



# Desert Retributivism: A Deweyan Critique

Andrei Poama<sup>1</sup>

Received: 24 September 2020 / Accepted: 1 February 2023 / Published online: 15 February 2023  
© The Author(s) 2023

## Abstract

In this article, I argue that Michael Moore's (1997), and other similar formulations of desert retributivism – viz., the theory that holds punishment to be justified because of the deserved suffering it imposes on guilty offenders – are epistemically problematic. The argument draws on John Dewey's inchoate critique of retribution, and on Dewey's more general contention that the justification of ethical judgments and principles proceeds *ex post* – viz., that it depends on the experiences elicited by acting on those judgments and principles. Based on Dewey's *ex post* take on justification, I more specifically argue that, given its commitment to moral naturalism and to coherentism, Moore's version of desert retributivism is epistemically unwarranted. This is because we have evidentially grounded reasons to think that many of the retributive judgments that underlie the principle of retributive justice are not clearly supported *ex post* in the relevant, desert retributivist sense.

**Keywords** Retributivism · Desert retributivism · Ex post justification · John Dewey · Moral naturalism · Coherentism · Michael S. Moore

Whether a judgment or principle justifies an action or practice depends, to a significant extent, on the experiences elicited by acting on that judgment or principle. Call this the *practical test*. John Dewey takes the practical test to be central to moral epistemology in general, and to the justification of coercive government policies in particular. Terminological differences to the side, Dewey's commitment to the practical test is widely recognized.<sup>1</sup> The test has been used by Dewey's supporters to make

---

<sup>1</sup> See Pappas (1997), Welchman (1997), Anderson (2023).

---

✉ Andrei Poama  
a.poama@fgga.leidenuniv.nl

<sup>1</sup> Faculty of Governance and Global Affairs, Leiden University, Den Haag, Netherlands

sense of why some morally repugnant practices (e.g., slavery<sup>2</sup>) have been abandoned in the past or to advance arguments that favor specific policies (e.g., urban integration<sup>3</sup>) for the future.

In this article, I bring the practical test to bear on desert retributivism, which is one of the most prominent justifications of state punishment.<sup>4</sup> Desert retributivism, in the version I consider, argues that retributive justice is grounded in particular judgments that guilty wrongdoers deserve to suffer and that, thus construed, retributive justice justifies punishment. My argument proceeds in two steps. First, I make explicit Dewey's use of the practical test in his critique of retribution. I then argue that, although Dewey's application of the critique begs the question, it can be suitably applied to desert retributivism. This is because, unlike the version of retributivism targeted by Dewey, desert retributivism posits a theory of justification that renders it sensitive to the practical test.<sup>5</sup>

The article is structured as follows. In Section I, I briefly reconstruct Dewey's critique of retribution, show how it inchoately relies on the practical test, and explain why Dewey's anti-retributive critique is question-begging. In Section II, I offer a close characterization of desert retributivism and its underlying theory of justification. In Sections III and IV, I apply Dewey's practical test to desert retributivism – first, by clarifying why the test aptly applies to it (Section III) and, second, by arguing that desert retributivism fails the practical test (Section IV). In Section V, I consider some objections to the Deweyan critique. Section VI concludes.

## 1 Dewey's Critique of Retribution

John Dewey's critique of retribution draws on his more general question about whether and, if so, how coercive force can be justified within legally sanctioned practices – in particular, war and punishment (1916). To answer this question, Dewey picks a mid-point between two justificatory extremes: “the Tolstoian, to whom all force is violence and all violence is evil” (1916: 295) – viz., the pacifist position on matters of war and the abolitionist position on penal matters – and those wedded to “the glorification of force” – viz., positions that defend force in the name of a general moral end or principle, such as the “common will and consciousness” (1916: 295). Dewey argues that those who embrace the first extreme ignore that principles do not

<sup>2</sup> Anderson (2014).

<sup>3</sup> Anderson (2010).

<sup>4</sup> Hereafter, I use the terms *justification*, *judgment* and *principle* in their ethical sense.

<sup>5</sup> Importantly, the argument presented here does not constitute an overall rejection of retributivism. Nor is it obvious that such a rejection can be advanced. This is because retributivism covers a heterogenous variety of views that differ both substantively (in terms of their specification of retribution), and theoretically (in terms of the theoretical desiderata – in particular, epistemological and methodological ones – that a retributive account needs to satisfy). Cottingham (1979) makes an inventory of nine varieties of retributivism and Walker (1999) later adds five varieties to the list. Even so, desert retributivism seems to be *distinctively* retributive not least because of its self-professed contrast with other justifications of punishment. In this sense, undermining the case for desert retributivism has broader implications the normative theory (or theories) of punishment that we should take to be conclusively warranted.

exclude, but often require at least *some* use of force; those who side with the second extreme ignore that principles cannot *as such* guarantee that force is justified. To put it in Dewey's terms:

No ends are accomplished without the use of force. It is consequently no presumption against a measure, political, international, jural, economic, that it involves a use of force. Squeamishness about force is the mark not of idealistic but of moonstruck morals. But *antecedent and abstract principles cannot be assigned to justify the use of force*. The criterion of value lies in the *relative efficiency and economy of the expenditure of force as a means to an end*. With advance of knowledge, refined, subtle and indirect use of force is always displacing coarse, obvious and direct methods of applying it. This is the explanation to the ordinary feeling against the use of force. What is thought of as brutal, violent, immoral, is a use of physical agencies which are gross, sensational and evident on their own account, in cases where it is possible to employ with greater economy and less waste means which are comparatively imperceptible and refined. (Dewey 1916: 364; *emphases added*)

This passage cautions against two mistakes. The first mistake is to deny *that* force matters for realizing at least some principles. The second mistake consists in an inadequate specification of *how* force matters for assessing principles— viz., it ignores that how we come to experience force when it is exerted on the basis of a principle has normative import *for* that principle. To see this as a mistake, we need to see that justification cannot proceed only on an *ex ante* basis, but *also* on an *ex post* basis: “in no case,” notes Dewey, “can antecedent or a priori principles be appealed to as more than presumptive: the point at issue is concrete utilization of means for ends.” (1916: 367; *emphasis added*). This, and other equivalent formulations, capture the crux of the practical test – viz., that whether a principle is justified depends on the experiences elicited by acting on it, not only on the experiences which initially generate it.<sup>6</sup>

The claim that justification proceeds *ex post* is central to Dewey's ethics in particular, and to ethical pragmatism in general.<sup>7</sup> At its core, it holds that principles are not absolute, but subject to revision, and that some of the main reasons for revision reside in the experiential impact principles bring about. On this view, the ethical cogency of a principle is *conditional* on the experiences it has for those who act on or who are affected by those who act on it. An *ex post* standard of justification is thus tied to a moral epistemology which denies that principles can be grounded on a purely *a priori* or strictly *ex ante* basis, and holds that principles are *both* grounded (or formed) *and* tested (or evaluated) *a posteriori* – viz., based on concrete practical experiences.<sup>8</sup>

<sup>6</sup> Take, for instance, Dewey's contention that “an end is something which concerns results rather than aspirations.” (Dewey 1916: 296).

<sup>7</sup> See fn. 1–3 for the relevant references.

<sup>8</sup> The *a priori/a posteriori* and *ex ante/ex post* distinctions cover conceptually different territory. *A priori* justificatory accounts do not grant any justificatory import to experience, whereas *a posteriori* justificatory accounts give experience an essential justificatory role. The *ex ante/ex post* distinction is best (albeit not necessarily) construed as an additional distinction between two different modes of *a posteriori* justification: *ex ante* justification refers to the justificatory role experiences play in the *formation* of a

Dewey takes the *ex post* standard of justification as an indication that retribution is an inadequate penal principle:

Upon the whole, the opinion seems to be current that in [penal] matters force is hallowed by the mere fact that it is the State which employs it, or by the fact that it is exercised in the interests of “justice” – *retribution in the abstract*, or what is politely called “vindicating the law.” *When the justification of force is sought in some kind of abstract consideration of this sort, no questions are to be raised about the efficiency of the force used, for it is not conceived as a specific means to a specific end.* It is the sacrosanct character thus attributed to the State’s use of force which gives pungency to the Tolstoian charge that the State is the arch-criminal, the person who has recourse to violence on the largest scale. I see no way out except to say that *all depends upon the efficient adaptation of means to ends.* The serious charge against the State is not that it uses force – nothing was ever accomplished without using force – but *that it does not use it wisely or effectively.* Our penal measures are still largely upon the level which would convince a man by knocking him down instead of by instructing him. (Dewey 1916: 363–364; *emphases added*)

Like in the previous passage, Dewey’s use of the term *efficiency* is somewhat *sui-generis*. Its meaning is not strictly economic, but more generally normative – viz., the efficiency of a coercive practice is to be construed in terms of the practical relation between the expected value professed by the principle purportedly justifying that practice – in this case, retribution – and the actual value produced by the principle when applied in practice – in this case, punishment. On this understanding of efficiency, acting on a principle is efficient if the coercive action undertaken on its basis secures *ex post* the value or values the principle posits *ex ante*. A principle might, in this sense, turn out to be inefficient either absolutely – viz., if acting on it tends to engender *some* disvalue in addition to the value that it secures – or relatively – viz., if acting on it comparatively engenders more disvalue than value.

Dewey’s critique of retributive punishment *qua* inefficient practice is more suggestive than systematic, but its main contention seems to be that we have reasons to reject retribution because of its relative inefficiency – viz., because acting on it generates more disvalue than value. The ratio between value and disvalue is something we can assess *ex post*, on the basis of our experiences elicited from punishing retributively:

That punishment is suffering, that it inevitably involves pain to the guilty one, there can be no question; this, whether the punishment is externally inflicted or is in the pangs of conscience, and whether administered by parent, teacher, or

---

general principle, whereas *ex post* justification refers to the role experiences play in the *evaluation* of that principle. In this sense, *ex ante* justification can be said to be constructive and *ex post* justification critical. This specification of the *ex ante/ex post* distinction is arguably the most fitting one for Dewey’s overall pragmatist account of justification – after all, Dewey talks about “antecedent or a priori principles” (Dewey 1916: 367). For a recent discussion defending the idea that *ex post* justification is best understood as a distinct mode of a *posteriori* justification, see Neta (2017).

civil authority. But that suffering is for the sake of suffering, or that suffering can in any way restore or affect the violated majesty of law, is a different matter. (...) Those who are responsible for the infliction of punishment have, as well as those punished, to meet the requirements of justice; and failure to employ the means and instrumentalities of punishment in a way to lead, so far as possible, the wrongdoer to reconsideration of conduct and re-formation of disposition, cannot shelter itself under the plea that it vindicates law. Such failure comes rather from thoughtless custom; from a lazy unwillingness to find better means; from an admixture of pride with lack of sympathy for others; from a desire to maintain things as they are rather than go to the causes which generate criminals. (Dewey 1932/2022: 417)

The target of Dewey's critique is not any one specific retributive conception, but a particular *kind* of conception – viz., one whose underlying theory of moral justification is purely *a priori*. For such conceptions, the experiences elicited by acting on its principles are entirely irrelevant for assessing whether they are value-tracking or value-realizing. Plausibly, these are the kinds of conceptions held by Kant or some of the British idealists, for whom ethical justification is not a matter of *a posteriori* experience, but one of *a priori* grounding by reason alone.<sup>9</sup>

Dewey's critique, then, hinges on his wider rejection of *a priori* justification, not just only on his more specific concerns with *ex ante* justification. If this is correct, his critique of retribution discards *all* penal theories that take justification to be a matter of *a priori* reasoning alone, not only retributive ones. Because of this, supporters of retributive justice can simply argue that Dewey's critique begs the question: the critique simply posits that justification proceeds on an *a posteriori* basis, and thus entirely avert the *ex post* justification demand. Even so, the failure of Dewey's critique remains instructive, since it suggests that retributivist accounts which justify punishment on (partly) *a posteriori* grounds should be sensitive to the practical test, unless they have an independent argument for rejecting the *ex post* justification demand. Michael Moore's desert retributivism is one such account (1997).

## 2 Desert Retributivism: Justificatory Commitments & Critiques

Moore grounds punishment in the value of retributive justice<sup>10</sup>, which he derives from a “psychological fact” (Moore 1997: 105) that consists in our putatively widespread judgments that it is just (or good<sup>11</sup>) that guilty offenders are made to suffer by being punished for their offenses. Moore offers some evidence for this psychological

<sup>9</sup> For a discussion of Kantian, and more generally idealist accounts of retributive justice, see Brooks (2010). For a systematic defense that normative principles are justified *a priori*, see Boghossian (2021).

<sup>10</sup> Moore's articulation of desert retributivism belongs to a philosophically wider family, which includes Davis (1972), Kleinig (1973), Ten (1987: 79–81), and Berman (2008). Insofar as they share the same (or sufficiently similar) justificatory commitments, my critique is meant to apply to the entire family of accounts, not just Moore's.

<sup>11</sup> Given his deontological understanding of both predicates, I do not take issue with Moore's indistinct use of the predicates *good* and *just*.

fact through thought experiments that depict criminal offenses – in particular, serious offenses such as rape or murder. One such thought experiment portrays a modified version of the case in *State v. Chaney* (1970), where the defendant and his friend robbed, beat and raped the victim several times.

The thought experiment that such a case begins to pose us is as follows: Imagine in such a case that the defendant after the rape but before sentencing has got into an accident so his sexual and aggressive desires are dampened to such an extent that he presents no further danger of violence against women; if money was also one of his problems, suppose further that he has inherited a great deal of money, so that he no longer needs to rob. Suppose, because of both of these facts, we are reasonably certain that he does not present a danger of either forcible assault, rape, or robbery or related crimes in the future. Since Chaney is (by hypothesis) not dangerous, he does not need to be incapacitated, specially deterred, or reformed. Suppose further that we could successfully pretend to punish Chaney, instead of actually punishing him, and that no one will find out. (Moore 1997: 101–102)

Another, equally gruesome, but fictitious example is that of the nobleman in Dostoyevsky's *Brothers Karamazov*, who lets his dogs loose so that they savagely kill a boy in front of the boy's mother. The corresponding thought experiment runs as follows:

Imagine that circumstances are such that no non-retributive purpose would be served by punishing this offender. Now imagine two variations: (1) you are that offender; (2) someone else is that offender. Question: should you or the other offender be punished, even though no other social good will thereby be achieved? The retributivist's 'yes' runs deep for most people. (Moore 1997: 163)

The function of these, and other similar thought experiments is twofold. The first one is to modify the corresponding real-life cases such that consequentialist considerations – for instance, prevention, incapacitation or rehabilitation – are rendered practically irrelevant. The second function is to tease out our moral experiences of such cases, such that, building on the judgments informed by these experiences, we can properly infer general penal principles. Thought experiments thus serve as evidence collection methods meant “to sort one's true reasons for believing that certain propositions are true” (Moore 1997: 94) – viz., they are used to assess whether there is any evidence that distinctively supports a retributive justice account *if* we bracket the evidence that is distinctively relevant for consequentialist accounts.

Moore's retributivism is committed to moral naturalism for its justification of particular judgments, and to coherentism for the justification of the principles derived from those judgments. Moral naturalism holds that cogent beliefs are obtained *a posteriori* – here, through actual experiences prompted by pertinent (actual or counterfactual) criminal offense cases. Moore's own “naturalized version of moral

epistemology” (Moore 1997: 131)<sup>12</sup> focuses on the *emotional* experiences elicited by the kind of thought experiments listed above. It claims that generally widespread emotions such as guilt and indignation indicate that retribution – which he construes as the proposition that *guilty offenders deserve to suffer* – is true.<sup>13</sup> Like emotional experiences in general, guilt and indignation “are heuristic guides for us, an extra source of insight into moral truths beyond the knowledge we can gain from sensory and inferential capacities alone” (Moore 1997: 132).

Moore’s coherentism further holds that the principles whereby we justify a practice – in this case, punishment – should be inferred inductively or abductively from “those of our more particular judgments that we also believe to be true” (Moore 1997: 106) rather than derived from more general principles.<sup>14</sup> Given the putatively widespread nature of the emotional experiences of guilt and indignation as reactions to criminal offenses, and given that these emotions ground retributive judgments – viz., judgments that guilty offenders deserve to suffer – Moore argues that the principle of retributive justice – viz., the principle that “punishment is justified if it is given to those who deserve it” (Moore 1997: 154) coherently captures a moral truth. This is the crux of Moore’s desert retributivism, which can be schematically rendered as follows:

$$DR \rightarrow Ds1 \ \& \ Ds2 \ \& \ \dots \ Dsn.$$

Here, *DR* refers to desert retributivism, which is the general proposition (and corresponding principle) that guilty offenders deserve to suffer for their offenses. *Dsn* refers to particular judgments that individual offenders deserve to suffer for their individual offenses – for instance, on judgment *Ds1*, *Offender 1* deserves to suffer for *Offense 1*.<sup>15</sup>

*DR* is grounded by, and therefore implies *Dsn* – viz., the way we establish the truth of *DR* is by knowing *Dsn* to be true. Thus, the truth of *DR* is reducible to the truth of *Dsn*. Furthermore, and relatedly, since there are enough cases like *Dsn*, *DR* is taken to be sufficient for justifying punishment – viz., it tracks a moral value (i.e., offenders’ deserved suffering) worth realizing in cases that prompt *Dsn*, even absent other moral

<sup>12</sup> Moore provides a concise comparison between his theory of justification with other contending theories – most notably, rationalistic theories (e.g., for Moore, Kant’s theory of justification), conventionalist theories (e.g., for Moore, Lord Devlin’s account of morality), and intuitionist theories (e.g., arguably, Ross’s account of morality) – see Moore (1997: 132–133).

<sup>13</sup> In particular, Moore takes the emotion of guilt to indicate that guilty offenders deserve to suffer when we envisage how offenders should be treated from a first personal perspective (i.e., from the offender’s standpoint), and the emotions of moral outrage and indignation to indicate that guilty offenders deserve to suffer when we envisage how offenders should be treated from a third personal perspective.

<sup>14</sup> Moore explicitly equates this account of justification with John Rawls’ reflective equilibrium approach, whereby we can “justify a moral principle by showing that it best accounts for those of our more particular judgements that we also believe to be true” (Moore 1997: 106). Note that, like reflective equilibrium, pragmatist accounts of justification insist on the experiential nature of particular judgments and on the revisability of both particular judgments and general principles in light of morally relevant experiences and evidence. On the similarities and differences between reflective equilibrium and pragmatist accounts of justification, see Anderson (2015).

<sup>15</sup> *DR* is inductively or abductively inferred from *Dsn* (see Moore 1997: 162). The implication relation denotes that the truth of *DR* depends on the truth of *Dsn*.



values that are not or could not be realized in those cases (e.g., offense prevention or offender rehabilitation).

Moore defends his retributive account against a series of objections. Some of these objections target its coherentism. One such objection is that, since the inference in  $(DR \rightarrow Ds1 \ \& \ Ds2 \ \& \ \dots \ Dsn)$  is circular,  $DR$  is empty. Call this the circularity objection. Another objection is that the consequent of in  $(DR \rightarrow Ds1 \ \& \ Ds2 \ \& \ \dots \ Dsn)$  cannot be consistently extended to  $Dnon-s1'$ ,  $Dnon-s2'$  &...  $Dnon-sn'$  – viz., to pervasively held judgments that people do not deserve to suffer. Given the *prima facie* inconsistency between  $Dsn$  and  $Dnon-sn'$ , securing the truth of  $Dsn$  requires us to show that  $Dnon-sn'$  is false, which seems implausible. Call this the inconsistency objection.

The circularity objection is that, if the justification of  $Dsn$  cases is provided by  $DR$  and given that  $DR$  is not derived from any other general principle,  $DR$  lacks content – viz., it purports to justify the  $Dsn$  which are meant to justify it. In short, it posits that  $(DR \rightarrow Ds1 \ \& \ Ds2 \ \& \ \dots \ Dsn)$  is equivalent to  $(DR \rightarrow DR)$ . Moore discards this objection by clarifying the role that  $Dsn$  and  $DR$  play in his retributive account. As indicated,  $Dsn$  have a strictly evidential role – viz., they capture particular data points from which we can infer a more general principle. This means that  $Dsn$  do not rely on  $DR$  for their content.

To use Moore's analogy, the relation between particular retributive judgments and a general retributive principle is not more viciously circular than the relation between “a general belief (for example, that all emeralds are green)”, and the “perceptual beliefs (that) numerous people at numerous times have formed (...) that an item before them was both green and an emerald” (Moore 1997: 109–110). On this analogy, the task of  $DR$  is not to establish its truth, but to offer the best – viz., on Moore's view, the most coherent – account for true  $Dsn$  judgments. Put differently, the truth of  $DR$  is secured (or not) at the level of  $Dsn$  judgments, not at the level of its inferential relation to them.

The inconsistency objection notes that we do not *generally* believe that people ought to suffer. Instead, we tend to believe the opposite – viz., that it is bad if people suffer. Since these beliefs conflict with Moore's retributive beliefs, Moore needs to show that the former are generally false before he can infer to the truth of  $DR$ .<sup>16</sup> To this objection, Moore can reply that general *non-retributive* beliefs typically obtain when we observe the suffering of persons who did not wrong anyone. Since retributive desert implies both culpability *and* wrongdoing, when one of these two elements does not obtain, it should be unsurprising that retributive beliefs don't obtain either (Moore 1997: 185–186). The inconsistency objection, then, has bite only if we over-extend the domain of retributive beliefs beyond the realm of culpable wrongdoing.

Other objections are aimed at Moore's moral naturalism. One such objection is that, since  $Dsn$  can be debunked,  $DR$  is false.<sup>17</sup> Call this the debunking objection. The objection is that  $Dsn$  are unreliable because of their psychologically or otherwise dubious origins – for instance, that they are deeply tied to vicious emotions like Nietzschean *ressentiment* (Moore 1997: 137 ff.). This means that, for many  $Dsn$

<sup>16</sup> For different formulations of this objection, see Rosebury (2011), Hanna (2019), and Nussbaum (2016).

<sup>17</sup> For different formulations of this critique, see Wiegman (2017), and Caruso (2020).



judgments, we also hold defeating judgments – call them *Dfsn* – that tend to negate *Dsn*, such that  $[(Dfsn \rightarrow \neg Dsn) \ \& \ Dfsn \rightarrow \neg Dsn]$ . In reply, Moore argues that “the Nietzschean critic (did not) show that retributive judgments are *inevitably* motivated by the dark emotions of *ressentiment*” (Moore 1997: 140). In particular, he suggests that there is scarce evidence, and so not enough reason to assert *Dfsn*. But there is some evidence that corroborates *Dsn* – in particular, evidence provided by experiencing virtuous emotions like guilt (if we counterfactually put ourselves in the offender’s position) or outrage and indignation (if we are directly or testimonially exposed to the offense).<sup>18</sup> If Moore’s evidential claim about *Dsn* is accurate, then  $[(Dfsn \rightarrow \neg Dsn) \ \& \ Dsn] \rightarrow \neg Dfsn$ , and so the debunking objection is discarded in the domain targeted by the evidential claim.

Moore’s rejection of these objections is persuasive if we accept his moral naturalism, and do not overstretch his coherentism. The circularity objection confuses the naturalistic grounding of the truth of particular retributive judgments with the derivation of general retributive principles, and is thus misdirected. The inconsistency objection overextends the reach of the retributive principle. The debunking objection fails to ground the truth of defeating judgments naturalistically, and is thus epistemologically question-begging. This suggests that any non-question-begging critique of *DR* needs to take his naturalism seriously, and to apply his coherentism within the relevant domain – viz., by not asking it to account for experiences or beliefs that are not relevantly connected to retributive judgments. In the following section, I argue that such a critique is available, and use Dewey’s *ex post* practical test to support it.

### 3 The Justificatory Relevance of the Practical Test

The crux of Dewey’s practical test is that the justification of a principle is conditional on the experiences elicited by acting on that principle. To put it in a slogan, the proof of a principle lies in its felt practice. Applied to Moore’s desert retributivism, the test requires that retributive punishment – viz., punishment imposed because we believe that guilty offenders deserve to suffer – elicits experiences that are, in some relevant sense, desirable *and* derived from acting on the retributive principle. Consequently, we should expect that *ex ante* judgments holding that it is intrinsically just if guilty offenders suffer are consistent with *ex post* judgments that it is intrinsically just if guilty offenders suffer. Put differently, if experiences matter *ex ante* for the justification of retributive judgments, experiences should also matter *ex post*, especially if they are generated by acting on the same judgments.

<sup>18</sup> Moore typically provides this evidence *via* thought experimental devices. One of the thought experiments builds on the Bonnie Garland murder case, where Richard Herrin bludgeoned Bonnie Garland with a hammer because she wanted to end their relationship. When confronted with such cases, Moore argues that it is justified to experience intense moral outrage and indignation (if we consider the case from a third personal perspective) or intense moral guilt (if we consider the case for the first personal perspective of the offender – in this case, from Richard Herrin’s standpoint), and that these emotional experiences ground the particular retributive judgments which support the general retributive principle that offenders deserve to suffer for their offenses. For a more detailed discussion, see Moore (1997: 145 ff.).

More specifically, insofar as *ex post* judgments are sensitive to the experiences elicited by retributive punishment, we should expect them to be both substantively different from *ex ante* judgments – viz., we should expect them to be based on updated experiences – and compatible with these *ex ante* judgments. This expectation is built into Moore’s moral naturalism. To see why, take, by way of analogy, the *ex ante* judgment that I need to drink some water, which informs my decision to drink some water. Now ask yourself: how do I ground the truth of this judgment naturalistically? Part of the answer to this question is that I experience an *ex ante* emotion (i.e., thirst) that offers evidence about what I should do, given this experience: the natural experience of thirst informs – *via* a specific judgment – my decision to drink some water.<sup>19</sup> But this answer is largely insufficient when it comes to assessing the truth value of the *ex ante* judgment. An adequate test for whether this judgment is justified is to see if my thirst is eased when I drink some water – viz., to see if there is an *ex post* experience (i.e., the pleasure of eased thirst) that is both substantively different from the *ex ante* experience (i.e., being thirsty) and compatibly connected with it, in that both experiences are corroborated by me drinking some water.

Building on this analogy, and turning to Moore’s desert retributivism, the question concerns the naturalistic grounding of retributive judgments. Part of the answer to this question is that we experience *ex ante* emotions – most relevantly for Moore, indignation and guilt – that inform our retributive judgments. More specifically, considered from the offenders’ first personal perspective, “to feel guilty is to judge that we must suffer” (Moore 1997: 148). Similarly, considered from a third personal perspective, to feel indignant is to judge that offenders must suffer.

But an additional, and at least desirable if not necessary way of assessing whether *ex ante* retributive judgments are justified is to look at the *ex post* experiences prompted by acting on the said judgments. This test fits straightforwardly into Moore’s naturalism, since *ex post* experiences and the judgments based on them are both integral to the moral experience of punishment – viz., they are based on how we feel about offenders being punished – and directly relevant for assessing their truth value.

Moore leaves *ex post* emotions largely unspecified. But the specification seems straightforward: it is plausible to expect that the retributive emotions of guilt or indignation should be suitably gratified if and insofar guilty offenders suffer for their crimes. In this sense, the emotions experienced *ex post* punishment ought to be positive, at least when considered in relation to the *ex ante* emotions that favor punishment. To come back to the previous analogy, the *ex post* experience that I have when drinking water is that of eased thirst, which is a positive experience, at least when compared with the experience of being thirsty. Similarly, the *ex post* experience of offenders being punished ought to be relevantly positive – viz., we should expect some gratification or suitably specified form of satisfaction when offenders suffer penally for their offenses. Considered from a retributive standpoint, it would be psychologically odd if all one could feel when offenders are retributively punished is the

<sup>19</sup> Those who think that construing thirst as an emotion (in particular, as a homeostatic emotion) is conceptually misguided or empirically unsubstantiated can replace this example with a different one – for instance, the prudential belief that I should run away, which I infer from the fear that I feel when I am attacked by a pack of wolves.

same (or some other equivalent) negative emotion of indignation or guilt. Should that be true, retributive punishment would either be gratuitous or, if non-retributive penal practices could better gratify our feelings of guilt and indignation, less justified than these latter practices.<sup>20</sup>

This gratification expectation can be partly inferred from Moore's discussion of cases where innocent people are punished, and of cases where guilty offenders are not punished. In relation to wrongful conviction cases, Moore claims that it is "outrageously unfair to punish an innocent person" (Moore 1997: 118) – viz., that outrage and indignation are fitting emotions when experiencing the innocent being punished. In relation to impunity cases, Moore claims that we experience "emotional outrage when [guilty offenders] do not get [punished]" (Moore 1997: 118). If these two claims are experientially cogent, our experience of guilty offenders being punished should not be one of outrage, indignation or guilt; rather, it should be one of *gratified* (or otherwise suitably satisfied) outrage, indignation or guilt. In short, we should be morally satisfied when guilty offenders suffer.

To cast this expectation formally, one can assert that ( $Dsn \rightarrow Dgsn$ ), where  $Dsn$  refers to *ex ante* judgments based on emotions such as guilt and indignation, and  $Dgsn$  refers to *ex post* judgments based on gratified emotions such as gratified guilt and gratified indignation. Given this, evidence that  $\neg Dgsn$  negates  $Dsn$ , and, since ( $DR \rightarrow Dsn$ ), undermines  $DR$ .

Formulated in Dewey's normative efficiency language outlined in the previous section, the claim is that evidence is adduced in favor of  $DR$  if acting on  $Dsn$  generates positive experiences that justify  $Dgsn$ , and precludes negative experiences that justify  $\neg Dgsn$ .<sup>21</sup> This befits Moore's moral naturalism, whereby the evidence we get about the moral value(s) captured by a principle – here, the value of retributive justice – takes an experiential form – viz., we access the truth of retributive justice *via* specific emotional experiences. On this reading, acting on  $DR$  is justified *qua* efficient if it practically elicits positive experiences otherwise unattained, and minimizes retributively relevant negative experiences.

As indicated in Section I, there are at least two types of standards we might use to assess whether there is enough evidence that counts against  $DR$ . One standard is absolute: it holds that, for  $DR$  to be justified, acting on it should *only* elicit positive experiences like those captured by  $Dgsn$ . On this standard,  $DR$  needs to be *absolutely* efficient – viz., contrasted to other penal principles, its application elicits no negative experiences. This standard is violated if there is *any* evidence that  $\neg Dgsn$ . Another standard is relative: it holds that, for  $DR$  to be justified, it should elicit *more* positive experiences than negative ones. On this standard,  $DR$  needs to be *comparatively* efficient – viz., relative to other penal principles, its application tends to elicit fewer negative experiences. In the following section, I present and discuss the empirical evidence that suggests  $DR$  is *relatively* undermined.

<sup>20</sup> Retributive punishment here refers to penal practices that are deliberately designed and succeed to bring about *only* the suffering of guilty wrongdoers – viz., they do generate any other penally relevant good like crime prevention or rehabilitation.

<sup>21</sup> Remember that, for Dewey, a principle is inefficient if acting on it (or applying it) generates more disvalue than it does value, with both value and disvalue being accessed and assessed experientially – viz., in terms of the positive or negative experiences that acting on that value generates.

#### 4 Putting the Practical Test to Work: Desert Retributivism, and its Ex Post Justification

As mentioned, Moore does not examine retributively relevant *ex post* experiences; nor does he adduce evidence about them. His case for *DR* is thus based on a strictly *ex ante* basis. Remarkably, most empirical studies which conclude that people are penal retributivists are similarly based on *ex ante* evidence. Two general findings come out of these studies. First, when confronted with particular crime cases, individuals tend to reason retributively – viz., they are more sensitive to retributively relevant information (e.g., the seriousness of the criminal offense), as compared to non-retributively relevant information (e.g., the probability to reoffend). Second, individuals tend to decide retributively – viz., they are significantly more likely to favor sanctions which track retributive aims (e.g., giving offenders their deserved suffering), as compared to non-retributive ones (e.g., preventing offenses or rehabilitating offenders).<sup>22</sup>

These studies inform us about the penal justification “in the mind of the punisher” (Darley et al. 2000), and corroborate the evidential claim that underlies Moore’s *DR*, which is that people are *generally* retributive in the desert sense.<sup>23</sup> However, the findings cover only *ex ante* responses about justified punishment – viz., the studies ask people only about the sanctions that they proximately support (and about their reasons for doing so), not *also* about their retrospective support (or about their reasons) for these sanctions. The studies thus remain largely silent on people’s *ex post* responses to retributive punishment.

To address this evidential gap, other studies examined people’s experience when confronted with the retributive punishment favored *ex ante*. Using an experimental set-up where participants are encouraged to cooperate by confederates who subsequently free-ride, Carlsmith, Wilson & Gilbert’s (2008) study found that the free-riders’ victims do not obtain their *expected* psychological satisfaction when and insofar as they impose a strictly retaliatory response to the free-riders. Participants to three different studies were divided into three groups: *punishers* (those who had the opportunity to sanction free-riding confederates), non-punishers (those who did not have the opportunity to punish), and *forecasters* (participants who did not have the opportunity to punish, but were asked to predict their feelings *had* they been able to do so).

The first of these three studies found that the direct infliction of retributive punishment by victims made the victims feel worse – i.e., less satisfied – not better, and that victims were not able to predict such negative feelings *ex ante*. The second study examined how witnesses of punishment (i.e., those who merely observed punishment without imposing it themselves) felt compared to punishers, and found the latter felt worse in the post-punishment condition. The third study replicated the findings of the second one, and further found that punishment witnesses did not feel any better – i.e., they did not experience more closure – than forecasters who did not know whether punishment had been imposed. This latter finding is particularly relevant, as it sug-

<sup>22</sup> Here, see, most notably Darley, Carlsmith, Robinson (2000); Carlsmith, Darley, Robinson (2002); Carlsmith (2006) and, for a more extensive review, Carlsmith & Darley (2008).

<sup>23</sup> Typically, Carlsmith, Darley, Robinson (2002) construe and operationalize retribution as being mediated through moral outrage.

gests that third personal knowledge about *retributively* successful sanctions provides no clear retributive gratification benefits. Taken together, the studies show that “predicted and experienced emotion are substantially different” (1323).

The contrast between victims’ expectations and *ex post* assessments of offenders’ retributive suffering is documented by other studies as well. For instance, Gollwitzer & Denzler (2009), and Gollwitzer, Meder & Schmitt (2011) test two competing hypotheses about victims’ satisfaction in a retributive context. On the first, *comparative suffering hypothesis*, seeing the offender suffering from fate or from agent-inflicted suffering was predicted to be “sufficient for evoking satisfaction and perceptions of deservingness among victims (Gollwitzer et al. 2011: 364). On the second, *understanding hypothesis*, retributive suffering was predicted to be psychologically satisfactory only when offenders showed that they understood it as a response to transgressive behaviour. Overall, the first hypothesis was either weakly or not corroborated, while the second one was substantially corroborated. These studies suggest that the expectation that retributive suffering is *intrinsically* satisfying is empirically mistaken, and that suffering is rather *instrumentally* desirable – viz., victims “might want to make the offender feel bad because feeling bad indicates that the offender has learned his or her lesson” (Gollwitzer et al. 2011: 372).<sup>24</sup>

Aharoni et al. (2022) reach a similar conclusion with more realistic crime scenarios. Using a survey experimental set-up, they tested the effects of suffering and understanding on people’s satisfaction with punishment across different offenses (theft or aggravated robbery) and perspectives (participants were asked to imagine being the victim or to imagine that the victim was a member of their community). Similar to previous formulations, the *suffering hypothesis* holds that offenders’ penally induced suffering is intrinsically satisfying, whereas the *understanding hypothesis* holds that offenders’ penally induced understanding is intrinsically satisfying, with suffering serving, at best, as its instrument. Overall, the results show that “getting perpetrators to understand the wrongfulness of their actions and feel remorse might serve as a primary goal of punishment, whereas making the perpetrator suffer serves a secondary goal which becomes activated when the more primary goal of understanding has failed” (11). More specifically, punishment was deemed unsatisfying when both suffering and understanding were absent. But the study additionally found that suffering alone had a *negative* effect on punishment satisfaction when understanding was absent.

Closer to Moore’s retributive scenarios, Bauer and Poama (2020) assessed the comparative effects of the offender’s suffering and moral change on ordinary people’s justice perceptions. The study presented participants with a stylized version of Moore’s thought experiment based on *State v. Chaney* (1970), and experimentally manipulated whether the rape offender suffered or changed morally as a result

<sup>24</sup> These studies are corroborated by Funk, McGeer & Gollwitzer (2014) who conducted a series of in-lab experimental studies and found that participants who sanctioned offenders experienced justice-related satisfaction only when offenders displayed remorse for their transgression, but not when offenders only suffered for it. For a philosophical analysis of these findings, see McGeer & Funk (2017). Lambert et al. (2014) found that the contrast between expectations about and experiences of retributive suffering holds when we move from a second personal, victim-centred to a third personal, spectator perspective.

of being punished.<sup>25</sup> The results show that the effects of suffering are considerably smaller than the effects of moral change, with a noticeably small difference in justice perception between the version of the story where the offender displays moral change but does not incur retributive suffering and the version where the offender both suffers *and* displays moral change.

Taken together, these findings show that the impact of retributive suffering on people's satisfaction with punishment is small at best, and insignificant *or* negative at worst. Formulated in Dewey's normative efficiency language, the findings suggest that Moore's *DR* principle – viz., that is intrinsically just if guilty offenders suffer – is relatively inefficient. As discussed in the previous section, a principle might fail *ex post* if acting on it is absolutely inefficient (as when it elicits *some* negative experiences, not only positive ones) or relatively inefficient (as when the positive experiences do not outweigh the negative ones). On the evidence presented here, *DR* is absolutely inefficient. However, the absolute efficiency standard seems practically unworkable, as the application of otherwise justified principles can sometimes fail, and thus generate negative experiences. Not all generally justified principles succeed all the time.

However, we should expect that generally justified principles succeed in clearly bringing about the predicted positive experiences most of the time. If we endorse a naturalistic account of *ex post* justification whereby a principle is justified if acting on it generates positive experiences of the relevant kind, we ought to expect that acting on a generally justified principle elicits more positive than negative experiences *overall*. Generally justified principles satisfy this relative normative efficiency standard. *DR* seemingly does not – given the empirical evidence, it is unlikely that the small, negligible or insignificant positive experiences prompted by acting on retributive beliefs can, on balance, outweigh the negative experiences.

Cast formally, the concern that emerges from submitting *DR* to the practical test is that there are clear cases where  $[(Dsn \rightarrow Dgsn) \ \& \ \neg Dgsn] \rightarrow \neg Dsn$ . Since  $\neg Dgsn$  includes cases where *Dgsn* does not obtain (missing positive experiences), and cases where the opposite of *Dgsn* obtains (present negative experiences), and since, on the evidence presented here, such cases occur often enough, the evidential basis for inferring *DR* seems significantly diminished. This casts serious doubt on whether *DR* – viz., the principle that it is *intrinsically* just if guilty offenders suffer – can offer a coherent account for *most* of the relevant cases.

Importantly, the Deweyan critique takes seriously both Moore's moral naturalism and his coherentism, and thus averts begging the question in the way Dewey's own deployment of the critique did. The critique is naturalistic in virtue of the extended epistemic role it gives to particular experiences – viz., it holds that moral judgments are not only derived from *ex ante* experiences, but need to be additionally sensitive to, and potentially revised in light of the *ex post* experiences elicited by acting on the said judgments. This gives experience both a constructive and a critical role in the justification of our judgments and principles inferred from them. The critique is coherentist in two respects. First, it insists on the importance of the logical consis-

<sup>25</sup> Moral change in this study is operationalized as the convicted defendant apologizing for the offense and engaging in offense-relevant reparatory work from prison.

tency between the *ex ante* and *ex post* judgments that inform *DR*. Second, it is inferentially coherentist, in that it holds that *DR* requires sufficiently many cases where these two types of judgments are consistent.

Equally importantly, given its naturalist-cum-coherentist orientation, the critique raises a concern that is epistemically friendly to *DR*. To see this, contrast the Deweyan critique with another type of epistemic argument standardly directed at retributivism. For illustration purposes, take Caruso's (2020) recent view that we should reject retributivism because the so-called *basic desert claim*, which is the contention that people are morally responsible for their actions in a way that is not reducible to other causal factors, lacks the "evidentiary support needed to justify retributive legal punishment" (17). Caruso more specifically argues that the evidence retributivists need to secure should be both pervasive in its range – viz., it should be widely established – and justificatorily stringent – viz., it should satisfy an epistemic test whereby it is highly unlikely that it is mistaken or otherwise relevantly mistaken.<sup>26</sup>

To this, retributivists might reply that Caruso's epistemic critique is too demanding. First, if sound, it seems that the critique does not only require that we should discard retributivism, but also many other normative theories of justice which assume that people are free and responsible in a basic sense *and* have practically burdensome implications. This, retributivists might contend, suggests that Caruso's argument demands too much, and seems implausible. Second, given the stringency of its epistemic standard, the critique might require that we heavily discount ordinary judgments-based evidence that people are morally responsible in a basic sense.<sup>27</sup> This, retributivists can further contend, demands too much, and seems false.

Compared to Caruso's critique, the Deweyan one raises no special overdemandingness concerns. This is, first, because the critique can straightforwardly endorse the claim that there is enough evidence that people are free in a basic desert sense, and so leaves many of our theories of justice and ordinary judgments about freedom and responsibility unscathed. Second, and more importantly, the central tenet of critique – viz., that retributive judgments should be justified experientially – is already built into *DR*. Consequently, the critique is demanding only at the relevant, and epistemically charitable level.

## 5 The Deweyan Critique: Four Objections

Here, I briefly consider what I take to be some of the most important objections against my critique of *DR*. The first objection is that the available empirical findings do not seriously undermine *DR*'s evidential basis. Call this the *insufficient evidence* objection. The contention behind this objection is that, since the studies which find that *ex ante* retributive judgments are undermined *ex post* are scarce, we do not have

<sup>26</sup> As indicated by Caruso, the argument belongs to a wider family of philosophical views, which includes Pereboom (2006), Vilhauer (2015), and Corrado (2017). For an argument that focuses on justificatory stringency for *specific* retributive beliefs (i.e., that retributive beliefs that it is just for wrongdoers to suffer need to satisfy a beyond a reasonable doubt test), not general desert beliefs, see Hanna (2014).

<sup>27</sup> For an analysis of this evidence, see Cova & Kitano (2014).



strong reasons not to support *DR*. This objection is right to emphasize the scarcity of the relevant empirical findings, but it falls short in at least two respects. First, it misrepresents the point of the Deweyan critique, which is not that we should reject *DR*, but that, *for all we know*, we lack an adequate evidential basis for accepting it. Second, the objection somewhat disingenuously fails to acknowledge that the *ex post* case against *DR* is better substantiated than the *ex post* case in its favour.

The second objection is that many of the empirical findings discussed above cannot *validly* target *DR*. Call this the *inadequate evidence* objection. The claim is that many of the studies examined in the previous section examine the experiences people get from exacting *revenge*, not from obtaining *retributive justice*. This objection is correct to note that some of the studies at issue are formulated in the language of revenge. But the objection seems overstated. This is because not all of the studies cited above look at the experiential impact of revenge, but also because the studies that focus on revenge define it in terms which are equivalent with or substantively similar to how retributive justice is construed under *DR*.<sup>28</sup>

The third objection is that, given its pragmatic orientation, the practical test only captures the pragmatic *consequences* of holding retributive judgments for those who hold such judgments, not the evidential weight of retributive judgments as such. Call this the *anti-pragmatist* objection. The objection is correct to insist that we can *generally* draw a distinction between evidential and pragmatic reasons for holding a judgment.<sup>29</sup> But, irrespective of one's epistemological stance on this distinction, we can doubt whether the objection can be advanced from a naturalistic standpoint. As noted, Moore's defense of *DR* is *epistemically* naturalistic in that it attaches evidential weight to the experiences we undergo when particular offenses are committed – viz., it holds that the emotional experiences of first-personal guilt and third-personal indignation count as evidence about the value which warrants punishing these offenses. Given this, *DR* explicitly holds that evidence is strongly experience-sensitive, and so the epistemic exclusion of *ex post* experiences elicited by retributive judgments seems *ad-hoc*.

The fourth objection is that the Deweyan framework deployed here is not necessary for advancing my substantive critique. Call this the *redundancy* objection. The objection contends that the terminological idiosyncrasies of Dewey's own argument against retribution – for instance, his claims about retribution being inefficient in its application – distract from some of the central insights of the critique. This objection is correct to note that the critique does not depend on Dewey's admittedly *sui generis* normative vocabulary.<sup>30</sup> But the objection neglects at least two the advantages that the Deweyan framework enables. First, and more generally, my attempt to explicitly derive the critique from Dewey's inchoate case against retribution is, in part, meant

<sup>28</sup> For instance, Gollwitzer & Denzler (2009), takes revenge to involve retribution in the deserving suffering sense (840).

<sup>29</sup> For an analysis of this distinction, see Vahid (2010).

<sup>30</sup> For instance, the critique could be formulated in the language of defeaters – viz., by arguing that the *ex post* experiences elicited by acting on retributive beliefs count as rebutting reasons against *DR*. This reformulation would allow for a smoother comparison with epistemic critiques that *DR* is defeated because evidentially undercut. For an example of such a critique, see Wiegman (2017). For a discussion of the distinction between rebutting and undercutting defeaters, see Pollock (1987).

to illustrate how occasional dives into the history of normative thought can benefit systematic normative analysis.

Second, and more specifically, the Deweyan account – most notably, the emphasis it sets on *ex post* justification and normative efficiency – offers a terminologically clearer and more straightforward framework for assessing whether and, if so, how retributivism might be *practically* involved in excessive or otherwise deficient forms of punishment.<sup>31</sup> The central insight that Dewey offers against retribution – viz., that the value of retributive justice is *ex post* undermined by *retributively* elicited experiences – thus allows us to specify, and assess how retributive theories might, as a matter of actual practice, drive unjustified punishment.

VI. In this article, I argued that theories such as *DR* which contend on naturalistic-cum-coherentist grounds that punishment is justified because it exacts the suffering guilty offenders deserve, lack an epistemically relevant evidential basis. More specifically, I argued that they do not offer evidence that *ex ante* retributive judgments are corroborated by *ex post* experiences elicited by acting on these judgments and further argued that there is evidence that the latter either do not significantly corroborate or contradict the latter. My critique, then, is that the case for *DR* is evidentially incomplete at best, and implausible at worst. If this critique is cogent, the value of retributive justice needs to be either defended on different grounds or revised to accommodate the available evidence – viz., in terms of an account whereby deserved suffering is a necessary but not sufficient consideration for justified punishment.

## Declarations

**Conflict of Interest** There is no conflict of interest that I have with the Journal of Ethics editors or other members of the editorial board. Nor do I have any other conflict of interest that I am aware of.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aharoni, E., D. Simpson, E. Nahmias, and M. Gollwitzer. 2022. A painful message: testing the effects of suffering and understanding on punishment judgments. *Zeitschrift für Psychologie*.
- Anderson, E. 2010. The imperative of integration. In *The imperative of integration*. Princeton University Press.
- Anderson, E. 2014. Social movements, experiments in living, and moral progress: case studies from Britain's abolition of slavery, available online at: <https://kuscholarworks.ku.edu/handle/1808/14787>.

<sup>31</sup> For a recent discussion (and critique) of claims that retributivism is practically responsible for over-punishment, see Husak (2022).

- Anderson, E. 2015. Moral bias and corrective practices: A pragmatist perspective. In *Proceedings and Addresses of the APA* 89: 21–47.
- Anderson, E. 2023. *Dewey's Moral Philosophy*, Stanford Encyclopedia of Philosophy
- Bauer, P. C., and Poama, A. 2020. Does suffering suffice? An experimental assessment of desert retributivism. *Plos One*, 15 (4), e0230304.
- Berman, M. N. 2008. Punishment and justification. *Ethics* 118(2): 258–290.
- Boghossian, P. 2021. Normative principles are synthetic a Priori. *Episteme* 18(3): 367–383.
- Brooks, T. 2010. Punishment and british idealism. In *Punishment and Ethics*, ed. J. Ryberg, 16–32. London: Palgrave Macmillan.
- Carlsmith, K. M., J. M. Darley, and P. H. Robinson. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of personality and social psychology* 83(2): 284.
- Carlsmith, K. M. 2006. The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology* 42(4): 437–451.
- Carlsmith, K. M., and J. M. Darley. 2008. Psychological aspects of retributive justice. *Advances in experimental social psychology* 40: 193–236.
- Carlsmith, K. M., T. D. Wilson, and D. T. Gilbert. 2008. The paradoxical consequences of revenge. *Journal of personality and social psychology* 95(6): 1316.
- Caruso, G. D. 2020. Justice without retribution: an epistemic argument against retributive criminal punishment. *Neuroethics* 13(1): 13–28.
- Corrado, M. L. 2017. Punishment and the Burden of Proof. *UNC legal studies research paper*.
- Cottingham, J. 1979. Varieties of retribution. *The Philosophical Quarterly* 29(116): 238–246.
- Cova, F., and Y. Kitano. 2014. Experimental philosophy and the compatibility of free will and determinism: a survey. *Annals of the Japan Association for Philosophy of Science* 22: 17–37.
- Darley, J. M., K. M. Carlsmith, and P. H. Robinson. 2000. Incapacitation and just deserts as motives for punishment. *Law and human behavior* 24(6): 659–683.
- Davis, L. H. 1972. They deserve to suffer. *Analysis* 32(4): 136–140.
- Dewey, J., and J. H. Tufts. 1909. *Ethics*. New York: Henry Holt and Co.
- Dewey, J. 1916. Force and coercion. *The International Journal of Ethics* 26(3): 359–367.
- Dewey, J. 1932/2022. *Ethics (John Dewey's Ethical Theory: The 1932 Ethics)*, Frega, R., Levine, S. (eds), Routledge.
- Funk, F., V. McGeer, and M. Gollwitzer. 2014. Get the message: punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin* 40(8): 986–997.
- Gollwitzer, M., and M. Denzler. 2009. What makes revenge sweet: seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology* 45(4): 840–844.
- Gollwitzer, M., M. Meder, and M. Schmitt. 2011. What gives victims satisfaction when they seek revenge? *European journal of social psychology* 41(3): 364–374.
- Hanna, N. 2014. Retributivism revisited. *Philosophical Studies* 167(2): 473–484.
- Hanna, N. 2019. Hitting retributivism where it hurts. *Criminal Law and Philosophy* 13(1): 109–127.
- Husak, D. 2022. Retributivism and Over-Punishment. *Law and Philosophy* 41(2): 169–191.
- Kleinig, J. 1973. *Punishment and Desert*. Dordrecht: Springer.
- Lambert, A. J., S. A. Peak, F. R. Eadeh, and J. P. Schott. 2014. How do you feel now? On the perceptual distortion of extremely recent changes in anger. *Journal of Experimental Social Psychology* 52: 82–95.
- Moore, M. S. 1997/2010. *Placing blame: a theory of the criminal law*. Oxford University Press.
- McGeer, V., and F. Funk. 2017. Are 'optimistic' theories of criminal justice psychologically feasible? The probative case of civic republicanism. *Criminal Law and Philosophy* 11(3): 523–544.
- Neta, R. 2017. Two legacies of Goldman's epistemology. *Philosophical Topics* 45(1): 121–136.
- Nussbaum, M. C. 2016. *Anger and forgiveness: resentment, generosity, justice*. Oxford University Press.
- Pappas, G. F. 1997. Dewey's moral theory: experience as method. *Transactions of the Charles S Peirce Society* 33(3): 520–556.
- Pereboom, D. 2006. *Living without free will*. Cambridge University Press.
- Pollock, J. L. 1987. Defeasible reasoning. *Cognitive science* 11(4): 481–518.
- Rosebury, B. 2011. Moore's moral facts and the gap in the retributive theory. *Criminal Law and Philosophy* 5(3): 361–376.
- State v. Chaney 1970 477 P.2d 441
- Ten, C. L. 1987. *Crime, guilt, & punishment*. Oxford: Oxford University Press.
- Vahid, H. 2010. Rationalizing beliefs: evidential vs. pragmatic reasons. *Synthese* 176(3): 447–462.

- Vilhauer, B. 2015. Free will and the asymmetrical justifiability of holding morally responsible. *The Philosophical Quarterly* 65(261): 772–789.
- Walker, N. 1999. Even more varieties of retribution. *Philosophy* 74(4): 595–605.
- Welchman, J. 1997. *Dewey's ethical thought*. Cornell University Press.
- Wiegman, I. 2017. The evolution of retribution: intuitions undermined. *Pacific Philosophical Quarterly* 98(2): 193–218.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.