Retributivism, Justification and Credence: The Epistemic Argument Revisited

Sofia M. I. Jeppsson

Received: 9 December 2019 / Accepted: 12 May 2020 / Published online: 29 May 2020
© The Author(s) 2020

Abstract Harming other people is *prima facie* wrong. Unless we can be very certain that doing so is justified under the circumstances, we ought not to do it. In this paper, I argue that we ought to dismantle harsh retributivist criminal justice systems for this reason; we cannot be sufficiently certain that the harm is justified. Gregg Caruso, Ben Vilhauer and others have previously argued for the same conclusion; however, my own version sidesteps certain controversial premises of theirs. Harsh retributivist criminal justice can only be morally right if the following three propositions are true: Moral responsibility exists, retributivism is right, and we can find out how much punishment offenders deserve for their crimes. Suppose that we initially assign a high credence to each of the three propositions; I assume for the sake of argument that there are good arguments in support of each. Nevertheless, these arguments ultimately depend on intuitions. Since we have philosophical peers whose intuitions differ from ours, we ought to downgrade our credence in each. However, even slightly less credence in each proposition means far less credence in a conjunction of all three. Since the stakes are high and there are morally safer options for a criminal justice system, we ought to dismantle harsh retributivist ones.

Keywords Moral responsibility · Desert · Retributivism · Criminal justice · Peer disagreement · Uncertainty

The Epistemic Argument Against Harsh Retributivism

Retributivism is the view that the criminal justice system ought to give offenders the punishment that they deserve. But *should* we have a desert-based system? The answer might depend on how we fill in the details. Philosophers who self-apply the label ‘retributivist’ can have very different views; an anti-retributivist and a retributivist might even have more in common with each other, at least when it comes to practical suggestions, than one retributivist with another. For example, if retributivist Antony Duff and anti-retributivist Derk Pereboom were put in charge of creating a country’s criminal justice system, they might get along just fine, as long as they focused on institutional design and practice rather than the reasons behind them.

In this paper, I will focus specifically on *harsh retributivist criminal justice systems*, or Harsh Retributivism for short. A state has Harsh Retributivism if the primary justification for punishing offenders is the dealing out of just deserts. Deterrence and other considerations might still play some role, but desert is considered most important. Furthermore, crimes (at least the more serious ones) are punished harshly. Offenders receive long prison sentences in an environment designed to be punitive rather than rehabilitative; some

S. M. I. Jeppsson (✉)
Department of Historical, Philosophical and Religious Studies,
Umeå University, 901 87 Umeå, Sweden
e-mail: sofia.jeppsson@umu.se

crimes might even be punishable by death. Should we have Harsh Retributivism?

Ben Vilhauer, Gregg Caruso, Michael Louis Corrado and Elizabeth Shaw have all advanced versions of an *epistemic argument* against Harsh Retributivism [1–6]. I will primarily focus on Vilhauer and Caruso; they have full arguments that I believe are on the right track. Nevertheless, Caruso slides from an epistemic argument into standard skepticism, whereas Vilhauer relies on a controversial premise. The primary purpose of this paper is to present another version of an epistemic argument that sidesteps both these issues, even though my version unavoidably relies on other premises that are not wholly uncontroversial.

Caruso's Epistemic Argument

Caruso's argument, in a slightly abridged version, runs as follows:

1. Legal punishment inflicts harm on people; we ought not to harm others if it is significantly probable that we are unjustified in doing so.
2. The retributive justification for legal punishment assumes that agents have desert-entailing moral responsibility.
3. If this assumption is not justified according to high epistemic standards, Harsh Retributivism is *prima facie* seriously wrong.
4. Both compatibilism and libertarianism about desert-entailing moral responsibility face powerful and unresolved objections.
5. Therefore, Harsh Retributivism is unjustified and the harms it causes are *prima facie* seriously wrong.

There is much to discuss in this argument, and I will return to its premises later in this paper. For now, I will focus on the problems with premise four.

When Caruso elaborates on it, he first shows (by way of quotes) that certain prominent scholars in the American criminal justice debate have more doubts about the existence of moral responsibility than their self-applied compatibilist label would suggest. This is interesting, but does not show that Harsh Retributivism is unjustified. Caruso notes this, and moves on to argue that even scholars who are convinced that moral responsibility exists are so for irrational reasons, since the arguments for compatibilism and libertarianism are weak. At this

point, we no longer have a special *epistemic* argument against Harsh Retributivism. Readers who agree with Caruso about the weakness of compatibilist and libertarian arguments will of course be skeptics like him, and therefore oppose retributivism. However, the persistent liveliness of the moral responsibility debate is ample proof that you cannot turn compatibilists and libertarians into skeptics merely by insisting that their arguments are weak. A true epistemic argument should provide doubts-based reasons to oppose Harsh Retributivism to people who *do not* embrace skepticism. (Full disclosure: I am not a moral responsibility skeptic myself.) Ben Vilhauer attempts to do just that.

Vilhauer's Epistemic Argument

Here is Vilhauer's short and concise version of the epistemic argument [4]: 131:

1. If it can be reasonably doubted that someone had free will with respect to some action, then it is a requirement of justice to refrain from doing serious retributive harm to him in response to that action.
2. Anyone who believes the free will debate to be philosophically valuable must accept that it can be reasonably doubted that anyone ever has free will.¹
3. Therefore, anyone who believes the free will debate to be philosophically valuable must accept that it is a requirement of justice to refrain from serious retributive harm.

Both Caruso and Vilhauer refer, in their respective papers, to the epistemic standard of certainty beyond a reasonable doubt. Nevertheless, many jurisdictions put the burden of proof on the defense if they claim that the defendant was not in control of his actions. This is a *prima facie* problem for philosophers who want to argue that moral responsibility skeptics need not prove their thesis, only raise doubts, to undermine Harsh Retributivism. Corrado responds that it makes sense to demand that the defense presents proof of duress or legal insanity, because they are both rare, and usually relatively easy to prove if they do occur. Since neither

¹ Vilhauer talks about the free will debate rather than the moral responsibility debate, on the assumption that free will is necessary for desert-entailing moral responsibility. I will continue to write directly about moral responsibility, but this makes no difference for my arguments.

consideration supports placing the burden of proof on moral responsibility skeptics, the burden should remain with those who try to justify intentionally harming others [2]: 3, 9–10.

Vilhauer, unlike Caruso, does not insist that compatibilist and libertarian arguments are weak. He asks only that we see the moral responsibility debate as ‘valuable’ or ‘truth conducive’. A libertarian or compatibilist can believe they have the right view, but still find skeptical counter-arguments valuable; by responding to those, compatibilist and libertarian theories might improve over time. Compatibilists and libertarians who think this way thereby accept that skeptical doubts are reasonable – *unreasonable* objections to their theories would be of no help. Therefore, they are not certain *beyond a reasonable doubt* that moral responsibility exists.

Vilhauer’s argument thus appeals to a bigger audience than Caruso’s: not just to skeptics, but to compatibilists and libertarians with an optimistic view of the debate as well. However, a more pessimistically inclined philosopher might doubt that their theories really improve over time, as compatibilists, libertarians and skeptics take each other’s counter arguments into account. Corrado writes:

In a philosophical debate that has gone on for millennia without any of the various sides making much progress in convincing the others, it is a little surprising to find one side planting its flag and declaring victory on the basis of the burden of proof... In the free will debates none of the sides, skeptical or realist, libertarian or compatibilist, has made much progress in persuading the others. In spite of complex and ingenious devices like the Consequence Argument and the Conditional Analysis, the debates seem to end in a dead heat. [2]: 4

Now, it is possible that our moral responsibility theories are getting better over time, even though we do not get any closer to a consensus. However, one might be pessimistic about that as well. For instance, many arguments in the debate build on thought experiments. The original versions are simple, give rise to strong intuitions in the audience, but are vulnerable to objections. Over time, increasingly sophisticated and complicated versions are invented to escape said objections, but in the

process, they also lose much of their intuitive appeal. Frankfurt examples, to which I will return later in the paper, provide a good example of this kind of thought experiment evolution. Have Frankfurt examples *improved* over time? It is not obvious that the answer is ‘yes’.

My arguments do depend on there being philosophical value *of sorts* in these debates, a point I will return to later. Still, I do not assume that there is progress towards increasingly good arguments and theories, or that these theories gradually get closer to the truth about free will and moral responsibility.

The primary purpose of this paper is to develop an epistemic argument against Harsh Retributivism that rests neither on the assumption that compatibilist and libertarian arguments are weak, nor on the premise that the moral responsibility debate makes progress towards better and better arguments and theories. In doing so, I will inevitably wade into a few *other* philosophical controversies. However, I believe it is worthwhile to show that you can reach the same conclusion via different paths. Although none of them proceeds from uncontroversial premises, the arguments given by Caruso, Vilhauer and me rely on *different* ones; there is no single highly controversial premise that everyone must accept in order to conclude that we ought to abandon Harsh Retributivism.

Furthermore, I show that at least arguably (it depends on whether *moral* uncertainty plays the same role in decision-making as other kinds), a number of uncertainties will multiply, and make the epistemic argument even stronger.

Harsh Retributivism, Credence and Justification

Caruso and Vilhauer focus on what we subjectively ought to do, and so will I. As already stated, I will ignore the distinction between *moral* uncertainty and other kinds for the time being. I will thus ignore, for now, the difference between what Brian Hedden [7] calls *subjective ought* and *super-subjective ought*. Whereas the former is merely sensitive to descriptive uncertainty, the latter is sensitive to moral uncertainty as well. If I am a doctor with a seriously ill patient, I might be uncertain of which shot to give her for two different reasons. In the first scenario, drug A will cure her, B kill her, and C improve her condition, but not cure her completely. Although I know that C will improve her, I do not know

which of A or B is the cure drug and which is lethal. Objectively, I ought to give her drug A, but subjectively, given my lack of *descriptive* knowledge, I ought to give her C. In the second scenario, I know what all the drugs do, and I can either prolong her life somewhat or kill her. The patient asks for euthanasia to end her suffering, and as a matter of objective moral fact (let us imagine, purely for the sake of argument) I ought to comply. Nevertheless, I *believe* that euthanasia is always morally wrong; I *super-subjectively* ought to abstain.

I will come back to the issue of moral uncertainty later in the paper, but I will ignore this distinction for now, lump all uncertainties together, and simply discuss what people subjectively ought to do.

Still, we are not interested in what people ought to do given *whatever* irrational beliefs and ideas they might have. Caruso agrees that given certain beliefs held by certain scholars, Harsh Retributivism can be justified, but he argues that they hold these beliefs on irrational, or at least insufficient, grounds and ought to change them. Certainly, legislators and policy makers have a general, standing duty to do their research and deliberate in as impartial a manner as possible before making important decisions.

Is the ‘ought’ we are interested in, then, what an *ideal agent* would conclude is her duty, if she found herself in circumstances in which she only had access to whatever information actual agents have at their disposal? No, this will not work for our current purposes. Later in the paper, I will delve into the thorny issue of intuitions. A perfectly rational agent would presumably have perfect and consistently truth-tracking intuitions, and furthermore know this. For this reason she would rarely be uncertain about philosophical issues, even if she lacks empirical information. Nevertheless, we are interested in what to do precisely under such uncertainty.

I will therefore discuss ‘subjective ought’ in the sense of what a *semi-idealized* agent would do: an agent who lacks omniscience and suffers the same informational restrictions as real agents, including less-than-perfect intuitions. While not perfect, she is idealized in the sense that she does not suffer from the biases, prejudices, wishful thinking and general irrationality that real agents do. Furthermore, she knows herself fairly well, and justifiably believes that she does not suffer from any serious bias or the like. She is idealized enough that it no longer makes sense to say “Well, her irrational beliefs do imply that X is right, but this does not mean that she subjectively ought to do X; it simply means she ought to stop being so irrational”. There

is an interesting sense in which this agent really ought to do X in case her beliefs (desires, values and so on) imply it.

With this ‘subjective ought’ analysis in place, let us picture a group of people in position to influence the criminal justice system, in a nation which currently has Harsh Retributivism. The details of their position can be worked out in different ways. They might be politicians, members of a law institute, public intellectuals with a lot of clout, etc. There is some controversy about Harsh Retributivism in the public debate; moral responsibility skeptics and others attack it, but it has many defenders as well. Our influencers therefore ponder whether they ought to support Harsh Retributivism or work towards radical reform.^{2, 3}

The influencers’ initial stance, before they begin to seriously deliberate about the matter, is one of support for Harsh Retributivism. They realize that Harsh Retributivism cannot be morally right unless offenders can deserve to be punished for their crimes; i.e., unless desert-entailing moral responsibility exists. The influencers, however, are convinced compatibilists.

Philosophers debate how best to understand the term ‘moral responsibility’; some argue that we can meaningfully distinguish between several different kinds. For instance, Gary Watson popularized the term ‘attributability’, a kind of responsibility concerned with evaluating people on a moral scale; said evaluation need not have any implications for deserved punishment [14]. For the purposes of this paper, I will focus wholly on the desert-entailing kind relevant to debates about retributivism. When people are morally responsible in this sense, it can be morally right to blame or even punish them for wrongful behavior; not because doing so will have good consequences, but merely because they did what they did and deserve some hard treatment for it. The influencers are compatibilists about *this* kind of moral responsibility, not some watered-down concept.

² It is important to note that the influencers’ decision is about *the criminal justice system*. For the purposes of this paper, I completely ignore everyday, interpersonal blame. For all I say, blaming each other for wrongful behavior might be an inescapable part of ordinary, human relationships [8, 9]. Within the confines of this paper, I take no stand on this issue at all. However, even if this were the case, it would be absurd to claim that *Harsh Retributivism* is an inescapable part of society, since criminal justice systems differ between nations and over time.

³ The purpose of this paper is not to develop and defend an alternative to Harsh Retributivism; many other philosophers have done so already [10–13]. It is nevertheless important to note that we do not have to choose between on the one hand Harsh Retributivism, and on the other hand, a crude utilitarian system where we do whatever it takes to deter criminals.

Caruso (and Waller [15], to whom he refers) argues that people who believe in moral responsibility do so irrationally, despite the weakness of the arguments in favor of their position. Vilhauer, on the other hand, does not make these assumptions, and neither will I. Instead, I *postulate* that the influencers did not become compatibilists through irrational, motivated reasoning. They have read the relevant philosophical literature, thought conscientiously about the matter, and carefully weighed the arguments. Furthermore, they know themselves well enough to justifiably believe that they have not engaged in biased, motivated reasoning.⁴ Eventually, they became compatibilists. Let us say they were convinced that John M. Fischer's brand of moral responsibility compatibilism is, if not 100% true, at least strongly on the right track. Fischer argues that in order to be morally responsible for what she does, an agent must be responsive to and act for reasons [16, 17] (see also Fischer and Ravizza [18]).⁵ She need not be perfectly rational (if that were the case, no one would be responsible for anything, since no one has perfect rationality), but she must be rational enough. The influencers find it eminently plausible that something along the lines of this non-mysterious psychological capacity grounds moral responsibility. Caruso writes that compatibilists cannot satisfyingly respond to famous incompatibilist arguments, like Pereboom's four-case manipulation argument [19, 20], van Inwagen's consequence argument [21], Galen Strawson's basic argument [22, 23], and Levy's luck argument [24]. The influencers disagree. Within the confines of this paper, I can obviously not repeat the relevant debates: two brief examples will have to do. Pereboom's four-case manipulation argument begins with an agent who is blatantly manipulated into doing what he does, and therefore, Pereboom writes, obviously not morally responsible. Through a series of morally irrelevant changes to the scenario, Pereboom eventually arrives at an ordinary agent in a deterministic universe. Since all the changes were morally irrelevant, Pereboom writes, ordinary agents under determinism lack moral responsibility as well [19]: chapter 4 [20]: chapter 4. He later adds that the argument goes through

even if we tweak the scenario to a plausible version of indeterminism, which shows that we lack moral responsibility regardless of whether the universe is deterministic or indeterministic [20]: 83. Michael McKenna objects that we can just as well run the argument in the opposite direction [25]. Start with the ordinary agent, who is surely responsible for what he does, go through the morally irrelevant changes, and conclude that the manipulated agent is actually responsible as well. Let us imagine that the influencers have read the important papers in this debate, thought long and hard about it, and come to side with McKenna. Moving on to the consequence argument, it purports to show that under determinism, no one can do otherwise than they do [21]: 69. Therefore, the moral responsibility skeptic claims, moral responsibility is at the very least incompatible with determinism. However, Harry Frankfurt famously argued that an agent could be morally responsible for what he does even if he lacks the ability to do otherwise [26]. A classic Frankfurt case looks like this: Black and Jones both want Smith dead. Black prefers for someone else to do the dirty work, so unbeknownst to Jones, he implants a device in the latter's head that allows him to monitor Jones' thoughts and actions and take control of them if need be. As it happens, Jones murders Smith on his own; Jones is presumably responsible for doing so, even though he could not have done otherwise. The influencers have read the relevant literature in this area as well, thought long and hard, and come to side with Frankfurt and his followers. And so on. Every skeptical argument in the literature has its counter arguments. Whether these are *satisfying* or not is, of course, a controversial topic (if it were not, we would no longer have a moral responsibility debate, but agreement among the philosophers in this field). But the influencers side with the compatibilists, and not because they suffer from irrationality or distorting biases.

Let us therefore accept that the influencers ought not to abandon compatibilism and instead embrace skepticism, in the here used subjective sense of 'ought'. However, the influencers need more than compatibilism to justify a continued support of Harsh Retributivism.

Three Necessary Conditions

Vilhauer and Caruso both focus on whether free will, moral responsibility and desert exist. True enough, Harsh Retributivism rests on the assumption that

⁴ If the influencers had reached a compatibilist conclusion through rational arguments but did not know this, and rather suspected that they had fallen prey to unconscious bias or some other type of irrationality, this would merely strengthen my arguments, since I later argue that they ought to downgrade their initial high credence.

⁵ Or, rather, the agent must act from a reasons-responsive *mechanism*, but these details need not concern us here.

offenders can deserve to be punished; if there is no moral responsibility, it is wrong to have Harsh Retributivism. Nevertheless, the existence of moral responsibility is not *sufficient* to justify Harsh Retributivism. There are at least two additional necessary conditions: retributivism must be true, and we must be able to find the right punishment for people's crimes.

According to retributivism, the criminal justice system ought to give offenders the punishment that they deserve. However, the thesis that wrongdoers can deserve blame and possibly punishment does not, in itself, have any implications for the criminal justice system. Maybe people can only deserve social, informal punishments (like cold shoulders, not being invited to events, etc.), but no more than that. Maybe other considerations ought to have much more weight than desert in criminal justice. Robert Kane is a libertarian, but argues that everyone has a strong pro tanto right to liberty. Therefore, interfering with dangerous criminals in order to protect others is a necessary evil rather than a good [27]: 19–30. Out of all the reasons moral responsibility skeptic Derk Pereboom enumerates against Harsh Retributivism, most are actually independent of skepticism, and could be embraced by compatibilists and libertarians as well [20]: 158–160. One could even object to Harsh Retributivism on the grounds that criminals often grew up under harsh circumstances, and arguably ended up in the criminal justice system through events largely beyond their control, without embracing the view that moral responsibility is *metaphysically* impossible and does not exist at all.⁶

Harsh Retributivism needs both the existence of desert-entailing moral responsibility and the truth of retributivism for its moral justification – and, finally, that people can find out how much punishment offenders deserve for their crimes.

Few philosophers would argue that there is an exact science of how much punishment each offender deserves.⁷ It would be strange to suggest that there is, for

instance, a precise number of days that offender O deserves to spend in prison P for having committed crime C. If O is moved from P to prison P*, which has a slightly more pleasant environment, there is no mathematical formula for adjusting the number of deserved days upwards. Even if an objectively true answer to the question of what O deserves exists, although we cannot find it, this need not be much of a problem for retributivism as long as we know how to stay in the ballpark. Nevertheless, if criminals deserve to be punished and the criminal justice system ought to deal out just deserts, but our estimates of deserved punishments are *wildly* off since criminals deserve *much* milder punishments than are dealt out within Harsh Retributivism, having this kind of system is still morally wrong.

Let us abbreviate the thesis that moral responsibility exists MRE, the thesis that retributivism is true simply R, and the thesis that we can find out how much punishment offenders deserve FPD. The influencers ought not to support Harsh Retributivism unless they ought to assign a high credence to the truth of $((MRE \wedge R) \wedge FPD)$. When I discuss the influencers assigning credence to different propositions, I use 'ought'/'should' in the same subjective sense as when discussing whether they ought to support Harsh Retributivism or work towards changing it. The credence they ought to assign to a proposition is the credence a semi-idealized agent in the same position, with access to the same evidence, but undisturbed by bias, prejudice, wishful thinking and the like, would assign. I will argue that the influencers should *not* assign a high credence to $((MRE \wedge R) \wedge FPD)$, first because they ought to downgrade any initial high credence they have in each of the constituents, and second because rational credence in the whole conjunction is far lower than the credence in any single constituent.

However, it might seem as if FPD implies both R and MRE, so all we actually have to worry about is how much credence to give the former. But this is too quick. Suppose we conclude from FPD and people's best estimates of deserved punishments that some offenders can deserve long, harsh prison sentences. Nevertheless, FPD does not imply retributivism. As mentioned above, other considerations (cost efficiency, crime reduction, etc.) might outweigh the importance of people getting what they deserve; even if some offenders do deserve prison, it might still be better, *all things considered*, to have them do, e.g., community service. Furthermore, we

⁶ Arguably, retributivism does not imply that moral responsibility exists either. Perhaps we ought to define retributivism conditionally: as the thesis that *if* there are morally responsible criminals, *then* the criminal justice system ought to give them the punishment that they deserve. If so, retributivism and moral responsibility skepticism could both be true. This view might still be an odd one, though; unlike the combination of libertarianism/compatibilism and non-retributivism, I do not know of anyone who has actually defended it.

⁷ On the view that offenders *do* deserve very precise amounts of punishment, and diverting from these precise amounts is unjust, it would be even harder to justify Harsh Retributivism.

cannot defend FPD against direct counter arguments (like those found in Braithwaite and Pettit [13]: 149–151), and thereby get MRE as a free lunch. The thesis that moral responsibility exists must be separately defended against skeptical arguments attacking it.

A high rational credence in $((\text{MRE} \wedge \text{R}) \wedge \text{FPD})$ thus requires, just for starters, good arguments for compatibilism or libertarianism, good arguments for retributivism, *and* good arguments for the thesis that we can find out how much punishment offenders deserve.

So how much credence should the influencers have in $((\text{MRE} \wedge \text{R}) \wedge \text{FPD})$? I have already stated that they are compatibilists, and ought not to abandon this position for skepticism, since they became compatibilists through careful reading of the literature and weighing of the arguments. Now, let us further postulate that they became convinced that retributivism is true through a similar, unbiased process; we might imagine, for instance, that Michael Moore's arguments convinced them [28]. Our retributivist intuitions do not rest, Moore writes, on pathological and vengeful emotions we would be better off without, but rather on a virtuous commitment to morality, and we should trust them. The virtuous person who have committed a serious wrong has strong feelings of guilt and believes she deserves to be punished; if she respects other people, she extends the same judgment to them as she applies to herself [28]: 104–188. The influencers think carefully about the matter, and come to believe that Moore is right in all of this. Finally, they trust that we have truth-tracking intuitions about which punishment fits which crime. Moore argues that we can reliably find out what someone deserves by asking what punishment we would deserve had we done the same thing, and Paul Robinson points out the cross-cultural stability of judgments of relative seriousness for different crimes [29]: chapter 1.

Given all I have said in this section, it might seem as if the influencers should *not* try to dismantle Harsh Retributivism and replace it with a softer alternative. The influencers believe, for good reasons, in compatibilism, retributivism and our ability to find out which punishment offenders deserve for which crime; they are not rationally obligated to abandon these beliefs. However, if they start out highly certain, they still ought to downgrade their credence a bit. If they believe all these philosophical debates to be truth conducive in the sense that they produce better and better theories as the debates go on, they should downgrade their credence

for the reason mentioned by Vilhauer: If there is continuous improvement, this is likely because all sides are reasonable and argue well. Their opponents present *reasonable* doubts about compatibilism, retributivism and the possibility of finding the right amount of punishment. If, on the other hand, they consider the debates utterly stuck, this does *not* imply that their opponents are *unreasonable*. Rather, the best explanation points to the role of intuitions in the arguments employed. When relying heavily on intuitions, peer disagreement ought to be taken seriously.

Peer Disagreement

The claim that the influencers should take peer disagreement into account is immediately controversial. After all, there is a lively debate about which role, if any, peer disagreement should play in epistemology. Nevertheless, few philosophers would defend the position that discovering disagreeing peers is *never* relevant for the justification of one's beliefs [30]: 4. Here is the standard textbook scenario invoked to prove this point: I have dinner with a group of friends, and we decide to split the bill evenly. I do the math in my head and conclude that we ought to pay 43 dollars each, but all of my (mathematically competent) friends arrive at the number 47. Even if I am normally good at mental calculations, I ought to downgrade my credence in the proposition that we ought to pay 43 dollars each, and conclude that I might have made a mistake. It would be irrational to insist that I *know* it is 43 dollars since I did the math, and so it *must* be the case that everyone else made the exact same mistake whereas I got it right.

Some philosophers still argue that if we downgrade our credence in a proposition every time we discover peer disagreement, we will slide straight into a wholesale skepticism about everything. The usual reply is that serious disagreement between suitably educated and informed experts is actually not *that* common, and no one argues we must take disagreement seriously when it comes from people who lack the relevant competence. Often, there will be a lot of initial disagreement, with more consensus developing as time goes by and people learn more and more about the issue at hand. Even within philosophy, there might be less persistent disagreement than what seems to be the case at first glance

(see Kornblith on logic and decision theory [31]: 40–41). Katia Vavova argues that there is less peer disagreement than one might think even in ethics, at least if we focus on the more applied level, where moral philosophers discuss *what to do* [32]. For instance, animal ethicists converge over time on how to treat non-human animals, even though a utilitarian like Peter Singer and a Kantian like Christine Korsgaard reach their conclusions through different routes. Furthermore, we can dismiss *some* ethical disagreement; if someone argues that we ought to torture children for fun, it is safe to say that this person is not our ethical peer. Vavova goes on to argue that when it comes to *truly* peer contested issues, where philosophers who all seem competent enough within their field of expertise still disagree and continue to do so as time goes by, it is often *reasonable* to downgrade one's credence and move closer to agnosticism. With these issues, it is often independently plausible that our epistemic situation is poor, and thus downgrading our credence makes more sense than insisting that we *must* be right while our opponents *must* be wrong. Taking peer disagreement into account thus implies a reasonable amount of agnosticism when it comes to difficult moral matters, not skepticism about all or almost all of morality.

The moral responsibility debate has certainly evolved from cruder to increasingly sophisticated theories, and we have seen widespread, general trends come and go over time. Nevertheless, there does not seem to be any emerging consensus regarding the existence of desert-entailing moral responsibility. Moving on to retributivism, it seems quite obvious that criminal justice philosophers are not converging on retributivist beliefs as time goes by. We would certainly be hard pressed to argue that everyone who disagrees with retributivism thereby shows that she has beliefs that are just *too odd* for her to count as a proper peer in the criminal justice debate [3]: 167–169. The influencers could still be justified in dismissing this persistent peer disagreement if it did not depend on a “poor epistemic situation” as Vavova says, but rather on some fault in their opponents which prevents them from seeing reason. However, this is an implausible explanation.

Compatibilist retributivists might justifiably dismiss the objections of skeptics and anti-retributivists if it turned out that they all suffered from some brain disorder that negatively affects philosophical

reasoning, or that billionaires paid them to argue for these views regardless of their real opinions. But as it is, we have no such grounds to dismiss whole sides of these debates. What we have are *error theories*, but those *start* from the assumption that our opponents are wrong, and then try to explain why they persist in their wrongness. For instance, if I find Björnsson and Persson's compatibilism compelling already, I might consider it a lingering problem that people often exhibit incompatibilist intuitions. Where do these intuitions come from, if they do not track the truth? Björnsson and Persson arguably solve this problem through their error theory of incompatibilist intuitions, which explains why people mistakenly focus on the wrong causal factors when presented with certain thought experiments [33]. However, if I start out a skeptic, I have no reason to think that focusing on factors that give rise to incompatibilist intuitions is a *mistake*. The error theory explaining away these intuitions seems to stand and fall with their compatibilist theory; it does not provide us with *independent* reason to dismiss incompatibilists, the way a discovery that all incompatibilists are brain-damaged might.

It should furthermore be noted, that *if* an error theory like Björnsson and Persson's *did* provide some independent reason to dismiss incompatibilism (including skepticism), we have a *prima facie* reason to downgrade our credence whatever our position in the moral responsibility and retributivism debates. There are error theories for them all, including the influencers' compatibilism and retributivism [15, 34]. I believe, however, that it is more plausible to see all of these error theories as dependent on the positions they are meant to support.

Nevertheless, it has been argued that we can justifiably dismiss peer disagreement even if we lack independent reason to judge our opponents untrustworthy. According to this line of argument, if I know that I have conscientiously considered and weighed all the evidence before arriving at my conclusion, I can conclude that someone who disagrees with me must have made *some* mistake, even if I do not know what it is. Peter van Inwagen, somewhat infamously, even talks about relying on the assumption that my supposed peer lacks some special insight that I have access to myself [35]. However, even if this attitude can be warranted in *some* situations (I leave this open, for the purposes of this

paper), it would be absurd to take this attitude towards my *intuitions* and judgments that rely on them; to simply assume that whereas *my* intuitions offer true insight, *your* intuitions do not.^{8, 9}

When Vilhauer writes that compatibilists and libertarians should acknowledge that skeptics bring reasonable doubts to the debate, he essentially says that we should take peer disagreement into account. He writes that this is conditional, though, on us finding the debate truth conducive, and when elaborating on what this means, talks about arguments and theories improving over time in response to one's opponents' objections. I argue that compatibilists and libertarians ought to downgrade their credence in the face of skepticism *even if* they consider the philosophical debate completely stuck. The same goes for retributivists in general, and those who believe we can accurately determine the amount of punishment that offenders deserve. If the philosophical debates on these issues do not make real progress, the most plausible explanation is *not* that one's own reasonable side is relentlessly bargained by stupid objections from unreasonable opponents. Rather, it is because the arguments ultimately bottom out in intuitions, which differ between different philosophers.¹⁰

For the purposes of this paper, it does not matter how we conceive of intuitions more precisely. They might be

a special class of beliefs that are not inferred from any other beliefs, a kind of disposition to believe something rather than beliefs in themselves, or a kind of propositional attitude, presenting a proposition as true to the philosopher entertaining it (see Pust [38] for an overview). Regardless, we lack any method for ultimately resolving disagreements that depend on differing intuitions. If I believe, based on intuition, that X, and you believe, also based on intuition, that $\neg X$, we can certainly discuss the matter, draw out the implications of X and $\neg X$ respectively, see how they fit with our other philosophical commitments, engage in fanciful thought experiments, and everything else that philosophers do when disagreeing. Nevertheless, if it remains the case after all these discussions and exercises that I still believe that X and you that $\neg X$ because our intuitions *just differ*, there is no way to finally decide whether your intuition or mine tracked the truth. We do not have a truth-track-o-meter. Since intuitions are not based on any other evidence or inferred from other beliefs, we cannot evaluate them by looking at how well grounded they are either.

David Christensen agrees that assuming, based on disagreement alone, that I must have some special insight that you lack, is epistemologically irrational. Still, he concedes that doing so might sometimes have good consequences; perhaps academia needs over-confident scholars and researchers to flourish, and perhaps we would not produce much knowledge in the end if we were as quick to downgrade our credence as would be rational on purely epistemic grounds [39] (see also Everett for a similar argument [40]). Even so, Christensen goes on, there are situations where such pig-headedness is clearly terrible.

Suppose once again that I am a doctor. I believe that my patient suffers from illness A, for which I ought to give him a number of painful treatments. My colleague and peer objects, after reviewing the exact same evidence, that the patient is more likely to suffer from illness B, in which case he requires a different kind of treatment in order to save his life. In this situation, where the stakes are high, I ought not to stay perfectly confident, thinking that my colleague must lack the special insight into illnesses that I enjoy. The question of whether we ought to have Harsh Retributivism in society has very high stakes; stakes-wise, it is much closer to Christensen's doctor example than, say, Hilary Kornblith's musings on whether to remain steadfast in one's externalist beliefs upon encountering internalists

⁸ Cocchiaro and Frances point out that when philosophers define 'peer', they often include access to the exact same evidence in their definition [36]. Arguably, this means that two philosophers whose intuitions differ cannot be peers in this sense; if the content of their intuitions forms part of the evidence, their evidence differs. However, they could plausibly still be peers if they have read all the same literature, are familiar with the same arguments, the same thought experiments, and both know that different people have different intuitions regarding said thought experiments – even if their own intuitions differ.

⁹ Maja Spener makes a similar point regarding philosophical disagreement based on introspection [37]: 279. We ought to take peer disagreement seriously and downgrade our credence when it comes to introspective disagreement, even if we think it is rational to remain steadfast in the face of disagreement in most other contexts.

¹⁰ Thus, I still accept the premise that there is *some* philosophical value to be found in these debates. If a group of intelligent, sensible, well-read and honestly truth-seeking philosophers debate these matters, and get completely stuck because their intuitions just differ, we have still learnt a valuable lesson from the whole debate. We have learnt that our own intuitions, which we might initially have thought of as obvious, perhaps nearly self-evident, actually were not. My argument is still independent of the premise that the free will and moral responsibility debate is philosophically valuable in the specific sense of producing better and better arguments and theories.

[31]: 31.¹¹ Caruso makes a point similar to Christensen's in his own paper on the epistemic argument: When we ponder whether to deliberately harm other people, the epistemic standards ought to be particularly high [1].

Do We Have to Rely on Intuitions?

In the previous section, I argued that we ought to take peer disagreement seriously when the stakes are high *and* we base our beliefs on intuitions. A defender of Harsh Retributivism might respond that even if the stakes are high, we could provide arguments for having Harsh Retributivism that do not depend on intuitions, in which case the view that we ought to remain steadfast in the face of disagreement becomes more plausible. However, I do not believe this response succeeds, for two reasons: First, excising intuitions from one's philosophical arguments might be impossible, or nearly so. Second, a philosophical argument's dependence on intuitions that vary between peers is a *sufficient* reason to downgrade one's credence in the face of persistent peer disagreement, but not necessary.

There is no denying that the bulk of the moral responsibility literature appeals to intuitions in a pretty straightforward way. The influencers in my example are compatibilists in part because Frankfurt cases convinced them that moral responsibility does not require the ability to do otherwise. Upon hearing the story of Black, Jones and Smith, we are supposed to *intuit* that Jones is morally responsible for the murder despite being unable to do otherwise. This intuition is a widespread one, but not universal (David Widerker, for instance, denies Jones' responsibility [42]). Considering the rest of Fischer's compatibilism, the reasons-responsiveness conditions for moral responsibility are mostly argued for by looking at cases where someone satisfies or does not satisfy them, and then intuiting whether that person is responsible or not [18]: 69–70, 93–95. Moving on to retributivism, Moore also appeals to intuition, or the

emotions we feel upon encountering wrongdoing, in his arguments [28]: chapter 3.

Admittedly, there are scholars whose arguments are less obviously intuition-dependent. Philosophers have spun Immanuel Kant's writings on free will into both compatibilist and libertarian arguments. There are likewise less blatantly intuition-dependent arguments for retributivism in the literature [43].

Still, philosophical arguments might be ultimately intuition-dependent even when they do not explicitly appeal to intuition, as Berit Brogaard argues in her critique of Herman Cappelen [44]. Cappelen claims that analytic philosophers often do fine by relying on arguments rather than intuitions, and supports his claim with famous examples from the literature [45]. Brogaard shows that even though the philosophers Cappelen appeals to do provide arguments in support of their theses, the arguments, in turn, rely on intuitions.

I believe, with Brogaard, that philosophical arguments usually (perhaps always) bottom out in intuitions, but I will not insist on this. Suppose it could be shown that some arguments in the moral responsibility debate are truly independent of intuitions. Suppose, for instance, that we can find a Kantian transcendental argument for the existence of both free will and moral responsibility that does not rely on any claims about when people *just seem* morally responsible for their actions. There will still be massive disagreement among philosophical peers about whether this transcendental argument works. The argument will be complicated and non-obvious; it will *not* be a situation where one side can point to empirical falsehoods assumed by their opponents or clearly invalid inferences they make, and then dismiss the disagreement on these grounds. Albeit not relying on intuitions in this hypothetical case, we would remain in a similarly problematic epistemic situation; we would be stuck with a disagreement we cannot solve by conclusively proving who is right and who is wrong, and then justifiably dismiss philosophers who refuse to accept these proofs as irrational.

To sum up: If it is ever appropriate to take peer disagreement seriously, we ought to do so when our convictions depend on intuitions (or at least complicated and non-obvious arguments) while, at the same time, the stakes are extremely high.

So far, I have focused on MRE and R, the intuitions that support them, and how we should downgrade our confidence when we see that intuitions differ. I now turn to the amount of punishment that offenders deserve for

¹¹ It has been pointed out to me that in medicine, doctors normally operate under time constraints. For this reason, they might have no choice but doing the best they can with what limited amount of information they have access to. However, the same actually goes for the criminal justice system as a whole. Stephen Morse points out that we cannot wait until philosophy has resolved the moral responsibility debate before deciding what to do with criminal offenders, since there is no end in sight. We have to do *something* with criminals, even though we have not solved philosophy [41].

their crimes, and how our intuitions in this area might be wrong as well.

Braithwaite and Pettit discuss the possibility of accurately matching punishment and crime in a retributivist criminal justice system [14]: 148–151. In order to do so, they write, we would need

1. A scale of seriousness for crimes
2. A responsibility/culpability scale
3. A way of combining seriousness and culpability into a single desert metric
4. A scale for how burdensome different kinds of punishments are (for instance, 2 years of imprisonment is not necessarily twice as burdensome as 1 year)
5. And, finally, a way to match up the punishment scale with the desert scale.

They write that point 5 is the “fundamental stumbling block” (although one can certainly raise serious questions about, e.g., 2 as well [46]). They mention, only to reject, John Kleinig’s suggestion for matching up the scales: Anchoring the end points by giving the least serious crime the mildest punishment, and the most serious crime the worst morally acceptable punishment [47]. Declaring some kind of terrible punishment “still morally acceptable” and then pairing it with the worst possible crime is simply arbitrary, Braithwaite and Pettit write.¹² As previously mentioned, Robinson points out that there is cross-cultural convergence on intuitions about the *relative* seriousness of crimes [29]: chapter 1, but even if we assume that such intuitions are trustworthy, this takes us no further than point 1 on Braithwaite and Pettit’s list.

As already mentioned, few would suggest that Harsh Retributivism, in order to be justified, must find the precise right amount of punishment for each offender. However, Harsh Retributivism cannot be justified unless we can, at the very least, find the right ballpark. All of the influencers might feel very strongly that the worst crimes ought to be punished by, say, life in prison, or even a swift death, whereas no one can deserve to, say,

be maimed or tortured. However, many philosophers have much milder views on what kind of punishments offenders can deserve, and thus do not share the influencers’ intuitions here [48, 49]. Once again, this kind of disagreement should give them pause, and prompt them to entertain the possibility that their intuitions about how to match up offenders’ crimes with appropriate punishments could be way off.

Where does all this leave us? Suppose our group of legal influencers have a high level of initial credence in all three propositions: moral responsibility exists (MRE), retributivism is true (R), and we are able to find out how much punishment offenders deserve (FPD). Suppose, as well, that their credence does not reflect wishful thinking, irrational bias or the like; they have arrived at these beliefs through careful and conscientious weighing of the arguments for each position. However, considering how intuition-dependent their beliefs still are (or, at the very least, dependent on complicated, non-obvious and controversial arguments), and considering the fact that they have peers whose intuitions differ from their own, they rationally ought to downgrade their credence at least a bit. Say, for instance, that they downgrade their credence in MRE to 90%, and give a somewhat lower credence to R and FPD, reflecting the fact that they probably have a higher proportion of dissenting peers in philosophical debates about these latter issues – say, 80% for R and FPD respectively.¹³ The influencers’ combined credence in $((MRE \wedge R) \wedge FPD)$ is now down to 58% – that is, rationally, they ought to give nearly as much credence to the moral wrongness of Harsh Retributivism as to its moral rightness. If the three propositions are not completely independent, say, if the truth of MRE makes R more probable, the rational credence the influencers put in $((MRE \wedge R) \wedge FPD)$ would be a bit higher, but still worryingly low. Of course, I picked those numbers somewhat arbitrarily; nothing says it has to be exactly 90-80-80. Still, in order to get a high credence for the whole conjunction, the influencers would have to be *very* certain of each constituent, and such certainty is not reasonable in the light of philosophical peer disagreement. We consider punishment to be such serious business that we ought not to punish an individual

¹² I will not delve into their discussion of von Hirsch’s suggestions for matching punishments with crimes. Not because he mixes deterrence considerations with retributivism; in my definition of Harsh Retributivism, I allowed for the possibility that non-desert considerations play *some* part in the criminal justice system. Von Hirsch does not advocate a *harsh* system of punishment, though, and so falls outside of this discussion.

¹³ According to Philpapers’ survey, <https://philpapers.org/surveys/results.pl>, a majority of philosophers believe that free will exists; they likely believe in moral responsibility as well. My impression is that more people are skeptical of retributivism than moral responsibility. However, the exact numbers are not important.

offender unless we are convinced of his guilt beyond a reasonable doubt. If the entire system of Harsh Retributivism is in serious doubt, this ought not to be taken lightly either.

However, retributivism is a moral theory, thus making uncertainty of retributivism a case of moral uncertainty. This brings new controversies to the discussion, which I will delve into in the penultimate section of this paper.

Moral Uncertainty and Other Kinds

As with peer disagreement, the debate about moral uncertainty and what to do in the face of it is much too large to do justice to in one section of a paper. Nevertheless, once again as with peer disagreement, I will argue that if we ever ought to try to err on the side of caution in the face of moral uncertainty, we ought to do so regarding Harsh Retributivism. But first a few caveats.

Even if we disregard moral uncertainty, a version of the epistemic argument remains standing. In the face of peer disagreement, the influencers should not be too certain that moral responsibility exists. This uncertainty is not directly moral, but rather metaphysical in nature. Suppose, taking peer disagreement seriously, that they assign a 90% credence to MRE, implying a 10% risk that no one is responsible and no one deserves to be punished. This uncertainty alone should give them serious pause regarding Harsh Retributivism, considering how high the stakes are. Nevertheless, I will defend the thesis that the influencers ought to take moral uncertainty into account, making their total uncertainty *very* high, and thus the epistemic argument even stronger.

Uncertainty about retributivism, i.e., about whether the criminal justice system ought to focus primarily on dealing out just deserts, *is* clearly moral. Uncertainty of how to match up punishments and crimes seems moral as well. And trying to play it safe when faced with moral uncertainty might *often* be a bad idea; I grant that much.

Suppose, for instance, that a friend of mine values honesty, truth and fidelity. She has told me that she would want to know if a partner cheated on her, rather than living in blissful ignorance. Now I find out that her current boyfriend *has* cheated – but she is happy in her relationship, and would be miserable if I told her. I feel seriously torn; I want to do right by my friend, but I do not know whether I ought to tell her or keep quiet. Even

if I am no philosopher and do not think in these terms, we might say that I am uncertain of whether something like hedonistic utilitarianism or something more along the lines of Kantianism is right.

I think this kind of case shows that one can be morally uncertain and want to do the right thing without being a moral fetishist, as has sometimes been claimed (e.g., Weatherson [50] for the fetish argument, and Sepielli [51] for an objection). I respect my friend and her principles, but I also care deeply about her happiness; I do *not* worry about ‘rightness’ in some purely abstract sense. Still, I cannot ‘play it safe’, since there *is* no safe side: I either tell her and hurt her, or keep quiet and override her expressed wishes. It seems plausible, as well, that I cannot do a ‘moral utility calculation’ in this scenario, and maximize expected moral value [52]. How do we assess whether it is worse to Kantian-lie or to hedonist-hurt someone? If such an assessment is impossible, trying to play it safe or maximize expected moral utility in scenarios like this is just *pointless*.

Nevertheless, the legal influencers’ uncertainty about whether to support or work against Harsh Retributivism is different from my uncertainty about whether to lie or tell a hurtful truth. Even if we focus on the undisputedly moral part of their uncertainty, there is clearly a *safer*, albeit not completely safe, side for them to err on. Shaw argues that we ought to minimize moral risk by punishing someone only when the main moral theories agree that we should [3]: chapter 6. Similarly, adopting a criminal justice system that is justified (albeit for different reasons) according to soft retributivists, quarantine theorists, expressivists, consequentialists and others seems *prima facie* safer than having a Harsh Retributivist one. The softer system based on a large overlapping consensus is not *obviously* morally permissible; if the harsh retributivists are right, it would be unjust, due to giving offenders less punishment than they deserve. Still, dealing out too much punishment is plausibly worse than dealing out too little [53]. This judgment is reflected in how we regard (descriptive) uncertainty about individual offenders’ guilt in the courtroom; we ought not to punish anyone unless we are certain of their guilt beyond a reasonable doubt. If undeserved punishment and an undeserved lack of punishment were equally bad, it would make sense to abandon the ‘beyond a reasonable doubt’ standard completely, and convict and punish people as soon as their guilt was more likely than not. The thesis that too much punishment is worse than too little when the offender

does deserve *some*, is simply a reasonable extension of the thesis that undeserved punishment is worse than an absence of deserved punishment. Thus, there is a *safer* side, if not completely safe.

Caruso points out that deliberately harming people – and we all agree that this is *prima facie* wrong, we are not uncertain about *that* – is something we ought not to do unless we are very certain that it is justified under the circumstances. It is possible to agree with Caruso here, even if we think that it is normally a bad idea to try to ‘play it safe’ when morally uncertain.

Conclusion

Caruso argues that Harsh Retributivism is unjustified from the premise that compatibilists and libertarians have relatively weak arguments at their disposal. Vilhauer argues for the same conclusion, relying instead on the premise that the moral responsibility debate is truth conducive, focusing on supposed improvement over time. I have given another version of the epistemic argument, according to which Harsh Retributivism is even more difficult to justify than previously thought. I do not rely on either Caruso’s or Vilhauer’s controversial assumptions, but on the relevance of peer disagreement in debates where the arguments depend heavily on intuitions. Thus, we all have more or less controversial premises in our arguments, but in order to refute our common conclusion, the defender of Harsh Retributivism has to show that they are *all* false.

The one assumption we all rely on, is that the stakes are high in criminal justice, so the epistemic standards should be similarly high. *This* assumption, at least, should not be a controversial one.

Acknowledgements I want to thank Torfinn Huvenes, Thomas Hartvigsson and Daphne Brandenburg for helpful comments, as well as my other colleagues at the University of Gothenburg and Umeå University.

Funding Information Open access funding provided by Umeå University.

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

1. Caruso, Gregg. 2020. Justice without retribution: An epistemic argument against retributive criminal punishment. *Neuroethics* 13: 13–28.
2. Corrado, Michael Louis. 2017. Retributivism and the burden of proof. UN C legal studies research paper. Available at SSRN: <https://doi.org/10.2139/ssrn.2997654>.
3. Shaw, Elizabeth. 2014. *Free will, punishment and criminal responsibility*. Dissertation. Edinburgh: School of Law, Edinburgh University.
4. Vilhauer, Ben. 2009. Free will and reasonable doubt. *American Philosophical Quarterly* 46 (2): 131–140.
5. Vilhauer, Ben. 2012. Taking free will skepticism seriously. *The Philosophical Quarterly* 62 (249): 833–852.
6. Vilhauer, Ben. 2015. Free will and the asymmetrical justifiability of holding responsible. *The Philosophical Quarterly* 65 (261): 772–789.
7. Hedden, Brian. 2016. Does MITE make right? Decision-making under normative uncertainty. In *Oxford studies in metaethics*, ed. R. Schafer-Landau, vol. 11, 102–128. New York: Oxford University Press.
8. Strawson, Peter. 1962. Freedom and resentment. *Proceedings of the British Academy* 48: 1–25.
9. Shabo, Seth. 2012. Where love and resentment meet: Strawson’s intrapersonal defense of compatibilism. *Philosophical Review* 121 (1): 95–124.
10. Pereboom, Derk, and Gregg Caruso. 2018. Hard-incompatibilist existentialism: Neuroscience, punishment and meaning in life. In *Neuroexistentialism: Meaning, morals and purpose in the age of neuroscience*, ed. G. Caruso and O. Flanagan. Oxford: Oxford University Press.
11. Vilhauer, Ben. 2013. Persons, punishment and free will skepticism. *Philosophical Studies* 162 (2): 143–163.
12. Bennett, Cristopher. 2008. *The apology ritual. A philosophical theory of punishment*. Cambridge: Cambridge University Press.
13. Braithwaite, John, and Philip Pettit. 1992. *Not just deserts: A republican theory of criminal justice*. Oxford: Oxford University Press.
14. Watson, Gary. 1996. Two faces of responsibility. *Philosophical Topics* 24 (2): 227–248.
15. Waller, Bruce. 2015. *The stubborn system of moral responsibility*. Cambridge, MA: MIT Press.
16. Fischer, John M. 1994. *The metaphysics of moral responsibility. An Essay of control*. Oxford: Wiley-Blackwell.
17. Fischer, John M. 2006. *My way. Essays on moral responsibility*. New York: Oxford University Press.
18. Fischer, John M., and Mark Ravizza. 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

19. Pereboom, Derk. 2001. *Living without free will*. New York: Cambridge University Press.
20. Pereboom, Derk. 2014. *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
21. Van Inwagen, Peter. 1983. *An essay on free will*. Oxford: Clarendon Press.
22. Strawson, Galen. 1986. *Freedom and belief*. Oxford: Oxford University Press.
23. Strawson, Galen. 1994. The impossibility of moral responsibility. *Philosophical Studies* 75 (1): 5–24.
24. Levy, Neil. 2011. *Hard luck: How luck undermines free will and moral responsibility*. New York: Oxford University Press.
25. McKenna, Michael. 2008. A hard-line reply to Pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research* 77 (1): 142–159.
26. Frankfurt, Harry G. 1969. Alternate possibilities and moral responsibility. *The Journal of Philosophy* 66 (23): 829–839.
27. Kane, Robert. 1996. *Through the moral maze*. New York: North Castle Books.
28. Moore, Michael. 1997. *Placing blame. A general theory of the criminal law*. Oxford: Oxford University Press.
29. Robinson, Paul. 2013. *Intuitions of justice and the utility of desert*. Oxford: Oxford University Press.
30. Feldman, Richard, and Ted A. Warfield. 2010. *Disagreement*. Oxford: Oxford University Press.
31. Komblith, Hilary. 2010. Belief in the face of controversy. In *Disagreement*, ed. Feldman and Warfield, 29–52. Oxford: Oxford University Press.
32. Vavova, Katia. 2014. Moral disagreement and moral skepticism. *Philosophical Perspectives* 28 (1): 302–333.
33. Björnsson, Gunnar, and Karl Persson. 2013. A unified empirical account of responsibility judgments. *Philosophy and Phenomenological Research* 87 (3): 611–639.
34. Wiegman, I. 2017. The evolution of retribution: Intuitions undermined. *Pacific Philosophical Quarterly* 98 (2): 490–510.
35. Van Inwagen, Peter. 1996. It is wrong, everywhere, always, for anyone, to believe anything upon insufficient evidence. In *Faith, freedom and rationality*, ed. J. Jordan and D. Howard-Snyder, 137–154. Savage: Rowman and Littlefield.
36. Cocchiaro, Mariangela Zoe and Frances, Bryan. 2019. Epistemically different epistemic peers. *Topoi*: 1–11. <https://doi.org/10.1007/s11245-019-09678-x>.
37. Spener, Maja. 2011. Disagreement about cognitive phenomenology. In *Cognitive phenomenology*, ed. T. Bayne and M. Montague, 268–284. Oxford: Oxford University Press.
38. Pust, Joel. 2017. Intuition. The Stanford encyclopedia of philosophy. <https://plato.stanford.edu/entries/intuition/>. Accessed 5 November 2019.
39. Christensen, David. 2007. Epistemology of disagreement: The good news. *The Philosophical Review* 116 (2): 187–217.
40. Everett, Theodore J. 2015. Peer disagreement and two principles of rational belief. *Australasian Journal of Philosophy* 93 (2): 273–286.
41. Morse, Stephen J. 2015. Criminal law and common sense: An essay on the perils and promise of neuroscience. *Marquette Law Review* 99: 39–72.
42. Widerker, David. 2005. Blameworthiness, non-robust alternatives, and the principle of alternative explanations. *Midwest Studies in Philosophy* 29 (1): 292–306.
43. Van der Rijt, Jan-Willem. 2009. Republican dignity: The importance of taking offence. *Law and Philosophy* 28 (5): 465–492.
44. Brogaard, Berit. 2014. Intuitions as intellectual seemings. *Analytic Philosophy* 55 (4): 382–393.
45. Cappelen, Herman. 2012. *Philosophy without intuitions*. Oxford: Oxford University Press.
46. Jeppsson, Sofia. 2018. Irrational option exclusion. *Ethical Theory and Moral Practice* 21 (3): 537–551.
47. Kleinig, John. 1973. *Punishment and Desert*. Hague: Martinus Nijhoff.
48. Duff, Antony. 1986. *Trials and punishment*. Cambridge: Cambridge University Press.
49. von Hirsch, Andrew. 1994. *Censure and sanctions*. Oxford: Oxford University Press.
50. Weatherson, Brian. 2014. Running risks morally. *Philosophical Studies* 167 (1): 141–163.
51. Sepielli, Andrew. 2016. Moral uncertainty and fetishistic motivation. *Philosophical Studies* 173 (11): 2951–2968.
52. Gustafsson, Johan, and Olle Torpman. 2014. In defense of my favourite theory. *Pacific Philosophical Quarterly* 95 (2): 159–174.
53. Duus-Otterström, Göran. 2013. Why retributivists should endorse leniency in punishment. *Law and Philosophy* 32 (4): 459–483.

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