/j.cognition.2017.05.004>
- Siegel, J. Z., Mathys, C., Rutledge, R. B., & Crockett, M. J. (2018). Beliefs about bad people are volatile. *Nature Human Behaviour*, 2(10), 750–756.
- Sinclair, S., Hack, T. F., Raffin-Bouchal, S., McClement, S., Stajduhar, K., Singh, P., ... Chochinov, H. M. (2018). What are healthcare providers' understandings and experiences of compassion? The healthcare compassion model: a grounded theory study of healthcare providers in Canada. *BMJ Open*, 8(3), Article e019701.
- Sinclair, S., Norris, J. M., McConnell, S. J., Chochinov, H. M., Hack, T. F., Hagen, N. A., ... Bouchal, S. R. (2016). Compassion: a scoping review of the healthcare literature. *BMC Palliative Care*, 15(1), 1–16.
- Singer, P. (2015). The logic of effective altruism. *Boston Review*, 1.
- Singer, P. (2016). *Famine, affluence, and morality*. USA: Oxford University Press.
- Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, 24(18), R875–R878.
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075), 466–469. <https://doi.org/10.1038/nature04271>
- Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: the impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143–153.
- Stellar, J. E., Anderson, C. L., & Gatchpazian, A. (2020). Profiles in empathy: different empathic responses to emotional and physical suffering. *Journal of Experimental Psychology: General*, 149(7), 1398.
- Strauss, C., Taylor, B. L., Gu, J., Kuyken, W., Baer, R., Jones, F., & Cavanagh, K. (2016). What is compassion and how can we measure it? A review of definitions and measures. *Clinical Psychology Review*, 47, 15–27. <https://doi.org/10.1016/j.cpr.2016.05.004>
- Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131(1), 159–171.
- Sznycer, D., Lopez Seal, M. F., Sell, A., Lim, J., Porat, R., Shalvi, S., ... Tooby, J. (2017). Support for redistribution is shaped by compassion, envy, and self-interest, but not a taste for fairness. *Proceedings of the National Academy of Sciences*, 114(31), 8420–8425.
- Tappin, B. M., & McKay, R. T. (2019). Moral polarization and out-party hostility in the US political context. *Journal of Social and Political Psychology*, 7(1), 213–245.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, 46(1), 35–57.
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10(1), 72–81. <https://doi.org/10.1177/1745691614556679>
- Van Dijk, W. W., Ouwerkerk, J. W., Goslinga, S., & Nieweg, M. (2005). Deservingness and Schadenfreude. *Cognition and Emotion*, 19(6), 933–939.
- Van Kleef, G. A., Oveis, C., Van Der Löwe, I., LuoKogan, A., Goetz, J., & Keltner, D. (2008). Power, distress, and compassion: turning a blind eye to the suffering of others. *Psychological Science*, 19(12), 1315–1322.
- Västfjäll, D., Erlandsson, A., Slovic, P., & Tinghög, G. (2017). Commentary: empathy and its discontents. *Frontiers in Psychology*, 8, 542.
- Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: affect and charity are greatest for a single child in need. *PLoS One*, 9(6), Article e100115.
- Von Kellenbach, K. (2013). *The mark of Cain: guilt and denial in the post-war lives of Nazi perpetrators*. Oxford University Press.
- Wang, Y. A., & Todd, A. R. (2021). Evaluations of empathizers depend on the target of empathy. *Journal of Personality and Social Psychology*, 121(5), 1005–1028. <https://doi.org/10.1037/pspi0000341>
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford.
- Weiner, B. (2010). The development of an attribution-based theory of motivation: a history of ideas. *Educational Psychologist*, 45(1), 28–36.
- Williams, C. R. (2008). Compassion, suffering and the self: a moral psychology of social justice. *Current Sociology*, 56(1), 5–24.
- Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, 29(26), 8525–8529.
- Yang, F., Knobe, J., & Dunham, Y. (2021). Happiness is from the soul: the nature and origins of our happiness concept. *Journal of Experimental Psychology: General*, 150(2), 276.
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104(20), 8235–8240.
- Yu, H., Siegel, J. Z., Clithero, J. A., & Crockett, M. J. (2021). How peer influence shapes value computation in moral decision-making. *Cognition*, 211, Article 104641.
- Zaki, J. (2019). *The war for kindness: Building empathy in a fractured world*. Broadway Books.
- Zanon, M., Novembre, G., Zangrandino, N., Chittaro, L., & Silani, G. (2014). Brain activity and prosocial behavior in a simulated life-threatening situation. *NeuroImage*, 98, 134–146.