Status and Moral Education: On the Philosophy and Psychology of Punishment

Michał Kłusek

Draft paper presented at **VSPEE - Research seminar in law, economics, and empirics** on October 8, 2024

Please cite published version

Abstract:

Philosophers have debated the justification of punishment for a long time. More recently, psychologists began studying lay intuitions of punishment to find out whether they fit any of the philosopher's accounts. They firmly established that the intuitions are retributive (rather than consequentialist) and then that they are expressivist—punishment is meant to send a message. Lately, it has been argued that they fit Antony Duff's communicative theory of punishment. This article shows that this argument is off the mark. More importantly, it surveys a wide range of research to show that, on the contrary, punitive intuitions are best captured by a combination of status-based and moral education theories. The former claim that punishment is meant to raise the status of the victim and lower the offender's. This is perfectly in line with psychological research on the effects of victimization and punishment, on antisocial punishment, and on punishment's direct effect on social standing. The latter claim that we care about the moral change of the punished offender – in line with the latest psychological experiments.

Keywords: punishment, moral intuitions, communicative theory, social standing, moral education

Status and Moral Education: On the Philosophy and Psychology of Punishment

> "As soon as men began to value one another, and the idea of consideration had got a footing in the mind, every one put his claim to it, and it became impossible to refuse it to any with impunity. Hence arose the first obligation of civility even among savages; and every intended injury became an affront; because, besides the hurt which might result from it, the party injured was certain to find in it a contempt for his person, which was often more insupportable than the hurt itself."

Jean-Jacques Rousseau, "On the Origin of Inequality of Mankind" (1913, pp. 197–198)

Introduction

The desire to punish is both baffling and universal. We find punishment in all human cultures, but it is hard to pinpoint why it makes sense. Like St. Augustin pondering the

nature of time, we all know why we punish until we are asked to explain it.[1] Why do we feel *morally obliged* to harm[2] wrongdoers?[3]

Philosophers have suggested many justifications[4], and a small psychological industry has grown up to test which of them best fit our intuitions. Psychologists have long focused on the two best-known accounts: retributive justice and deterrence theory (Aharoni & Fridlund, 2012; Baron et al., 1993; Baron & Ritov, 1993, 2009; Carlsmith et al., 2002; Carlsmith, 2006; Carlsmith & Darley, 2008; Darley et al., 2000; Kłusek, 2023; Sunstein et al., 2000; Twardawski et al., 2020; Twardawski & Hilbig, 2020). The former justifies punishment as what the wrongdoers deserve. The latter, with reference to deterrence of future crime. Retributivism has been declared the clear winner in terms of intuitiveness (Darley, 2009). We may verbally justify punishment with positive consequences, but we are all retributivists at heart (Hoffman, 2014, p. 343).

That said, later studies showed that punitive intuitions are closer to *expressive* theories, which claim that punishment is justified as the best or the only way to express something important (e.g., Duff, 2003; Feinberg, 1965; French, 2001; Hampton, 1991, 1992; Hirsch,

---

[1] "What then is time? Provided that no one asks me, I know. If I want to explain it to an inquirer, I do not know." (Augustine, 2008, p. 230)

[2] Whether punishment must involve the intention to harm is a topic of some philosophical controversy (see e.g.: Hanna, 2017, 2020; Wringe, 2013, 2019).

[3] There are many ways to understand the "why" question, depending on the discipline. This article focuses on the psychological mechanisms underlying punitiveness. Different answers would be given by evolutionary biology, sociology, economics, etc.

[4] There are also thinkers who claim that punishment cannot be justified - it is a basic, nonrational instinct or emotion (classically (Holmes, 1963; Stephen, 1883), more recently (Barash & Lipton, 2011; Henberg, 1990; Solomon, 1990)).

1996; Kahan, 1996; Metz, 2000; Primoratz, 1989; Skillen, 1980; Wringe, 2016). We generally want the offender to know *why* he was punished and *by whom* (classically (Heider, 1958; Miller, 2001; Vidmar, 2001), more recently: (Cushman et al., 2019; Gollwitzer et al., 2011; Gollwitzer et al. 2014; Gollwitzer & Denzler, 2009; Molnar et al., 2020; Sarin et al., 2021; Sjöström et al., 2017).

Recently, Nahmias and Aharoni went a step further and claimed our intuitions may be best captured by Duff's communicative account (Duff, 2003, 2022; Nahmias & Aharoni, 2017). On this theory, punishment seeks to persuade the offenders to "repent their crimes, to reform themselves, and to reconcile themselves (…) with those they have wronged." (Duff, 2003, p. 116)

I disagree. This article shows that our punitive intuitions are best captured by a combination of what I call *status-based* expressive theories and *moral education* theories of punishment. The former claim that we punish to send a message about the *status* of the wrongdoer and her victim. Wrongdoing diminishes the victim's social standing and raises the perpetrator's. Punishment lifts the victim up and brings the wrongdoer down. The latter theories claim that punishment is justified as a way to teach a *moral* lesson.

In what follows, I first discuss Nahmias and Aharoni's attempt to explain punitive intuitions using Duff's communicative account and demonstrate where I think they failed. I then present several status-based theories of punishment and psychological research that links punishment and social standing. Next, I complicate the picture by discussing psychological research demonstrating that punitive intuitions go beyond empowerment. I show that,

rather than Duff's communicative account, they are most in-line with *moral education* theories, as developed by Hampton (1984) and Morris (1981). Finally, I argue that a combination of status-based and moral education accounts allows us to capture all of our psychological intuitions.

Communicative Theory

The recent contribution by Nahmias and Aharoni is noteworthy for moving beyond the classic retributivist/deterrence duo (Nahmias & Aharoni, 2017). They argue that the communicative theory of punishment may best fit our punitive psychology (Duff, 2003).

Antony Duff's communicative account sees punishment as a sort of dialogue, a two-way communication between the offender and the community. It is built on the view that criminal law is a collective declaration that certain actions are wrong. Such declaration, if it is sincere, must be followed by criticism of those who engage in such actions. Punishment communicates such censure. For or some crimes, only penal hard treatment can express it appropriately.

However, criminal punishment should also pursue the three aims of repentance, reform and reconciliation. According to Duff, prison is supposed to provide an opportunity for the offenders to "examine their souls" (Duff 2003, p. 87), it is to be "a space for reflection and reformation." (Brooks 2012, p. 104). Punishment must be burdensome and painful to "go deep with the wrongdoer and (...) occupy his attention, his thoughts, his emotions, for considerable time." (Duff 2003, p. 108). The offenders ought to endure "secular penance,"

because "a (mere) apology cannot heal the moral wound done by the wrong." (Duff 2003, p. 95).

In short, "Punishment should be understood, justified, and administered as a mode of moral communication with offenders that seeks to persuade them to repent their crimes, to reform themselves, and to reconcile themselves through punishment with those they have wronged." (Duff 2003, pp. 115-116).

Nahmias and Aharoni argue that Duff's theory requires paying close attention to the mental states of the offender before, during, and after the crime. This is necessary to assess what the appropriate punishment is and whether it served its purpose. Nahmias and Aharoni suggested that, if Duff's theory is descriptively on track, people's judgements of punishment should be responsive (among other things) to whether the crime was pre-planned or opportunistic; whether the criminal had served a sentence for a similar crime before; and whether he offered a sincere apology. Pre-planned crimes demonstrate that the offender cares less about the community's norms. A pattern of recidivism suggests that he cannot internalize them. A sincere apology shows that she understands how wrong her crime was.

They carried out a survey study based on the prediction that "punishers will be responsive to information about criminal intent, criminal history, and the perceived sincerity of apology." (Nahmias & Aharoni, 2017, p. 150) They presented their participants with vignettes describing a robbery and asked for their sentence recommendations. They varied whether the crime was planned or not; whether the criminal was a first- or second-time offender; and whether he offered a sincere or insincere apology. The primary dependent measure was the

length of the recommended sentence. As hypothesized, both high intent and prior criminal history evoked more severe punishment. Both sincere and insincere apology decreased punishment, but sincere apology had a larger effect under the 'high intent' condition.

A later study tested the effects of perpetrator suffering and understanding on people's satisfaction with punishment and punishment recommendations.[5] (Aharoni et al., 2022) Then, the authors hypothesized that, first, "the signal that the perpetrator understands why he has been punished [would] increase the satisfaction with the prospect of parole and reduce additional sentence recommendations" (understanding hypothesis); second, that evidence of suffering would have the same effect (suffering hypothesis); and third, that the combination of understanding and suffering will have the greatest effect on punishment goal fulfillment (understanding by suffering hypothesis) (Aharoni et al., 2022, p. 141). The "understanding by suffering" hypothesis would be consistent with Duff's theory (Aharoni et al., 2022, p. 140).

Again, they presented their participants with vignettes describing crimes of varying seriousness and asked for sentence recommendations. The participants then read evidence of "either or both understanding and suffering being present or absent" (Aharoni et al., 2022, p. 143). The fictional defendant served a 10-week jail term while waiting for trial. The suffering in question involved anxiety, losing one's job, and the worsening of

[5] It must be pointed out that the later paper does not claim that Duff's theory *best* captures lay punitive intuitions. It only broadly argues that the "understanding hypothesis" is consistent with, and suggests the psychological plausibility of, expressive and communicative theories. I thank the reviewer for raising this point.

asthma or a stomach ulcer. The evidence of understanding involved a jail therapist's testimony that the offender deeply regretted his actions. The participants were asked whether they wanted to change their initial punishment decision in light of the new information. Finally, they were informed that the perpetrator received immediate parole and asked how satisfied they were with this being his only sentence.

Sentence recommendations decreased when the perpetrator understood that what he did was wrong, which confirmed the understanding hypothesis. The participants were also more satisfied with the prospect of parole in such a situation. The exact same effect was observed when the offender suffered, confirming the suffering hypothesis. However, there was no synergistic interaction between suffering and understanding. "Participants were no more likely to reduce their sentences when understanding and suffering were both present compared to when just one of these was present." (Aharoni et al., 2022, p. 145)

Crucially, the vast majority of the participants *did not* change their original punishment judgement in response to new information about understanding or suffering. As Aharoni and colleagues wrote, "it might suggest that perpetrator suffering and understanding (…) are not sufficient to satisfy these individuals' punishment goals." (Aharoni et al., 2022, p. 147) Uniform punishers were, however, more satisfied with the prospect of parole when there was evidence of perpetrator understanding. The effects of understanding and suffering on satisfaction with the prospect of parole were strongest for the less serious crimes.

The empirical work of Aharoni, Nahmias, and colleagues is second to none, and their results are valuable regardless of their relationship to Duff's theory. However, I want to argue that they failed to show Duff's account to be descriptively accurate.

Their studies show that we want the offenders to understand why they are punished; that their suffering matters for punishment goal fulfillment; that people are more punitive in response to pre-planned crimes and second-time offenders, and less punitive for offenders who offer (sincere) apology. But these findings have little to do with Duff's theoretical work.

If one wants to see whether people's intuitions track Duff's communicative theory, examining sentence recommendations is not the way to go. In his view, punishment must be, first and foremost, proportional (Duff, 2003, sec. 4.1). Punishment is *for an offense*, "its character and severity must (…) be determined by the offense for which it is imposed (…) We must determine not just that an offender deserves censure but how severe that censure should be: the more serious the crime, the more severe the deserved censure." (Duff, 2003, p. 132) The severity of the penal hard treatment serves to communicate the severity of censure. To punish somebody disproportionately is to communicate more, or less, censure than they deserve. This is "dishonest and unjust" (Duff, 2003, p. 132).

That punishment should be proportional to the moral wrongfulness of an act (usually taken to be a function of harm and culpability) is a core tenet of most theories of punishment. That the criminal intent affected the participants' sentence recommendations is unsurprising and in line with Duff's account. But this fact matches the 'predictions' of many other philosophical theories as well.

The second reason why focusing solely on sentence recommendations does not get to the core of Duff's account is that his theory puts the *meaning* of punishment center stage, not its severity. As Duff wrote, "What matters about crimes is not just their seriousness but their character as public wrongs. What matters about punishment is not just their severity but their character as responses to such wrongs." (Duff, 2003, p. 139) A study design that compared different *forms* of punishment would be more in line with Duff's account.

Also, let us turn to the notion of suffering employed in the second study. That "guilty deserve to suffer" is an intuition well-deserving of psychological exploration (Davis, 1972). However, the suffering described in Aharoni and colleagues' study is quite irrelevant from the point of view of Duff's theory. As Duff wrote, "It would be absurd for sentencers to try to calculate how much offenders had already suffered in their lives and how much extra suffering, if any, must be inflicted by punishment to give them what they deserve; and not much less absurd to reduce a burglar's sentence because he caught pneumonia while carrying out the crime." (Duff, 2003, p. 20) Communicative punishment "…aims to bring offenders to suffer what they deserve to suffer - the pains of repentance and remorse" (Duff, 2003, p. 107) — not the pains of a stomach ulcer or asthma. Penal hard treatment forces the offender to focus his attention on his crime, which must be unpleasant. The pain of repentance is the only one that matters.

On Duff's account, it is *punishment itself* that communicates a public, reparative apology. Whether the offender himself apologizes and whether they apologize sincerely is of secondary importance. Penal hard treatment itself "constitutes a forceful and public

apology" (Duff, 2003, p. 119). Even more importantly, the aim of punishment is not to exact repentance but to persuade the offender in a way that leaves them free to remain unpersuaded (Duff, 2003, sec. 3.7.4.). Even the offender who is already repentant (as, presumably, is the one who sincerely apologizes after a 10-week jail time)[6] must be punished on Duff's account (Duff, 2003, sec. 3.7.3). On the other hand, a refusal to apologize (let alone an insincere apology) should in no way alter our judgements of punishment. As Duff wrote: "[The] severity of the punishment [must] be proportionate to the seriousness of the offense; and just as later repentance does not mitigate the seriousness of the offense, the offender's defiance does not aggravate it." (Duff, 2003, p. 122)

Finally, inquiring into whether the offender's apology is sincere fails to respect his autonomy - a liberal polity must not be concerned with the offender's conscience. Again, on Duff's account, the offenders are forced to *hear* the message of punishment, but they are not forced to *listen* and be persuaded by it. (Duff, 2003, p. 126)

Nahmias and Aharoni's studies do not give us strong reasons to claim that Duff's theory is the most accurate as a *descriptive* account. Let me suggest other contenders.

Status-based Theories of Punishment

Status-based theories, such as Jean Hampton's expressive retributivism, justify punishment with reference to social standing or worth (Hampton, 1988, 1991, 1992).

[6] Also, on Duff's account, repentance requires time and effort - with serious wrongs, it likely requires more than 10 weeks.

According to Hampton's theory, crime requires a retributive reaction because it is (also) a *moral insult* causing a *moral injury*. It expresses the criminal's superiority. It carries the insulting message of: "I count but you do not", "I can use you for my purposes", or "I am here up high and you are there down below." (Murphy & Hampton, 1988, p. 28) A wrongdoer treats her victims worse than their personal worth requires.

"A person behaves wrongfully in a way that effects a moral injury to another when she treats that person in a way that is precluded by that person's value, and/or by representing him as worth far less than his actual value; or, in other words, when the meaning of her action is such that she diminishes him, and by doing so, represents herself as elevated with respect to him, thereby according herself a value that she does not have."(Hampton, 2006, p. 127)

Punishment shows the message to be false. It "takes down" the wrongdoer and elevates the victim. It reveals the wrongdoer's true value, which is *equal* to the victim's.[7] Punishment must be proportional to the crime to fully deny the insulting message, but not too severe so it does not falsely bring the victim 'above' the wrongdoer (Hampton, 2006, p. 140). Retributive punishment is about "asserting moral truth [of equality] in the face of its denial." (Hampton, 1991, p. 398)

Jonathan Wolff's "communicative retributivism" is a somewhat similar theory (Wolff, 2011). He argued that: "In the case of becoming a victim of crime, one loses the sense of being master of one's fate. Furthermore, one can become the object of pity, which many people

[7] Hampton subscribed to a Kantian theory of human worth, according to which people are "intrinsically, objectively and equally valuable." (Hampton, 1991, p. 397) The general framework of her account can be combined with other theories of worth as well.

find diminishing. But most of all, another person has treated you with contempt, and has succeeded in doing so. (…) crime seems, in at least some cases, to bring about a change in status and self-respect." (Wolff, 2011, p. 116)

When a criminal wrongs you, "they implicitly announce themselves as in some respect your superior. They have victimized you, and left you with lowered status." (Wolff, 2011, p. 125) The role of punishment is "to re-establish some sort of proper status between all the parties. If a criminal is caught and adequately punished (…) he can no longer implicitly claim to be of higher status, and those who were victims may feel that their victimhood is expunged, and they have their previous status restored to them." (Wolff, 2011, p. 125)

Finally, a roughly similar theory has been suggested by Whitley Kaufman (Kaufman, 2013). On his view, punishment defends *honor*. As Kaufman writes: "[I]t was the universal traditional assumption that the purpose of revenge was the defense of one's honor. An attack on one's person was an attack on one's honor, and honor had to be defended by engaging in a physical confrontation with the wrongdoer." (Kaufman, 2013, p. 121) He approvingly cites Nietzsche, writing that: "The revenge of restoration does not protect against further harm; it does not make good the harm suffered — except in one case. If our honor has suffered from our opponent, then revenge can restore it. But this has suffered damage in every instance in which our suffering has been inflicted on us deliberately; for our opponent thus demonstrated that he did not fear us. By revenge we demonstrate that we do not fear him either: this constitutes the equalization, the restoration." (Nietzsche, 1989, p. 181)

Each of the theories links punishment to an expression of the victim's value, status, or honor. In the following sections, I present research showing that this is a deeply intuitive notion.

Punishment and Status Intuitions

What follows is an empirical case for the status-based theories of punishment, which capture the psychological causes and effects of punitiveness ignored by all other accounts.[8] First, I show that victimization diminishes status-related psychological resources (self-esteem, power, and a sense of control), and punishment restores them. Second, I discuss antisocial or strategic punishment, which aims not at enforcing moral norms, but rather at altering the standing of the parties. Third, I discuss studies that directly "tested" Hampton's theory and showed that punishment affects social standing.

The effects of victimization and punishment

Victimization

Victimization strips us of status-related psychological resources (Shnabel & Nadler, 2008). It lowers our self-esteem and increases psychological distress (Brockner et al., 2003, 2008; Koper et al., 1993; Norris & Kaniasty, 1994; Schroth & Pradhan Shah, 2000). As Scobie

[8] That punishment aims at the restoration of the victim's lost status and power has been recognized by the *empowerment-focused* psychological accounts of punishment and revenge (Shnabel & Nadler, 2008; Okimoto & Wenzel, 2008; Fischer et al., 2022). This section reviews the standard arguments for such accounts as well as lesser-known studies.

and Scobie wrote, in a way that mirrors philosophical arguments, "[t]ransgressions (…) devalue the victim and may result in a lowering of their self-esteem and a consequent mismatch between the person's own self-image and the one the offender appears to hold." (Scobie & Scobie, 1998, p. 381) The effect on self-esteem is telling because, according to at least one theory, self-esteem itself measures social belonging and social standing (Leary, 2012).

Victimization also affects power, defined as "the sense that one has control over one's outcomes and, therefore, can resist the influence of others in particular situations" (Strelan et al., 2020, p. 447). Power, in turn, enhances self-esteem (Wojciszke & Struzynska-Kujalowicz, 2007).

Finally, being a victim of injustice is a shaming and humiliating experience. Feelings of shame and humiliation typically follow disruption of social bonds (Scheff, 2003, 2019).

The effects of victimization on status-related psychological resources suggest that the status-centered theories are onto something. This becomes clearer once we observe punishment's opposite effect.

Punishment

On the face of it, vengeance does not have a clear function. It does not undo past wrongs, and it can be harmful itself. Frijda suggests that vengeance serves power equalization. For him: [w]hen someone willfully harms another, he or she manifestly has the power to do so, and the other lacks the power to prevent it or do likewise. There is power inequality. The

offender has or had power over you, and you are or were powerless. (…) Through revenge, one gets even in power." (Frijda, 1994, p. 275)

Revenge also restores self-esteem, as suggested by Kim and Smith (1993) and a wide range of organizational behavior literature (e.g., Aquino & Douglas, 2003; Ferris et al., 2009). The self-restoration function is more pronounced among individuals high on vertical individualism, who have a strong need to surpass others (Cukur et al., 2004; Singelis et al., 1995). They also experience greater self-esteem threat as a result of victimization and a greater subsequent desire for revenge (Zdaniuk & Bobocel, 2012).

In general, aggression follows when people of high self-esteem are faced with serious threats to their self-image (Baumeister et al., 1996; Bushman & Baumeister, 1998), "[w]hen favorable views about oneself are questioned, contradicted, impugned, mocked, challenged, or otherwise put in jeopardy (…) aggression emerges from a particular discrepancy between two views of self: a favorable self-appraisal and an external appraisal that is much less favorable."(Baumeister et al., 1996, p. 8)

All this closely mirrors the philosophical arguments of status-based theories.

Antisocial Punishment

That punishment affects social standing is also not surprising in light of research on antisocial punishment. While most punitive behavior targets the wrongdoers (in the context of economic games — those who do not cooperate), surprisingly often people also punish those who did nothing wrong, or even the "good guys." This does not make sense on most

philosophical theories of punishment, but it is perfectly in line with status-centered accounts.

In "money burning" experiments, the participants get a chance to reduce other people's incomes at a cost (e.g., Zizzo & Oswald, 2001). It turns out people burn other's money for the hell of it. The lower the cost of burning, the more it happens. That said, most people turn out to be rank-egalitarian; the richer players' funds are burned much more than the poorer ones (Zizzo, 2003).

Other studies found that people pay both to reduce the income of the top earners and to increase that of those at the bottom, purely out of a desire to make the distribution more egalitarian (Dawes et al., 2007).

Psychologists distinguish between strategic (a.k.a. spiteful) and non-strategic punishment (Fehr & Fischbacher, 2002; Fehr & Gächter, 2000). The latter is driven by a desire to uphold cooperation norms based on negative strong reciprocity (Paál, 2021). The former aims at increasing the punisher's payoff, regardless of who is punished.

In highly competitive contexts, where one's standing matters more, strategic punishment becomes far more prevalent (Paál & Bereczkei, 2015). It no longer functions as a tool for fostering cooperation, but rather as a tool for rivalry.

Other studies found that the desire to punish is far greater when 'cheating' results in disadvantageous inequity — in other words, when the wrongdoer gets more than us (Raihani & McAuliffe, 2012). Spiteful punishment is also more prevalent in societies with

extreme social hierarchies. Fehr, Hoff, and Kshetramade (2008) found that in India, high-caste subjects were far more likely to engage in spiteful punishment, imposing severe sanctions on others whether or not it was fair. This was especially the case if they were behind in terms of payoff. Thus, their antisocial punishment was driven by a "concern for status and superiority and their strong aversion to disadvantageous inequality" (Fehr et al., 2008, p. 499). That said, antisocial punishment is present to a lesser or greater extent in all societies. In some societies, high cooperators are just as likely to be punished as those who don't cooperate (Gächter et al., 2010; Herrmann et al., 2008).

This brings us to "punishing the good guys." Even 20% of all punishment in economic games targets cooperators, and it's usually the least cooperative who punish them (Cinyabuguma et al., 2006; Kuběna et al., 2014). Those who do not cooperate also gladly remove high cooperators from their group (Parks & Stone, 2010).

This behavior in economic games is in line with social psychology research on "do-gooder derogation." People who are "too" helpful are sometimes ridiculed or criticized for their efforts (Minson & Monin, 2012; Monin, 2007).

One way to make sense of do-gooder derogation and antisocial punishment is to claim that there are moral norms about how much one should contribute, and those who help "too much" break them (Henrich, 2004; Van Dijk et al., 2015). The biological markets theory is a subtler explanation (Barclay, 2013; Noë & Hammerstein, 1994, 1995). When we decide whom to cooperate with, we choose the most helpful. There is an incentive to "show" how

cooperative you are, and so "competitive helping" or "competitive altruism" emerges (Barclay, 2011, 2013; Roberts, 1998; Sylwester & Roberts, 2010).

Another way to ensure that one is picked as a partner is to take others down. We don't need to be very cooperative; we just need to be more cooperative than everybody else. As Pleasant and Barclay concluded, after experimentally demonstrating this sort of punishment: "…antisocial punishment may be an attempt to stop high cooperators from looking too good, (…), and, by extension, to stop the antisocial punisher from looking selfish in comparison." (Pleasant & Barclay, 2018, p. 869) In other words: "…antisocial punishment and do-gooder derogation (…) prevent one's competitors from gaining relative reputation, which would make oneself look worse by comparison." (Pleasant & Barclay, 2018, p. 875)

Punishment affects social status

Bilz (2016) tested Hampton's theory directly. She carried out three studies to see whether punishment affects the victim's social standing. The first one tested "[t]he prediction (…) that third parties would regard successful criminal punishment as raising a victim's standing, and nonpunishment as lowering it." (Bilz, 2016, p. 364).

The participants watched a film depicting a rape trial. After watching the first half, which depicted the rape's aftermath, they were asked about how the residents of the area where the crime took place would rate the people involved, on different traits. The traits: "admired," "valuable", and "respected" were combined into a "social standing" scale. The participants then watched the second half, which depicted either a conviction or a plea

bargain, and answered the questions once more. The victim gained in social standing when the offenders were punished. The offender lost it. The plea bargain had the opposite effect.

The second study replicated the results using credit fraud.

The third tested whether the effects of punishment differ depending on the in-group or out-group status of the punisher. Punishment expresses both the victim's social standing and the standing of the group to which he belongs. As Bilz wrote, "Consider, for example, an all-black jury for a black victim. Its verdict could be understood to reveal something about whether the victim is well regarded within the black community, but it does not reveal much about how blacks are regarded by other racial groups. An all-white jury rendering a decision with a black victim, on the other hand, could reveal something about how well-regarded blacks are by whites, but would not reveal much about the victim's standing among fellow blacks." (Bilz, 2016, pp. 378–379). The study also measured the effect of punishment on self-esteem, based on Leary's "sociometer" theory of self-esteem (Leary, 2012), where self-esteem is a measure of social belonging.

The participants read and vividly imagined a scenario involving a confrontation where an offender in a car with diplomatic plates collides with their car, yells at them, and slaps their phone in such a way that it strikes them in the eye. The participants then learned that the offender was convicted or acquitted, by an American or a foreign court.

As predicted, punishment increased the victim's social standing, both when the punisher was an ingroup (American court) and an outgroup (foreign court). Punishment by an ingroup had no effect on the victim's *group* status, but punishment by an outgroup punisher

increased both the victim's and the group's social standing. The punishment's effect on self-esteem was fully mediated by social standing.

Other Psychological Research

There is a related line of research that, for brevity's sake, I did not discuss. According to the well-known dual-process account, moral judgements (including judgements of punishment) are formed in a process involving two distinct "systems" (Kahneman, 2011; Sloman, 1996, 2002), with the "fast, automatic, effortless, associative, implicit (…), and often emotionally charged" (Kahneman, 2003, p. 698) System 1 playing the main role (Greene, 2013; Haidt, 2001, 2013; Haidt & Hersh, 2001; on the intuitive character of judgements of punishment, see: Carlsmith & Darley, 2008; Darley, 2009; P. H. Robinson, 2013; P. Robinson & Darley, 2007).

A key role in shaping punitive intuitions is played by anger (Biaggio, 1980; Carlsmith et al., 2002). It motivates more punitive responses to crime (Gault & Sabini, 2000; Hartnagel & Templeton, 2012), increases punishment in economic games (Gummerum et al., 2016; Seip et al., 2014), and is a strong predictor of support for harsh criminal justice policies (Cassese & Weber, 2011; Hartnagel & Templeton, 2012; Johnson, 2009; Ray & Kort-Butler, 2020).

Crucially, anger is closely associated with notions of insult and disrespect — the very notions put center-stage by all three status-based punishment theorists (Averill, 1983; Lazarus, 1991; for a review, see: Miller, 2001).

Moral Change

The case for viewing status-based theories of punishment as most *descriptively* accurate is quite strong. However, they too fail to capture *all* of our punitive intuitions. As already mentioned in the introduction, more recent psychological studies convincingly demonstrated the essentially *communicative* or *expressive* aspects of intuitions of punishment.[9] We want punishment to send a message, and we want the message to be received.

For example, using an implicit measure of goal-fulfillment, Gollwitzer and Denzler (2009) demonstrated that punishment achieves its aim when the offender understands why they were punished. Later, Gollwitzer and colleagues (2011) showed that punishment can only bring satisfaction when the transgressor realizes why they were punished.

More recently, Molnar and colleagues (2020) showed that people have belief-based preferences when they punish. We want the offender to share our understanding of the situation. The participants in Molnar's studies had a strong desire for the offender to understand the reasons behind punishment. They also tended to punish less severely, when there was the option of sending a message to the offender. Sarin and colleagues (2021) showed that, in personal matters, people *prefer* communicative, figurative (i.e. costless)

[9] Of course, this aspect of punishment/revenge is quite obvious from the perspective of communication theory (Berlo 1960), which views them as ways of exchanging meanings and understandings with others. The studies discussed in this section seem to combine communication theory and social psychology, in the way advocated by Boon and Yoshimura (2020).

punishments to costly ones. We also tend to interpret ambiguous cases as cases of punishment.

It seems to be the case that one of the chief aims of punishment is to affect moral change. In a seminal article, Funk and colleagues (2014) demonstrated that people are *most* satisfied with punishment when the offender not only understands why they were punished, but also indicated a positive moral change.

It appears that there are two main motives for punishment: the restoration of status/power and moral education. But perhaps one is reducible to the other? Perhaps offender's feedback of moral change *empowers* the victim? If that were the case, we could simply conclude that, despite appearances, punishment is all about status after all. It is not. A recent series of studies by Fischer and colleagues (2022) demonstrated that the positive effects of offender feedback on victim's satisfaction cannot be reduced to empowerment. Rather, the desire to affect moral change is a core psychological motive irreducible to status/power restoration.[10] Thus, we arrived at a point where there are punitive intuitions that social-status theories of punishment cannot account for.

[10] The psychological picture was possibly complicated by another recent study. Hechler and colleagues (2021) demonstrated that people are satisfied with the offender's moral change even without punishment. As the authors wrote, "…punishment is not a prerequisite for victims' satisfaction with offender change—and does not even contribute much to it." This suggests that even if there are two principal motives for punishment, empowerment may be the more important (Hechler et al. 2021, p. 1029).

Moral Education Theory of Punishment

The fact that people care about the moral change of the offender could point us back in the direction of Duff's communicative account. But there is a simpler alternative.

Before Jean Hampton developed her status-based account, she defended the moral education theory, which claimed that the *only* justification for punishment is the *moral education* of the punished individual and of society (Hampton 1984).[11] The idea is as simple as it is controversial, despite strong endorsements from the likes of Plato and Hegel. Punishment involves threatening people with pain, also known as hard treatment, if they fail to follow the law. In that regard, "[p]unishments are like electrified fences." (Hampton 1984, p. 212). A cow learns to stay on the pasture, and we learn to obey the law. However, unlike an electrified fence, the law also provides us with *moral* reasons to obey it. On this view, punishment is an infliction of pain for moral reasons *designed to teach a moral lesson* about what is right and wrong. We want punishment to get the wrongdoer "to reflect on the moral reasons for [punishment], so that he will make the decision to reject the prohibited action for moral reasons, rather than for the self-interested reason of avoiding pain." (Hampton 1984, p. 212)

A roughly similar theory has been proposed by Herbert Morris (1981). He similarly argued that "a principal justification for punishment is the potential and actual wrongdoer's good." (Morris 1981, p. 264). The good in question is "essentially one's identity as a morally

[11] A roughly similar theory has been developed by Herbert Morris (1981).

autonomous person attached to the good…it is good for the person, and essential to one's status as a moral person, that the evil underlying wrongdoing be comprehended (…) in the way remorse implies comprehension of evil caused…It is a moral good (…) that one feel contrite, that one feel the guilt that is appropriate to one's wrongdoing, that one be repentant, that one be self-forgiving, and that one have reinforced one's conception of oneself as a responsible being." (Morris 1981, p. 265). Punishment is justified as a way "to further the realization of this moral good." (Morris 1981, p. 265). One's status as a moral being is "a non-waivable, non-forfeitable, non-relinquishable right" (Morris 1981, p. 270). We owe it to the offenders to treat them as such moral beings, even if they rather we wouldn't.

The Happy Marriage

Clearly, both moral education theories capture the essential psychological intuitions discussed above. It appears that we need a theory that combines them with the status-based aspects. A short reflection suffices to see that the two philosophical accounts are not in conflict.[12] Any punishment that seeks to teach a *moral lesson* about why the crime was wrong will also, necessarily, communicate the true value of the victim, and deny the insulting message of the crime. In Hampton's terms, a well-crafted retributive response contains "the expressive elements that both vindicate the value of the victim and also act like radioactive elements inside the heads of the abusers, killing their taste for disrespect

[12] Jean Hampton suggested as much in a 1998 article (Hampton 1998).

and domination." (Hampton 1998, p. 44). This view, a combination of status-based expressivism and moral education, seems to map most closely onto lay punitive intuitions.[13]

Summary

Nahmias and Aharoni failed to demonstrate that Duff's communicative account works best as a *descriptive* theory of lay intuitions. However, their findings are perfectly in line with status-based theories, and moral education theories. The fact that sincere apologies decrease punitiveness fits the latter (Nahmias & Aharoni, 2017).[14] The fact that most people did not change their initial sentence recommendations in Aharoni's more recent studies fits status-based theories very well (Aharoni et al., 2022). Punishment that is too lenient does not reestablish the victim's worth.

Psychological research suggests that our intuitions are best captured by a combination of status-based theories and moral education. We intuitively punish to teach a moral lesson, and expressing the true worth of the victim is part of that. It is important to recognize, however, that most philosophical accounts of punishment capture *some* of our intuitions. There is a reason why many people explicitly endorse deterrence theory (e.g. Carlsmith et

[13] Herbert Morris also recognized elsewhere that wrongdoing implies a claim of superiority of sorts, when he wrote that people guilty of breaking community norms place "themselves in a position of superiority to others who have complied with the norms." (Morris 1988, p. 64)

[14] Recall that the sincerity of apology is, for Duff, essentially not the state's business.

al. 2002) or why simple retributive thinking is attractive. Same goes for restorative justice or rehabilitation.

Finally, we must keep in mind that while *intuitiveness* is a virtue of a philosophical theory, it's not the only one.

References

Aharoni, E., & Fridlund, A. J. (2012). Punishment without reason: Isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and Law*, *18*(4), 599–625. https://doi.org/10.1037/a0025821

Aharoni, E., Simpson, D., Nahmias, E., & Gollwitzer, M. (2022). A painful message: Testing the effects of suffering and understanding on punishment judgments. *Zeitschrift Für Psychologie*, *230*, 138–151. https://doi.org/10.1027/2151-2604/a000460

Aquino, K., & Douglas, S. (2003). Identity threat and antisocial behavior in organizations: The moderating effects of individual differences, aggressive modeling, and hierarchical status. *Organizational Behavior and Human Decision Processes*, *90*(1), 195–208. https://doi.org/10.1016/S0749-5978(02)00517-4

Augustine, S. (2008). *The Confessions* (H. Chadwick, Trans.). Oxford University Press.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, *38*(11), 1145–1160. https://doi.org/10.1037/0003-066X.38.11.1145

Barash, D. P., & Lipton, J. E. (2011). *Payback: Why We Retaliate, Redirect Aggression, and Take Revenge*. Oxford University Press.

Barclay, P. (2011). Competitive helping increases with the size of biological markets and invades defection. *Journal of Theoretical Biology*, *281*(1), 47–55. https://doi.org/10.1016/j.jtbi.2011.04.023

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior*, *34*(3), 164–175. https://doi.org/10.1016/j.evolhumbehav.2013.02.002

Baron, J., Gowda, R., & Kunreuther, H. (1993). Attitudes Toward Managing Hazardous Waste: What Should Be Cleaned Up and Who Should Pay for It? *Risk Analysis*, *13*(2), 183–192. https://doi.org/10.1111/j.1539-6924.1993.tb01068.x

Baron, J., & Ritov, I. (1993). Intuitions about penalties and compensation in the context of tort law. *Journal of Risk and Uncertainty*, *7*(1), 17–33. https://doi.org/10.1007/BF01065312

Baron, J., & Ritov, I. (2009). The Role of Probability of Detection in Judgments of Punishment. *Journal of Legal Analysis*, *1*(2), 553–590. https://doi.org/10.1093/jla/1.2.553

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, *103*(1), 5–33. https://doi.org/10.1037/0033-295X.103.1.5

Berlo, D. K. (1960). *The Process of Communication: An Introduction to Theory and Practice*. Holt, Rinehart and Winston.

Biaggio, M. K. (1980). Assessment of anger arousal. *Journal of Personality Assessment*, *44*(3), 289–298. https://doi.org/10.1207/s15327752jpa4403_12

Bilz, K. (2016). Testing the Expressive Theory of Punishment: Testing the Expressive Theory of Punishment. *Journal of Empirical Legal Studies*, *13*(2), 358–392. https://doi.org/10.1111/jels.12118

Boon, S. D., & Yoshimura, S. M. (2020). Revenge as social interaction: Merging social psychological and interpersonal communication approaches to the study of vengeful behavior. *Social and Personality Psychology Compass, 14(9)*, e12554. https://doi.org/10.1111/spc3.12554

Brooks, T. (2012). *Punishment*. London: Routledge.

Brockner, J., De Cremer, D., Fishman, A. Y., & Spiegel, S. (2008). When does high procedural fairness reduce self-evaluations following unfavorable outcomes?: The moderating effect of prevention focus. *Journal of Experimental Social Psychology*, *44*(2), 187–200. https://doi.org/10.1016/j.jesp.2007.03.002

Brockner, J., Heuer, L., Magner, N., Folger, R., Umphress, E., Van Den Bos, K., Vermunt, R., Magner, M., & Siegel, P. (2003). High procedural fairness heightens the effect of outcome favorability on self-evaluations: An attributional analysis. *Organizational Behavior and Human Decision Processes*, *91*(1), 51–68. https://doi.org/10.1016/S0749-5978(02)00531-9

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. https://doi.org/10.1037/0022-3514.75.1.219

Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, *42*(4), 437–451. https://doi.org/10.1016/j.jesp.2005.06.007

Carlsmith, K. M., & Darley, J. M. (2008). Psychological Aspects of Retributive Justice. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 40, pp. 193–236). Elsevier. https://doi.org/10.1016/S0065-2601(07)00004-4

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). *Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment*. *83*, 284–299.

Cassese, E., & Weber, C. (2011). Emotion, attribution, and attitudes toward crime. *Journal of Integrated Social Sciences*, *2*(1), 63–97.

Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, *9*(3), 265–279. https://doi.org/10.1007/s10683-006-9127-z

Cukur, C. S., Guzman, M. R. de, & Carlo, G. (2004). Religiosity, Values, and Horizontal and Vertical Individualism-Collectivism: A Study of Turkey, the United States, and the Philippines. *The Journal of Social Psychology*, *144*(6), 613–634.

Cushman, F. A., Sarin, A., & Ho, M. K. (2019). *Punishment as communication* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/wf3tz

Darley, J. M. (2009). Morality in the Law: The Psychological Foundations of Citizens' Desires to Punish Transgressions. *Annual Review of Law and Social Science*, *5*(1), 1–23. https://doi.org/10.1146/annurev.lawsocsci.4.110707.172335

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior*, *24*(6), 659–683. https://doi.org/10.1023/A:1005552203727

Davis, L. H. (1972). They Deserve to Suffer. *Analysis*, *32*(4), 136–140.

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, *446*(7137), 794–796. https://doi.org/10.1038/nature05651

Duff, R. A. (2003). *Punishment, Communication, and Community*. Oxford University Press.

Duff, R. A. (2022). Punishment as Communication. In *Oxford Handbook of Punishment Theory and Philosophy*. https://papers.ssrn.com/abstract=4171062

Fehr, E., & Fischbacher, U. (2002). Why Social Preferences Matter – the Impact of non-Selfish Motives on Competition, Cooperation and Incentives. *The Economic Journal*, *112*(478), C1–C33. https://doi.org/10.1111/1468-0297.00027

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review*, *90*(4), 980–994. https://www.jstor.org/stable/117319

Fehr, E., Hoff, K., & Kshetramade, M. (2008). Spite and Development. *American Economic Review*, *98*(2), 494–499. https://doi.org/10.1257/aer.98.2.494

Feinberg, J. (1965). The Expressive Function of Punishment. *The Monist*, *49*(3), 397–423. https://doi.org/10.5840/monist196549326

Ferris, D. L., Brown, D. J., & Heller, D. (2009). Organizational supports and organizational deviance: The mediating role of organization-based self-esteem. *Organizational Behavior and Human Decision Processes*, *108*(2), 279–286. https://doi.org/10.1016/j.obhdp.2008.09.001

Fischer, M., Twardawski, M., Strelan, P., & Gollwitzer, M. (2022). Victims need more than power: Empowerment and moral change independently predict victims' satisfaction and willingness to reconcile. *Journal of Personality and Social Psychology*, *123*(3), 518–536. https://doi.org/10.1037/pspi0000291

French, P. A. (2001). *The Virtues of Vengeance*. University Press of Kansas.

Frijda, N. H. (1994). The Lex Talionis: On Vengeance. In S. H. M. van Goozen, N. E. van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 263–289). Psychology Press.

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent. *Personality and Social Psychology Bulletin*, *40*(8), 986–997. https://doi.org/10.1177/0146167214533130

Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1553), 2651–2661. https://doi.org/10.1098/rstb.2010.0135

Gault, B. A., & Sabini, J. (2000). The roles of empathy, anger, and gender in predicting attitudes toward punitive, reparative, and preventative public policies. *Cognition and Emotion*, *14*(4), 495–520. https://doi.org/10.1080/026999300402772

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, *45*, 840–844. https://doi.org/10.1016/j.jesp.2009.03.001

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology*, *41*(3), 364–374. https://doi.org/10.1002/ejsp.782

Gollwitzer, M., Skitka, L. J., Wisneski, D., Sjöström, A., Liberman, P., Nazir, S. J., & Bushman, B. J. (2014). Vicarious Revenge and the Death of Osama bin Laden. *Personality and Social Psychology Bulletin*, *40*(5), 604–616. https://doi.org/10.1177/0146167214521466

Greene, J. (2013). *Moral Tribes: Reason, and the Gap Between Us and Them*. The Penguin Press.

Gummerum, M., Dillen, L. F. V., Dijk, E. V., & Lopez-Perez, B. (2016). Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY*, *65*, 94–104. https://doi.org/10.1016/j.jesp.2016.04.004

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2013). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage Books.

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, *31*(1), 191–221. https://doi.org/10.1111/j.1559-1816.2001.tb02489.x

Hampton, J. (1984). The Moral Education Theory of Punishment. *Philosophy & Public Affairs*, 13(3), 208–238.

Hampton, J. (1988). The Retributive Idea. In *Forgiveness and Mercy* (pp. 111–161). Cambridge University Press.

Hampton, J. (1991). A new theory of retribution. In R. G. Frey & C. W. Morris (Eds.), *Liability and Responsibility* (1st ed., pp. 377–414). Cambridge University Press. https://doi.org/10.1017/CBO9780511527395.013

Hampton, J. (1992). An Expressive Theory of Retribution. In W. Cragg (Ed.), *Retributivism and Its Critics* (pp. 1–25). F. Steiner, Verlag.

Hampton, J. (1998). Punishment, Feminism, and Political Identity: A Case Study in the Expressive Meaning of the Law. *Canadian Journal of Law & Jurisprudence*, *11*(1), 23–45. https://doi.org/10.1017/S0841820900001673

Hampton, J. (2006). Righting Wrongs: The Goal of Retribution. In D. Farnham (Ed.), *The Intrinsic Worth of Persons* (1st ed., pp. 108–150). Cambridge University Press. https://doi.org/10.1017/CBO9780511618239.006

Hanna, N. (2017). The Nature of Punishment: Reply to Wringe. *Ethical Theory and Moral Practice*, *20*(5), 969–976. https://doi.org/10.1007/s10677-017-9835-9

Hanna, N. (2020). The Nature of Punishment Revisited: Reply to Wringe. *Ethical Theory and Moral Practice*, *23*(1), 89–100. https://doi.org/10.1007/s10677-019-10047-1

Hartnagel, T. F., & Templeton, L. J. (2012). Emotions about crime and attitudes to punishment. *Punishment & Society*, *14*(4), 452–474. https://doi.org/10.1177/1462474512452519

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Henberg, M. (1990). *Retribution*. Temple University Press.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, *53*(1), 3–35. https://doi.org/10.1016/S0167-2681(03)00094-5

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367. https://doi.org/10.1126/science.1153808

Hirsch, A. von. (1996). *Censure and Sanctions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198262411.001.0001

Hoffman, M. B. (2014). *The punisher's brain: The evolution of judge and jury*. Cambridge University Press.

Holmes, O. W. (1963). *The Common Law*. Little, Brown,; Company.

Johnson, D. (2009). Anger about crime and support for punitive criminal justice policies. *Punishment & Society*, *11*(1), 51–66. https://doi.org/10.1177/1462474508098132

Kahan, D. (1996). What Do Alternative Sanctions Mean? *University of Chicago Law Review*, *63*(2). https://chicagounbound.uchicago.edu/uclrev/vol63/iss2/4

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus; Giroux.

Kaufman, W. (2013). *Honor and Revenge*. Springer.

Kim, S., & Smith, R. H. (1993). Revenge and Conflict Escalation. *Negotiation Journal*, *9*(1), 37–43. https://doi.org/10.1111/j.1571-9979.1993.tb00688.x

Kłusek, M. (2023). People Want Optimal Deterrence – Just a Little Bit. *Review of Law & Economics*, *19*(1), 99–113. https://doi.org/10.1515/rle-2022-0050

Koper, G., Van Knippenberg, D., Bouhuijs, F., Vermunt, R., & Wilke, H. (1993). Procedural fairness and self-esteem. *European Journal of Social Psychology*, *23*(3), 313–325. https://doi.org/10.1002/ejsp.2420230307

Kuběna, A. A., Houdek, P., Lindová, J., Příplatová, L., & Flegr, J. (2014). Justine Effect: Punishment of the Unduly Self-Sacrificing Cooperative Individuals. *PLOS ONE*, *9*(3), e92336. https://doi.org/10.1371/journal.pone.0092336

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Leary, M. R. (2012). Sociometer Theory. In *Handbook of Theories of Social Psychology* (pp. 141–159). SAGE Publications Ltd. https://doi.org/10.4135/9781446249222.n33

Metz, T. (2000). Censure Theory and Intuitions about Punishment. *Law and Philosophy*, *19*(4), 491–512. https://doi.org/10.1023/A:1026548200351

Miller, D. T. (2001). Disrespect and the Experience of Injustice. *Annual Review of Psychology*, *52*(1), 527–553. https://doi.org/10.1146/annurev.psych.52.1.527

Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, *3*(2), 200–207. https://doi.org/10.1177/1948550611415695

Molnar, A., Chaudhry, S., & Loewenstein, G. (2020). *"It's Not About the Money. It's About Sending a Message!" Unpacking the Components of Revenge* [{SSRN} {Scholarly} {Paper}]. https://doi.org/10.2139/ssrn.3541450

Monin, B. (2007). Holier than me? Threatening social comparison in the moral domain. *Revue Internationale de Psychologie Sociale*, *20*(1), 53–68.

Morris, H. (1981). A Paternalistic Theory of Punishment. *American Philosophical Quarterly*, 18(4), 263–271.

Morris, H. (1988). The Decline of Guilt. *Ethics*, *99*(1), 62–76.

Murphy, J. G., & Hampton, J. (1988). *Forgiveness and Mercy* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511625121

Nahmias, E., & Aharoni, E. (2017). Communicative Theories of Punishment and the Impact of Apology. In C. Surprenant (Ed.), *Rethinking Punishment in the Era of Mass Incarceration* (1st ed., pp. 144–159). Routledge.

Nietzsche, F. (1989). *On the Genealogy of Morals & Ecce Homo* (W. Kaufmann, Ed.; W. Kaufmann & R. Hollingdale, Trans.). Vintage Books.

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

Noë, R., & Hammerstein, P. (1995). Biological markets. *Trends in Ecology & Evolution*, *10*, 336–339.

Norris, F. H., & Kaniasty, K. (1994). Psychological distress following criminal victimization in the general population: Cross-sectional, longitudinal, and prospective analyses. *Journal of Consulting and Clinical Psychology*, *62*(1), 111–123. https://doi.org/10.1037/0022-006X.62.1.111

Okimoto, T. G., & Wenzel, M. (2008). The symbolic meaning of transgressions: Towards a unifying framework of justice restoration. In *Advances in Group Processes* (Vol. 25, pp. 291–326). Emerald (MCB UP). https://doi.org/10.1016/S0882-6145(08)25004-6

Paál, T. (2021). Strong Reciprocity. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 8019–8022). Springer International Publishing. https://doi.org/10.1007/978-3-319-19650-3_1632

Paál, T., & Bereczkei, T. (2015). Punishment as a Means of Competition: Implications for Strong Reciprocity Theory. *PLOS ONE*, *10*(3), e0120394. https://doi.org/10.1371/journal.pone.0120394

Parks, C. D., & Stone, A. B. (2010). The desire to expel unselfish members from the group. *Journal of Personality and Social Psychology*, *99*(2), 303–310. https://doi.org/10.1037/a0018403

Pizarro, D. (2006). Hodgepodge Morality. In J. Brockman (Ed.), *2006: What is Your Dangerous Idea*. Harper Perennial.

Pleasant, A., & Barclay, P. (2018). Why Hate the Good Guy? Antisocial Punishment of High Cooperators Is Greater When People Compete To Be Chosen. *Psychological Science*, *29*(6), 868–876. https://doi.org/10.1177/0956797617752642

Primoratz, I. (1989). Punishment as Language. *Philosophy*, *64*(248), 187–205. https://doi.org/10.1017/s0031819100044478

Raihani, N. J., & McAuliffe, K. (2012). Human punishment is motivated by inequity aversion, not a desire for reciprocity. *Biology Letters*, *8*(5), 802–804. https://doi.org/10.1098/rsbl.2012.0470

Ray, C. M., & Kort-Butler, L. A. (2020). What you see is what you get? Investigating how survey context shapes the association between media consumption and attitudes about crime. *American Journal of Criminal Justice*, *45*(5), 914–932. https://doi.org/10.1007/s12103-019-09502-7

Roberts, G. (1998). Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 427–431.

Robinson, P. H. (2013). *Intuitions of Justice and the Utility of Desert*. Oxford University Press.

Robinson, P., & Darley, J. (2007). Intuitions of Justice: Implications for Criminal Law and Justice Policy. *Southern California Law Review*. https://scholarship.law.upenn.edu/faculty_scholarship/144

Rousseau, J.-J. (1913). *Discourse on the Origin of Ineqaulity Among Men* (G. D. H. Cole, Trans.). https://en.wikisource.org/wiki/Discourse_on_the_

Sarin, A., Ho, M. K., Martin, J. W., & Cushman, F. A. (2021). Punishment is Organized around Principles of Communicative Inference. *Cognition*, *208*, 104544. https://doi.org/10.1016/j.cognition.2020.104544

Scheff, T. J. (2003). Shame in Self and Society. *Symbolic Interaction*, *26*(2), 239–262. https://doi.org/10.1525/si.2003.26.2.239

Scheff, T. J. (2019). *Bloody Revenge: Emotions, Nationalism, and War* (1st ed.). Routledge. https://doi.org/10.4324/9780429038792

Schroth, H. A., & Pradhan Shah, P. (2000). Procedures: Do we really want to know them? An examination of the effects of procedural justice on self-esteem. *Journal of Applied Psychology*, *85*(3), 462–471. https://doi.org/10.1037/0021-9010.85.3.462

Scobie, E. D., & Scobie, G. E. W. (1998). Damaging Events: The Perceived Need for Forgiveness. *Journal for the Theory of Social Behaviour*, *28*(4), 373–402. https://doi.org/10.1111/1468-5914.00081

Seip, E., Dijk, W. van, & Rotteveel, M. (2014). Anger motivates costly punishment of unfair behavior. *Motivation and Emotion*, *38*. https://doi.org/10.1007/s11031-014-9395-4

Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology*, *94*(1), 116–132. https://doi.org/10.1037/0022-3514.94.1.116

Singelis, T. M., Triandis, H. C., Bhawuk, D. P. S., & Gelfand, M. J. (1995). Horizontal and Vertical Dimensions of Individualism and Collectivism: A Theoretical and Measurement Refinement. *Cross-Cultural Research*, *29*(3), 240–275. https://doi.org/10.1177/106939719502900302

Sjöström, A., Magraw-Mickelson, Z., & Gollwitzer, M. (2018). What makes displaced revenge taste sweet: Retributing displaced responsibility or sending a message to the original perpetrator? *European Journal of Social Psychology*, *48*(4), 490–506. https://doi.org/10.1002/ejsp.2345

Skillen, A. J. (1980). How to Say Things with Walls. *Philosophy*, *55*(214), 509–523. https://www.jstor.org/stable/3750319

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Sloman, S. A. (2002). Two systems of reasoning. In D. Kahneman, T. Gilovich, & D. Griffin (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.024

Solomon, R. C. (1990). *A Passion for Justice: Emotions and the Origins of the Social Contract*. Addison-Wesley.

Stephen, J. F. (1883). *A history of the criminal law of England* (Vol. 2). Macmillan.

Strelan, P., Van Prooijen, J., & Gollwitzer, M. (2020). When transgressors intend to cause harm: The empowering effects of revenge and forgiveness on victim well-being. *British Journal of Social Psychology*, *59*(2), 447–469. https://doi.org/10.1111/bjso.12357

Sunstein, C. R., Schkade, D., & Kahneman, D. (2000). Do People Want Optimal Deterrence? *The Journal of Legal Studies*, *29*(1), 237–253. https://doi.org/10.1086/468069

Sylwester, K., & Roberts, G. (2010). Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters*, *6*(5), 659–662. https://doi.org/10.1098/rsbl.2010.0209

Twardawski, M., & Hilbig, B. E. (2020). The motivational basis of third-party punishment in children. *PLOS ONE*, *15*(11), e0241919. https://doi.org/10.1371/journal.pone.0241919

Twardawski, M., Tang, K. T., & Hilbig, B. E. (2020). Is it all about retribution? The flexibility of punishment goals. *Social Justice Research*, *33*, 195–218.

Van Dijk, E., Molenmaker, W. E., & De Kwaadsteniet, E. W. (2015). Promoting cooperation in social dilemmas: The use of sanctions. *Current Opinion in Psychology*, *6*, 118–122. https://doi.org/10.1016/j.copsyc.2015.07.006

Vidmar, N. (2001). Retribution and revenge. In J. Sanders & V. L. Hamilton (Eds.), *Handbook of Justice Research in Law* (pp. 31–63). Kluwer Academic Publishers.

Wojciszke, B., & Struzynska-Kujalowicz, A. (2007). Power influences self-esteem. *Social Cognition*, *25*, 472–494. https://doi.org/10.1521/soco.2007.25.4.472

Wolff, J. (2011). *Ethics and Public Policy* (1st ed.). Routledge.

Wringe, B. (2013). Must Punishment Be Intended to Cause Suffering? *Ethical Theory and Moral Practice*, *16*(4), 863–877. https://doi.org/10.1007/s10677-012-9388-x

Wringe, B. (2016). *An Expressive Theory of Punishment*. Palgrave Macmillan UK. https://doi.org/10.1057/9781137357120

Wringe, B. (2019). Punishment, Jesters and Judges: A Response to Nathan Hanna. *Ethical Theory and Moral Practice*, *22*(1), 3–12. https://doi.org/10.1007/s10677-018-9966-7

Zdaniuk, A., & Bobocel, D. R. (2012). Vertical individualism and injustice: The self-restorative function of revenge: Vertical Individualism and Revenge. *European Journal of Social Psychology*, *42*(5), 640–651. https://doi.org/10.1002/ejsp.1874

Zizzo, D. J. (2003). Money burning and rank egalitarianism with random dictators. *Economics Letters*, *81*(2), 263–266. https://doi.org/10.1016/S0165-1765(03)00190-3

Zizzo, D. J., & Oswald, A. J. (2001). Are People Willing to Pay to Reduce Others' Incomes? *Annales d'Économie Et de Statistique*, *63/64*, 39–65. https://doi.org/10.2307/20076295