Ethical Theory and Moral Practice
https://doi.org/10.1007/s10677-020-10147-3

The Limits of Free Will: Replies to Bennett, Smith and Wallace

Paul Russell [1,2]

Accepted: 1 December 2020/Published online: 21 January 2021

1 I

I am very grateful to Christopher Bennett, Angela Smith and Robert Wallace for their insightful and illuminating commentaries on the essays collected in *The Limits of Free Will*. The exegesis and critical commentary provided not only identifies and clarifies a number of important issues, it has also further stimulated my own thoughts on this subject. In what follows I am not able to follow up on all the interesting and valuable points and objections raised but I will address several that are particularly significant and substantial.

Let me begin with Wallace's commentary and criticism, which is perhaps the most broadly based and comprehensive. Wallace presents two basic concerns that are closely related to each other Wallace 2019. There is, Wallace suggests, a fundamental tension between the two core components of critical compatibilism. The two components concerned are a neo-Strawsonian compatibilism combined with the claim that free and responsible agents are, nevertheless, subject to troubling forms of fate and luck. It is this combination that distinguishes *critical* compatibilism from more "optimistic" or "complacent" forms of compatibilism. According to Wallace, the more successful and credible its compatibilism "the less we have to be pessimistic about". On the other hand, if we press the case for pessimism then, Wallace argues, "the view risks collapsing into incompatibilism". The other basic concern that Wallace presents is that it is not obvious how this position differs from "hard incompatibilism", as developed by Derk Pereboom and others Pereboom 2007 and 2014. The two views, it is argued, share key features such that we may "wonder whether critical compatibilism is really a form of compatibilism after all". Clearly, this second concern about the distance between critical compatibilism and hard incompatibilism tracks one side of Wallace's concern that there is a fundamental instability in the position advanced (i.e. that it is liable to collapse into incompatibilism).

In order to address the two basic concerns that Wallace presents we may begin by saying more about the relationship between critical compatibilism and free will pessimism. Free will pessimism is comprised of three fundamental claims. These three claims *describe* interconnected and inescapable features of our human predicament.

✉ Paul Russell
paul.russell@ubc.ca

[1] Department of Philosophy, Lund University, Lund, Sweden

[2] Department of Philosophy, University of British Columbia, Vancouver, BC, Canada

(1) Human beings, as we observe them, possess normative capacities and abilities of a complex and robust kind such that we must recognize and acknowledge our status as morally responsible agents acting in the world. Any effort to deny or repudiate this is, indeed, a form of "bad faith". Call this the "responsibility claim".
(2) There are no credible accounts of the conditions of moral responsibility that serve to secure us against or leave us immune to the influence and operation of significant and relevant forms of fate and luck. Call this the "luck claim".
(3) Free and responsible moral agents are still subject to forms of fate and luck of a disturbing or troubling kind. Call this the "pessimism claim".

In order to locate free will pessimism in relation to other possible accounts, we may anticipate how the other parties in the (traditional) free will debate are likely to respond to this position. Each of them – the orthodox compatibilist, the libertarian and the sceptic – may argue that any such view involves a basic confusion. Both the orthodox compatibilist and the libertarian will grant the responsibility claim - although they interpret it very differently - but will deny the luck claim, since they also agree that conditions of responsible agency *cannot* be satisfied in circumstances where the agent is also subject to substantial or significant forms of fate and luck. While they both evidently reject scepticism - since they accept the responsibility claim - they agree that *if* the luck claim were true then scepticism would follow (i.e. the responsibility claim would then be false). Nevertheless, since the responsibility claim is true, they conclude that both the luck claim and the pessimism claim are false. In contrast with this, the sceptic is generally understood to accept the luck claim (under some relevant interpretation), and to conclude from this that the responsibility claim must be false. Since the responsibility claim is false, the pessimism claim must also be false. What all three parties are, nevertheless, agreed about is that the responsibility and the luck claim cannot be *jointly* true and that the pessimism claim, therefore, must be false. Clearly, then, on the face of it, all major parties in this debate, as generally understood, reject one or more of the core claims of free will pessimism.

How, then, does free will pessimism relate to critical compatibilism? According to critical compatibilism any plausible compatibilist theory must accept the core tenets of free will pessimism. Although all compatibilists accept the responsibility claim, few accept the luck and pessimist claim (and most explicitly deny them). Perhaps the most prominent example of this is found in Daniel Dennett's influential work.[1] Dennett's compatibilism describes forms of rational self-control that are in no way discredited or threatened by the thesis of determinism. It also describes a broadly "therapeutic" or "engineering" model for the correction of wrongdoers or those who might otherwise fail our moral standards. This pragmatic account of moral responsibility could be described as neo-Skinnerian in character.[2] While it provides a compatibilist-friendly way of interpreting the responsibility claim it also firmly rejects the luck claim.[3] Contrary to the "gloomleaders" of incompatibilism, Dennett argues, agents who enjoy the powers and abilities of rational self-control are not systematically or problematically subject to either fate or luck – these being familiar "bugbears" that Dennett is especially concerned to discredit.[4]

[1] Dennett 1994; and also Dennett 2003.
[2] Skinner 1971.
[3] Dennett 1994, Chaps. 4 and 5.
[4] Dennett 1994, Chap. 1

Critical compatibilism plainly rejects Dennett's (optimistic) claim that agents who enjoy powers of rational self-control are immune to general worries about fate and luck.[5] Neo-Skinnerian conceptions of responsibility, of the kind that Dennett describes, are not able to escape concerns about fate and luck in the way that justifies their optimism. Their various efforts to defuse these concerns are, the critical compatibilist claims, both evasive and unconvincing. Critical compatibilism, moreover, is not committed to a neo-Skinnerian understanding of moral responsibility. On the contrary, whereas Dennett's account dispenses almost entirely with reactive attitudes or moral sentiments, critical compatibilism retains a commitment to forms of normative competence and abilities that (crucially) include the reactive attitudes. An account of this kind could be described as neo-Strawsonian.[6]

It may be objected, with regard to this way of interpreting the responsibility claim, that any neo-Strawsonian conception of responsibility is committed to the assumptions and aspirations of what Bernard Williams has called "the morality system".[7] A crucial feature of conceptions of this kind is that they must meet the requirements of "absolute fairness", such that responsible agents somehow "transcend" all relevant forms of fate and luck.[8] There are, of course, neo-Strawsonian compatibilists who believe that the demands of "the morality system" can be met within compatibilist constraints – but this involves rejecting both the luck and pessimist claims.[9] Contrary to this view, critical compatibilism maintains that neo-Strawsonian accounts of moral responsibility may be satisfied consistent with the truth of the luck and pessimist claims. Indeed, the fact that critical compatibilism is committed to a neo-Strawsonian account of responsibility - in contrast with neo-Skinnerian views - actually serves to *amplify* the source of pessimism that is involved here. Moral agents may, on this view, be appropriate targets of (robust) reactive attitudes even though they are still subject to significant forms of fate and luck. It is this combination of claims that all adherents of "the morality system" dismiss out of hand.

With these preliminaries out of the way, let us now return to Wallace's objections, beginning with how we should understand the relationship between critical compatibilism and hard incompatibilism. Hard incompatibilism has two important components. The first is a scepticism about conceptions of moral responsibility understood in terms of what Pereboom refers to as "basic desert".[10] Basic desert conceptions involve strong metaphysical demands for ultimate agency or self-creation of some kind such that we can justify our established retributive attitudes and practices.[11] Hard incompatibilism maintains that, whether determinism is true or false, these conditions and requirements are not satisfied. This is the *sceptical* dimension of hard incompatibilism. The second aspect of hard incompatibilism is a set of arguments that are intended to justify a broadly *optimistic* understanding of the implications of this sceptical position. Scepticism about basic desert conceptions does not, Pereboom argues,

[5] Russell, *Limits*, Chaps. 10,11,12.
[6] Strawson 1992
[7] Williams 1985, Chap. 10.
[8] This is, of course, a central theme of Bernard Williams's critique of "the morality system". For an extended discussion of this see Russell, "Free Will and the Tragic Predicament". The classic defence of the view that conditions of fate and luck cannot coexist with the sort of control required for moral responsibility is presented in Nagel, "Moral Luck".
[9] See, for example, Wallace 1994: esp. 39–40, 64–6.
[10] For a full description of the elements of "basic desert" accounts of moral responsibility see, e.g., Pereboom 2007, 2–3, 127–30; and also Pereboom, "Hard Incompatibilism": 86–87
[11] For a related discussion and defence of scepticism see Galen Strawson, "The Impossibility of Ultimate Moral Responsibility".

carry any especially disturbing or troubling implications. On the contrary, we are better off without the cruel and destructive retributive attitudes and practices that accompany basic desert conceptions. We can, moreover, retain most of the features of our moral and personal lives that are actually worth preserving and protecting.

One feature that is common to hard incompatibilism and critical compatibilism is that both reject basic desert conceptions of moral responsibility. Clearly, however, it is possible to reject basic desert conceptions of moral responsibility without accepting free will pessimism. In the first place, some may argue that basic desert conceptions are the only "true" or "genuine" accounts of moral responsibility and, therefore, (local) scepticism about basic desert conceptions implies a total global scepticism.[12] Sceptical views of this kind, obviously, reject the responsibility claim. However, even those who reject basic desert conceptions but do not reject the responsibility claim, such as Dennett, may still resist the luck and pessimist claims. According this view, human agents are responsible (i.e. in pragmatic, neo-Skinnerian terms) but there is no reason to accept that we are somehow subject to luck and fate of a problematic or troubling kind. Clearly, any view of this kind also rejects free will pessimism, since it rejects the luck and pessimist claims.

So where do Pereboom and other hard incompatibilists stand on these matters? On one hand, Pereboom explicitly allows that there are conceptions of moral responsibility other than the basic desert conception. The alternative account that Pereboom is drawn to and is willing to countenance is, in fact, strikingly similar to Dennett's neo-Skinnerian (pragmatic, therapeutic) "compatibilism".[13] Despite this, Pereboom not only denies that he is a "compatibilist", he also claims that Dennett is not really a "compatibilist". They are not compatibilists, Pereboom argues, because, with respect to the traditional free will debate, the relevant issue concerns only the basic desert conception, about which both he and Dennett are *sceptics* (and incompatibilists). They are also both *optimistic* about the implications of this (local) scepticism. Given these considerations we may conclude that Pereboom converges on Dennett's "compatibilism", even though he resists being labelled this way.

It should be clear, in light of this analysis, that critical compatibilism should not be assimilated to the Dennett-Pereboom view (whatever label we put on it). It is true that they are all agreed about rejecting "basic desert" conceptions of moral responsibility. It is also true that they all still accept the responsibility claim *under some interpretation*. However, their particular interpretations of the responsibility claim vary greatly (i.e. between their respective neo-Skinnerian and neo-Strawsonian accounts).[14] More importantly, Dennett firmly rejects both the luck and pessimist claim – both of which are essential to free will pessimism. It might be argued that Pereboom, despite the strong affinities between his views and Dennett's, is not *explicitly* committed to rejecting the luck and pessimism claims.[15] Suffice it to say, however,

[12] Galen Strawson, "The Impossibility of Ultimate Moral Responsibility".

[13] Pereboom 2007, 131, 134,136, 171–4.

[14] It is worth emphasizing that while critical compatibilists are sceptical about *basic* desert, it remains committed to a desert-based view of responsibility, where desert is still interpreted in terms of its neo-Strawsonian elements and affective and motivational propensities associated with that. For relevant discussions of various complexities and difficulties relating to responsibility and the concept of desert see McKenna 2012, Chap. 6; and also Shoemaker 2015, 220–25.

[15] Although the "hard incompatibilist" *might* accept the luck claim (contrary to Dennett), accepting the pessimism claim cannot be squared with any unqualified or one-sided optimism. The hard incompatibilist might accept a *truncated* version of the pessimism claim, that allows that free and responsible agents (i.e. in non-basic desert terms) are subject to fate and luck but denies that this has any disturbing or troubling implications. Both these options, however, serve to show that hard incompatibilism and critical compatibilism diverge over significant issues. Moreover, to the extent that they converge, they articulate their respective positions in the very different terms. For a more extended discussion of these issues see Russell, "Moral Responsibility and Existential Attitudes".

that if Pereboom and other hard incompatibilists move in this direction, this would involve a *substantial* set of amendments and modifications to the hard incompatibilist position as currently articulated and defended. As things stand, it is not obvious that hard incompatibilists would endorse the free will pessimist position or are willing to amend their position accordingly.

Robert Wallace might still doubt the "compatibilist" credentials of critical compatibilism on the same general grounds that Pereboom and others have questioned Dennett's purported "compatibilism", since it too fails to "save" basic desert conceptions of moral responsibility. In reply to this, it may be argued that the traditional free will debate has not centred on basic desert conceptions in the exclusive way that Pereboom and other sceptics suggest. Clearly there are many compatibilists – including major historical figures such as Hobbes and Hume – who are not interested in preserving any narrow conception of responsibility of this kind but who, nevertheless, do *not* regard themselves as "sceptics" in relation to this matter. Having said this, the critical compatibilist can agree that the traditional free will debate, as it has evolved in modern Western philosophy, has indeed turned on a particular set of assumptions and requirements relating to this "problem" and the sort of "solution" that is allowed for. The crucial assumption – shared by most if not all the major parties involved in this debate – is that free and responsible moral agents cannot also be subject to significant or substantial forms of fate and luck. This is a "pessimistic" claim that all the established parties reject. Even the sceptic, on this view, is "optimistic" *in these terms*, since the sceptic denies that any agent is responsible *on the ground that* all such agents are subject to troubling forms of fate and luck. It is assumed, in other words, that the responsibility and luck claims *cannot both be true*. From the perspective of "the morality system" any understanding of our human predicament of the sort that *critical* compatibilism suggests is simply ruled out as an unacceptable and incoherent. Any acceptable "solution" to the free will problem must, according to this view, take the form of showing that free and responsible agents are *untainted* by any relevant forms of fate and luck.

It is this understanding of the free will problem, as generated by the assumptions and aspirations of "morality", that critical compatibilism aims to challenge. As the critical compatibilism sees it, it is quite misleading to present their view as simply "sceptical", since they plainly reject *global* scepticism. Drawing on the core elements of free will pessimism, what the critical compatibilist maintains is that there is no credible "solution" to the free will problem *as traditionally presented and framed* – and this *includes* the sceptical view. It is, more specifically, a mistake to present us as facing a choice between endorsing conditions of moral responsibility that avoid any taint of fate and luck *or* accepting unqualified, global scepticism. Although there is no form of moral responsibility that meets this standard (as per the luck claim) global scepticism does not follow (as per the responsibility claim). Accepting this as an accurate description of our human ethical predicament is, nevertheless, no basis for undiluted optimism (as per the pessimism claim). In sum, what we face is not a *problem* that requires a "solution" that will preserve the optimistic illusion that moral agency and conduct is immune to conditions of fate and luck. What we face is a troubling and disconcerting *predicament* that demands a more truthful description and acceptance. Although critical compatibilism does not sanction any sort of extreme (Pascalian) despair, we have no reason to suppose that because "free will is not an illusion" it follows that there is nothing troubling or disturbing when we confront and acknowledge the *limits* of free will.

2 II

Angela Smith provides a lucid and insightful commentary on a central theme of my work: namely, my critique of R. Jay Wallace's "narrow construal" of moral responsibility Smith 2020. Not only does her summary and exegesis pinpoint the crucial issues, Smith goes on to formulate and pose a challenge to my account and then provides a well-developed response to it. I not only agree with Smith that the challenge presented is important and needs to be addressed, I also find what she says in reply to it plausible and persuasive. My own remarks, therefore, will mostly go over these same points, not only to elaborate on Smith's remarks but also to use them to indicate the relevance of some earlier work of mine to the essays collected in *The Limits of Free Will*.

Smith's "Wallace-inspired objection" turns on an argument that holding people morally responsible is a matter of "deep appraisal" of a special kind. Although agents may be "responsible" for various other kinds of activity (e.g. artistic creations etc.) we do not expect them to be made *answerable* for their actions in the way that we do when *moral* action is at stake. In the case of moral responsibility, an agent's actions render them liable to a distinctive range of responses – including sanctions, when blame is involved. There is, according to this narrow account, a tight connection between moral responsibility and *retribution*, such that the forms of "adverse treatment" that agents are liable to raise distinct concerns about "fairness". A broader construal, of the sort that I have defended, Jay Wallace and others might argue, fails to capture this special and distinctive concern about "fairness" in relation to our reactive attitudes and retributive practices Wallace 1994.

Smith suggest that there is a "more powerful response" that can be provided to this challenge. This response turns on questioning Wallace's claim that the reactive attitudes of resentment and indignation, as occasioned by the voluntary violation of our moral obligations, are unique in their tendency to motivate harsh treatment and sanctions. There are two crucial points to be made about this. The first, Smith argues, is "that *none* of the reactive attitudes in *themselves*, can plausibly be understood as 'sanctions' [...] that we 'impose' [...] on others". Our reactive attitudes are not adopted for practical reasons but have their own "internal conditions of fittingness". Since these reactive attitudes are non-voluntary we should not think of them in terms of sanctions or rewards that are "fair" or "unfairly" distributed. On the other hand, it is also true, Smith says, that reactive attitudes do involve dispositions to treat people in a favourable or unfavourable manner. In the case of resentment and indignation, as Jay Wallace sees it, this involves a disposition to adverse treatment of some kind and it is *this* that raises the issue of fairness. In reply, Smith argues that resentment and indignation are not unique in this respect. The responses involved in reactive attitudes more broadly conceived (e.g. contempt, disdain, etc.) also involve "adverse treatment" of various kinds, even if they are not motivated by a desire to *sanction* those concerned. There are, therefore, no "special" concerns about fairness as it applies to the reactive attitudes of resentment and indignation that do not equally apply "to the full range of ethical reactive attitudes that we have towards others".

In my view, Smith's response to the "Wallace-inspired challenge" makes the crucial points that need to be made with respect to this line of reply on behalf of the "narrow construal". Since I have no criticism to make of Smith's response, what I would like to do is offer some further comment on two important points that she presents and relate them to earlier work of mine concerned with Hume's views on moral responsibility. Let me begin with Smith's interesting observations concerning the "internal" rationale of reactive attitudes. One point

of criticism Smith mentions in relation to my own account of critical compatibilism is that I need to say more about the range or scope of our reactive attitudes as this concerns finding a "middle path" between excessively narrow and overly inclusive views. What we are looking for, Smith suggests, is a "Goldilocks" set of conditions that properly identifies what *unites* our reactive attitudes – and distinguishes them from other forms of personal evaluation and emotional relationships. This is an issue that I discuss in some detail by way of criticism of Hume's views on this subject.[16]

According to Hume's theory – which I interpret as anticipating central features of Strawson's "Freedom and Resentment" – our moral sentiments (or reactive attitudes) are occasioned by our beliefs about a person's virtues and vices. Virtues and vices, Hume suggests, should be understood in terms of pleasant or painful qualities of mind. In criticism of this, I argue that Hume's understanding of virtue and vice is "too wide" and, in consequence of this, we cannot properly distinguish between appropriate and inappropriate "objects of moral sentiment". To be a proper object of moral sentiment an agent needs certain moral capacities of a more specific kind, including an ability to understand and interpret her own conduct and motivation and the emotional responses of others that such conduct may occasion. She must, in these terms, view herself as part of "the moral community".[17] On the basis of this criticism, I carry on to argue that Hume's theory is correct in encouraging us to look beyond the voluntary when we are assessing what an agent's virtues and vices may be – and that this includes, in particular, a person's feelings and desires.[18] Contrary to Hume, however, I argue that it is a mistake to construe the distinction between virtue and vice in utilitarian terms (i.e. *qua* pleasant or painful qualities of mind).

I have argued that, while it is correct to resist the "narrow construal" of moral responsibility, we should not accept Hume's particularly broad construal, which suggests that pleasant and painful qualities of mind are the relevant focus of interest for our moral sentiments. This criticism brings us back to Smith's "Goldilocks" worry: we need to say something more about what does unite the class of moral sentiments (i.e. if voluntary conduct is too narrow and pleasant/painful qualities of mind is too broad). In *Freedom and Moral Sentiment* I offer a brief account of what Hume (or we) should say about this.

> Moral sentiment is best understood, in general terms, as *reactive* value—valuing people for the way in which they do or do not manifest value for other people or themselves. Clearly this leaves a great deal unsaid. It does, however, suggest in general terms *why* we react to someone's involuntary feelings and desires no less than to their intentional actions—both can equally manifest the value which the person or agent places on the interests and concerns of other people as well as himself. This general analysis takes us in a very different direction from concern with pleasurable or painful qualities of mind. It nevertheless accords with Hume's views insofar as it does not limit moral concern to what we do or do not do voluntarily.[19]

These remarks, I believe, closely align with Smith's suggestion that what unites our reactive attitudes (or moral sentiments) is our general concern with the "evaluative assessments" an agent makes as this concerns the needs, interests, welfare and claims of others. Granted that we

[16] Russell 1995, Chap. 6: "The Content and Objects of Moral Sentiment".
[17] Russell 1995, 92.
[18] Russell 1995, 115. See also "Hume's Anatomy of Virtue".
[19] Russell 1995, 118; and see also my remarks about natural abilities at 126–7.

are concerned with evaluative assessments of this kind, there are, as Smith points out, "many things beyond an agent's voluntary choices and decisions [that] can reflect such assessments."

The only respect in which I, perhaps, demur from Smith's comments on this subject is that she describes agents as making evaluative "judgments" about what matters in these respects. It is this capacity, Smith suggests, that renders responsible agents morally answerable, since these "can be asked to *justify* these assessments" and give *reasons* for them. Although I entirely agree with the basic point that Smith is making here, which is that morally responsible agents need to be able to appreciate and understand ethical considerations and the priority or importance that they do or do not attach to them, this need not, I think, commit us to the ("optimistic") assumption that there are "reasons" available to all agents such that any failure to be moved or affected by them is a failure of (practical) *rationality*. That would be a stronger claim than I want to endorse, although I accept the weaker claim that morally responsible agents need to possess a (rational) capacity to understand and interpret ethical considerations and the weight and significance that they are giving them. A moral agent can be rational, in this weaker sense, and still fail to act on or be concerned with ethical considerations that matter (i.e. failing to give due weight to such considerations is not itself a failure of rationality).[20]

Another important issue that Smith raises in her commentary concerns the relationship between our reactive attitudes and the way in which we express or manifest them. As Smith argues, the (broader class of) reactive attitudes "motivate us to engage in, forms of 'adverse' or 'favourable' treatment".[21] On a narrow view of responsibility there is a close and unique relationship between indignation and resentment and harsh treatment and retributive practices. According to this view, harsh treatment or punishment plainly requires justification and this, in turn, requires that agents who are liable to such treatment must possess some sufficiently robust forms of control (although there is disagreement about how this condition should be interpreted). It is for this reason that the narrow view wants to restrict what we are morally responsible for to the voluntary (i.e. choices and decisions). Both Smith and I are agreed that it is a mistake to restrict moral responsibility in this way. We still have, nevertheless, several important questions that we need to address about the relationship between our reactive attitudes (broadly conceived) and the way that we express them. In particular, we need to say more about the relationship between responsibility and retribution.

Smith is right to pinpoint the general question about the relationship between responsibility and retribution as a matter of fundamental importance. It is also true that this is not an issue that I discuss in great detail in the essays collected in *The Limits of Free Will*. It is, however, an issue that I discuss in some more detail in my earlier book *Freedom and Moral Sentiment*.[22] By way of a very brief elaboration on this theme, I would like to simply note three important distinctions that are especially relevant to this issue and required for any adequate account of moral responsibility. They are:

(1) The distinction between the justification of our reactive attitudes and the justification of how our reactive attitudes may be expressed or "communicated".

[20] The point that I am concerned to make here is not that Smith's stronger interpretation of what is involved in "answerability" is inconsistent with critical compatibilism – it is not – but only that it is *not required* for it.
[21] See also Strawson, "Freedom and Resentment", 21: "Indignation, disapprobation, like resentment, tend to inhibit or at least limit our goodwill towards the objects of these attitudes, ..."
[22] Russell 1995, see especially Chap 10 ["Retributive Feeling and the Utility of Punishment"].

(2) The distinction between voluntary and involuntary expressions of our reactive attitudes, and the relevance of this to justificatory issues.
(3) The distinction between mere sanction or harsh treatment and (legal) punishment.

These distinctions not only overlap with each other, they are all relevant to Smith's aim to show that resentment and indignation do not stand in a unique relationship with "harsh treatment" and "sanctions".

In "Freedom and Resentment" Strawson suggests that being prepared "to acquiesce in the infliction of suffering on the offender" is an essential psychological component of the (negative) reactive attitudes.[23] He also points out that while there is a natural and emotional basis to retributive practices of this kind, we also "have some belief in the utility or practices of condemnation and punishment" and that this places limits on such practices when they are excessive or useless. Strawson's cursory remarks leave plenty of room for further interpretation and refinement.[24] What should be clear, in the first place, as Smith's commentary suggests, is that the considerations that serve to justify our reactive attitudes do not fully settle the question of how and when we may *express* these sentiments. For example, our sense of blame or resentment may be entirely justified but how we express it – or even if it should be expressed at all – needs to be guided by other considerations (e.g. is anything achieved by expressing these attitudes or imposing punishment?). There is, in other words, an important "justificatory gap" between the reactive attitudes (i.e. what Smith describes as their "fittingness conditions") and the practical considerations that help to inform when and how we express these feelings and attitudes.[25] The role of our reactive attitudes in these circumstances, I would argue, should be understood primarily in terms of how we *distribute* "harsh treatment", ensuring that it is not directed at the innocent or in an excessive form at the guilty (whatever the social utility of such treatment).[26]

Another important distinction arises out of the first distinction between the justification of reactive attitudes and the justification of the way in which they are expressed. Smith points out that reactive attitudes, like beliefs, are non-voluntary. To this we might add that, since reactive attitudes naturally affect and motivate us, the way in which we may *express* these sentiments may or may not be (fully) under our voluntary control. Sometimes we spontaneously and involuntarily express our emotional state, even when we are trying to curb or conceal them. Moreover, our emotions may affect our desires and motivations in a strong enough manner that it will prove very difficult not to act on them. Even though we may act voluntarily (e.g. running away due to fear in face of danger) our capacity for self-control in given circumstances may be tested – similarly with strongly felt reactive attitudes. It is also true, on the other hand, that we retain a considerable degree of control over the way in which our emotions are expressed and manifest in our behaviour and conduct. A person who is subject to moral anger, for example, may find her disposition to strike back when provoked hard to resist but we do not generally accept that retributive responses of this kind cannot be controlled, as clearly they are voluntary. It is for this reason that we need to distinguish voluntary and involuntary forms

[23] Strawson 1962, 22: "The making of this demand..."
[24] Russell 1995, 144–50, where I provide a more extended discussion of this.
[25] Russell 1995, 145.
[26] Russell 1995, 140: "The moral sentiments, it is claimed, provide us with a framework within which it may be determined who *deserves* to be punished and who does not..." The suggestion being made here is that moral sentiments do not tell us when we should punish but only when we should not punish. Desert, on this view, is a matter of *liability* to punishment or sanction.

of expressing our reactive attitudes. We expect agents to show a due degree of self-command with regard to such displays. It follows from these observations that the relationship between reactive attitudes and retributive practices or "harsh treatment" cannot be understand only with reference to the "fittingness" of the reactive attitude itself.[27] In assessing whether the response or expression was or was not itself justified we need to ask, at least, whether the expression involved was voluntary or involuntary and to what degree it could reasonably be expected to be controlled or directed in some other way.

The third distinction that may be noted here concerns the nature of the "harsh treatment" or forms of "sanction" that typically accompanies blame and other negative reactive attitudes. Smith points out that none of the reactive attitudes can, in themselves, be plausibly understood as "sanctions". Clearly it is possible to entertain reactive attitudes without expressing them at all. Moreover, we may express (voluntarily or involuntarily) our reactive attitudes without any intention or aim to sanction the offending party. Beyond this, even when we are motivated to express our reactive attitudes with a view to sanctioning the offender – making them suffer or uncomfortable in some way – such conduct does not itself satisfy all the necessary conditions of punishment. Punishment must not only involve the intentional administration of harsh treatment on a (perceived) offender for some violation or crime, it must also be administered by some relevant authority – a person or institution who enjoys the right standing.[28] Any (legal) practice or institutional system of this kind is, obviously, one that must be created and maintained and could, in principle, be radically modified or discarded. Our retributive practices and institutions should not, therefore, be viewed as a "given" or inescapable feature of human nature and society *in the same way* that our moral sentiments and reactive attitudes are (pace a neo-Strawsonian view). While we may concede that our reactive attitudes are motivationally tied to retributive dispositions, no strong conclusions follow from this that we are naturally committed to (legal) systems of *punishment*.

There is, nevertheless, a particularly close relationship between systems of legal punishment and a narrow concern with what agents do voluntarily or intentionally. The fundamental and essential (justificatory) point of systems and institutions of reward and punishment is that they encourage some forms of conduct and discourage others. Punishing individuals for what is not chosen or done intentionally would, as H.L.A. Hart has argued, come at an unacceptable cost. The legal system is a particular form of social control. It provides a framework within which "each individual is given a fair opportunity to choose between keeping the law required for society's protection or paying the penalty".[29] A system of this kind gives individuals the ability to decide beforehand when the law's punishments will interfere with them. The right lesson to draw from this, I suggest, is that we concede to the "narrow" theorist that (legal) *punishment* has a special connection with the voluntary and there is a rationale for this restricted concern. It does not follow from this, however, that moral responsibility, as such, should be narrowly understood in these terms. On the contrary, as Smith has argued, our reactive attitudes are concerned more broadly with an agent's "evaluative assessments" and so extend beyond the scope of the voluntary.[30]

[27] For related observations on this issue see Shoemaker 2015, 220–23.
[28] Hart 1968, 4–6.
[29] Hart 1968, 22–3, 46–9, 180–83. Hart comments on the relevance of understanding "legal threats" in terms of a "choosing system" and liberal politics. On this point, see also Williams 1993, 65–7.

[30] For reasons already explained, various forms of "harsh treatment" or "sanction", as associated with the expression of our reactive attitudes, are not unique or peculiar to our responses to voluntary conduct. An illuminating discussion of the difference between the harms involved in expressing our (negative) reactive attitudes and punishment, properly understood, is provided in McKenna 2012, esp. Chap.6 (esp. 141–6).

The three distinctions that I have described above are, I believe, all relevant to Smith's scepticism about the "narrow construal" of moral responsibility and support her general criticism of the suggestion that resentment and indignation have a "unique" connection with our retributive attitudes and practices. Our reactive attitudes, both narrowly and broadly conceived, present us with a "justificatory gap" between our reactive attitudes and retributive practices (i.e. understood in terms of the way we *express* our reactive attitudes). In the case of both broad and narrow accounts, we need to allow for the fact that reactive attitudes may be voluntarily or involuntarily expressed and we encounter similar issues of control and justification with regard to them all. Finally, although there may be a case to be made for a special connection between voluntary conduct (involving choice and decision) and the practice of (legal) punishment, it is a mistake to draw the conclusion that this is evidence that *moral responsibility* should be narrowly understood. We are, as the broad interpretation suggests, morally responsible for more than our choices and decisions and liable to "harsh treatment" – although not to (legal) punishment - when our failure to give due or proper value to the needs and interests of others, as may be manifest in our desires and feelings, occasion negative reactive attitudes.

What is provided above is the barest outline of my own understanding of these issues. The three distinctions that I have described are, nevertheless, entirely consistent with the central thrust of Smith's commentary and they also help to bring to light the close connections between my earlier study of Hume's theory of moral responsibility and the issues taken up in the essays collected in *The Limits of Free Will*.

3 III

Christopher Bennett offers a number of acute observations and criticisms relating to my discussion of motivational scepticism and practical reason ("Practical Reason and Motivational Skepticism"; *Limits*, Essay #7) Bennett 2019. One his particular concerns is my account of what Bennett describes as "moderate naturalism" – what I call type-naturalism - and how it relates to the issue of practical reason. According to "moderate" or type-naturalism, human beings are creatures of a certain kind and, as such, we are naturally liable or prone to emotional responses in relevant circumstances. (For example, like many other animals, we are prone to fear in circumstances where we perceive danger or a threat of some kind.) While particular tokens of our emotions may or may not be justified by our circumstances (i.e. "fit" or "appropriate"), there is no question of having to justify our more general liability or disposition to this emotion, which rests with our natural commitment to caring about or valuing certain things. This distinction between our general liability to a given emotion and the "fittingness" of the individual tokens of emotion shows where there may or may not be some relevant sceptical problem. While it is possible that our token emotions (e.g. fear, reactive attitudes, etc.) could be *systematically* discredited, there is no similar sceptical threat for our general liability to these emotion – as they are a "given" our human nature.[31]

Bennett elaborates on this distinction, in sympathetic terms, by pointing out that moderate naturalism helps us to understand the way in which reason, emotion and morality are

[31] This claim is, however, consistent, with recognizing that the particular *form* these emotions take can vary greatly from one society or culture to another. The norms that structure and inform these emotional responses can also change and evolve *within* any particular society or culture over time – even though the basic emotions remain a constant of our human nature. Our reactive attitudes are a notable example of this.

interconnected. While there "is no questioning of the rationality of the perspective presented by these emotions" [qua type], this is consistent with assessing emotions [*qua* token] as being fitting or appropriate to their situation or formal object. Bennett refers to this as a "restrictive" understanding of "moderate naturalism". He summarizes this view as follows:

> As I am understanding the restrictive version of moderate naturalism, the emotion is 'reasonable and appropriate' in virtue of the fact that the situation accords with the 'formal object' of the emotion. Furthermore, say that E typically prompts agents to action-type A (A is the action-tendency of the emotion). On this restrictive interpretation, it follows from the fittingness of the emotion that I have a pro tanto justification for doing A given the fittingness of feeling E.

Bennett calls this interpretation the Practical Reason from Formal Object Fittingness or PRFOF and he claims that PRFOF is false.

Why is PRFOF false? It is false, Bennett argues, because it is not sufficient to justify action. An emotion that satisfies PRFOF may not give us even a pro tanto reason to act on that emotion (e.g. to A) because, among other things, there may be emotions that we are susceptible to but "we ought not to have" (e.g. malice, spite, etc.). More importantly, given that we are susceptible to emotions of various kinds we need practical reason to adjudicate conflicts that may arise between them. There may be circumstances in which we experience emotions that have action-tendencies that are incompatible with one another or where "an emotion is inappropriate because it precludes some other emotional response". The version of moderate naturalism that I offer, Bennett suggests, fails to provide for the essential role that practical reason must play in educating and cultivating our emotions and adjudicating between them when they come into conflict.

The first thing to say in reply to this line of criticism is that it is not evident that "moderate" or "type- naturalism" is committed to PRFOF. In particular, while we may agree that our emotions (such as reactive attitudes) typically involve motivational responses of some relevant kind, the relationship between emotion and action is not as straightforward as PRFOF suggests. While some *range* of action-types may be associated with given emotions, there is no tight or obvious connection with a "pro tanto justification for doing A given the fittingness of feeling E". We need, as mentioned further above, to distinguish between justifying an emotion in terms of it being fitting to its formal object and justifying how that emotion may be expressed in our conduct and action. Some modes of expression may not be actions of any kind (e.g. fainting in face of danger or experiencing fear). Even when some voluntary and/or intentional action is motivated there may be no single type or kind of action that is correlated with it (e.g. fear may result in screaming, running away, grabbing the person beside us, and so on). Given these complexities, all that moderate naturalism is committed to is the claim that emotion of a given type involves a complex *syndrome* of cognition, affect and motivation, and this may serve to *make sense* of certain forms of behaviour or conduct that we associate with (experiencing) that emotion. But there is no "pro tanto justification" provided for (voluntary) action of any specific kind, even though the emotion may be fitting or appropriate to its formal object. For this reason, there remains scope, within the framework of moderate naturalism, as I understand it, to further assess and "adjudicate" how we may react (in action) in a given emotional situation.[32]

[32] I would agree with Bennett that *some* proponents of "moderate naturalism" appear to accept PRFOF. See, e.g., Smith 1759/1976, Part II; and Mackie, "Morality and the Retributive Emotions", who argue that there is no justificatory gap between resentment and retributive action. Hume, however, rejects this view, as I explain in *Freedom and Moral Sentiment*, Chap. 10.

Bennett, as we have noted, rejects PRFOF because it is not "sufficient to justify action". In order to adjudicate and resolve conflict between emotions, practical reason needs to be able to guide and inform us about "what really matters in a situation and what the right way to respond to what matters in it is." This process, as Bennett describes it, has two stages. It cannot, as Bennett points out, start from nowhere and must *begin* with existing attitudes and motivational attachments of some kind. Nevertheless, while reflection may start from emotions, it can "go beyond them". After we start moving along the track of assessing what really matters in situations of this kind, our reflections "get their own momentum" and this may "lead to the development of concepts and forms of argument the validity of which participants in the practice no longer take to rest on the emotions." In these circumstances, Bennett continues, our reflective inquiries acquire "autonomous momentum" and provide a "justificatory perspective" that is "external to the perspective of the emotions". Bennett claims that "Humean content-scepticism" involves denying that practical reason can perform these tasks and that it takes practical reason to be limited to "inquiring into the satisfaction of elements of our subjective motivational set". More specifically, understood this way, Humean content-scepticism involves setting a priori limits on the scope and power of practical reason and thus "blocks the road of inquiry".

Clearly, there is a great deal of material to cover in respect of these claims. For our present purposes the following points are especially relevant. First, it is significant, as Bennett concedes, that our inquiry into what really matters in situations of this kind must *begin* with our existing attitudes and emotional orientation. This is, from any point of view, the relevant platform from which all such reflection must proceed. Second, it is not obvious, as Bennett acknowledges, that "the Humean" lacks any resources to provide for the sort of "reflection" and "adjudication" in respect of "what really matters". Hume was certainly alive to the fact that our emotional responses in given circumstances can involve instability, ambiguity and conflict.[33] This is apparent, for example, in Hume's account of the "general point of view" as this concerns our efforts to find a shared perspective for the purpose of moral evaluation.[34] It is significant that Hume makes clear that "reason and sentiment concur in almost all moral determinations and conclusions", and that he is firmly of the view that reason has practical influence by way of altering and amending our emotions (passions) and, *through that channel*, it can influence human motivation and action.[35] To the extent that we take Hume to represent the classic statement of "Humeanism", it would misrepresent his position to read him as denying that reason serves as a guide to reflection and adjudication in circumstances where our emotional responses are unstable, variable or generate practical conflict.[36]

[33] There are two especially important sources of this on Hume's account: variations in our sympathies (from one person to another) and variations in how we may stand in relation to an object or situation (which can vary over time). One point that Hume is especially concerned to make in respect of these matters is that "so little are men govern'd by reason in their sentiments and opinions that they always judge more of objects by comparison than from their intrinsic worth and value" (*Treatise*, 2.2.8.1) Hume is no less clear, however, that reason can correct and adjust for these biases and conflicts.

[34] Hume 1739/2000, 3.3.1; and *Enquiry of Morals*, 5.41–3.

[35] Consider also the importance that Hume attaches to reason in respect of "the standard of taste", where reason and sentiment also operate together and influence each other (*Essays*, 226–49). It is worth noting, in this regard, that just as reason can influence emotion, so too emotion can influence reason (e.g. by making some considerations salient and significant in our reflections). I have argued that this two-way relationship is relevant to understanding the nature of moral competence and the conditions of moral responsibility ["Responsibility and the Condition of Moral Sense"; *Limits*, Essay #4].

[36] Hume's own views on this subject, as Bernard Williams points out, were more complex than his critics sometimes suggest. [Williams, "Internal and external reasons", 102.]

It is, perhaps, especially important to take note of Williams' observation that we should not regard an agent's "subjective motivational set", which constitutes the relevant basis for having a practical reason of some kind, as itself "statically given".

> The processes of deliberation can have all sorts of effect on S, and this is a fact which a theory of internal reasons should be very happy to accommodate. So also it should be more liberal than some theorists have been about the possible elements in S. I have discussed [set] S primarily in terms of desires, and this term can be used, formally, for all elements in S. But this terminology may make one forget that S can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent.[37]

A rational agent has a capacity, understood in these richer terms, to deliberate about a wide range of considerations and priorities, whereby she is able to find acceptable solutions to emotional and practical conflict or ambivalence. Such an agent may also recognize that there are "irresolvable conflicts" and that there is no single or entirely satisfactory solution available in all circumstances. In light of these considerations, we can provide a robust and satisfactory account of how we can "adjudicate" between conflicting motivations, and "educate" our emotional responses, without having to "reach beyond" our subjective motivational set for the purpose of critical reflection.

It may be said, in reply to this, that "moderate naturalism", so described, still fails to provide adequate (rational) resources for "deep evaluation" of our existing emotional and motivational orientation. Practical reason, understood in stronger terms, can place us in a position *outside* of these contingent commitments, and only then are we in a position to decide "what really matters" and what our practical priorities truly ought to be.[38] This ability requires us to "go beyond" our existing subjective motivational set in a more radical way, when practical reason operates "autonomously". The Humean content sceptic is, I would agree, sceptical about ambitions of this kind. However, I do not believe that this is an a priori prejudice. It is, rather, a matter of finding the proposals made along these lines obscure and/or psychologically implausible. More importantly, for the Humean content-sceptic our ability to reflect on emotions and consider what really matters does not depend on (mysterious) metaphysical powers of this kind.

I have suggested, in reply to Bennett's criticisms, that the moderate naturalist has adequate (psychological) resources to account for the "education" of our emotions and our ability to reflect on "what really matters". There remains, however, the issue concerning motivational scepticism, which needs to be treated separately. Bennett denies that there is any gap Korsgaard's argument concerning pure practical reasons and motivation. According to Korsgaard, *if* there are pure practical reasons (i.e. reasons that do not depend on pre-existing motivations of any sort), then such reasons serve to explain the agent's motivation to act on them. Grasping the *authority* of such a reason (*qua* its content) will be "sufficient to lead to changes in motivation". It is, Bennett agrees with Korsgaard, sufficient because such an explanation operates in "the logical space of reasons". To seek some further causal explanation or psychological mechanism when the agent is motivated by their grasp of reason is to fall prey

[37] Williams, "Internal and external reasons", 105.

[38] An illuminating discussion of "deep evaluation", understood in terms of the "re-evaluation of our most basic evaluations", is presented in Taylor, "Responsibility for Self".

to "a category mistake". Any such demand involves conflating two distinct orders of explanation: the logical space of reasons and the logical space of law-like causes. There is "no metaphysical mystery" about pure practical reasons motivating agents, Bennett maintains, so long as we take note of "a discontinuity between orders of explanation".

According to Bennett, the demand for some explanation for how pure practical reasons could motivate (rational) agents is a demand to find some further "non-rational" element or "new motivation" of some kind. This is not, however, the way I would formulate the relevant concern. It is not some extra non-rational element that is sought after here but rather an explanation for how the *content* of a pure reason (detached from the agent's pre-existing motivation) serves to motivate and produce activity and change in the world.[39] What is *about* pure practical reasons that makes this possible? Put another way, how does the *normative authority* of (pure) practical reason achieve *causal* traction in the world? To appreciate this concern consider the following analogy.[40] Let us allow that a flame will ignite and burn a flammable object. There are, of course, circumstances where, for some reason, the object may fail to ignite and burn and we may seek some relevant explanation for this (e.g. the object is wet or damp). We may, however, want to know how, more generally, flames cause things to ignite and burn. One theory that we might put forward is that flames contain heat and it is *this* that explains what it is about flames that can produce this effect. If it is suggested that there are *cold* flames that can ignite and burn objects it seems legitimate and reasonable to ask how this is possible. It will not suffice to reply that flames must ignite flammable objects, otherwise they would not be flames. We are looking for an account that explains what it is *about* (cold) flames that produces ignition and burning. Similarly, we are looking for a theory about (pure) practical reasons that explains how they produce motivation and action. In the absence of any relevant connection between reasons and the agent's existing motivation we have a puzzle or "metaphysical mystery".

I would agree with Bennett that there is a problem here concerning two "orders of explanation". However, the worry about motivational scepticism is not a matter of *conflating* these two levels of explanation but rather of asking how they are *related*. This brings us to the question of how much metaphysical baggage the concept of a pure practical reason needs to carry? The account provided may be read as relying on a distinction between two modes of interpreting human action: "understanding" as opposed to "explanation", where the former mode of interpretation is taken to rule out *causal* explanation altogether. According to views of this kind, there is a distinction to be drawn between *rational* explanation of action and *causal* explanation of action, in terms of which the former cannot be reduced to any form of the latter. Related to this point, it may be argued that reasons for action should not be conceived of in (law-like) causal terms.[41] Suffice it to say that if we need to rely on metaphysical apparatus of this kind then, as I see it, much more needs to be said about how to *integrate* these (distinct) orders of explanation into a unified and coherent world view. In its fundamentals, this is a problem about how – or whether - naturalism and normativity can accommodate each other.

[39] There are, obviously, other mental states that have *content* of some kind (beliefs, memories, etc.) but are, nevertheless, motivationally inert (i.e. do not satisfy any "internalist" requirement). In respect of practical reasons that serve to satisfy some pre-exiting desire, we have some sort of *theory* about *how* they differ from these other mental states and *why* they satisfy the internalist requirement. In the case of pure practical reasons we lack any such theory – and it is *this* that concerns the motivational sceptic.

[40] This analogy is presented and discussed in *Limits*, 129–31.

[41] A particularly influential Humean response to this claim, arguing that reasons can be causes, is provided by Donald Davidson in his *Essays on Actions and Events*.

Even if we accept the possibility of pure practical reasons, more needs to be said about this beyond rejecting motivational scepticism.[42]

I would like to close by pointing out that the problem of practical reason is, in my view, intimately connected with the problem of free will and moral responsibility – although this is not always fully acknowledged in the relevant philosophical literature. The contemporary debate has focused heavily on the question of how to interpret our abilities or powers of "rational self-control" and "reason-responsiveness", where abilities and powers of this nature are required for responsible agency. Several prominent statements of the "new compatibilism", as developed along these lines, have argued that, in respect of rational self-control and reason-responsiveness, we need to distinguish between the agent's ability to grasp (or recognize) a practical reason and the ability to apply (or translate) it into action.[43] It may be argued, however, that these accounts of rational self-control or reason-responsiveness are not entirely clear or explicit about how they understand the practical reasons that agents may or may not recognize and react to. It is here that the debate about the *nature* of practical reasons, and whether or not our responsiveness to them "is beholden to the motivational force of pre-existing desires and inclinations", is particularly relevant. Korsgaard (and Bennett) maintain that pure practical reasons are not subject to these (metaphysical or psychological) limitations. They also maintain that there is no distinct or independent problem about *motivation*, as long as some plausible account of the *content* of such reasons can be provided (i.e. if we can recognize reasons of this kind there is no further need to explain *how* the agent is *moved* by such a reason). My criticisms of Korsgaard's way of addressing this issue does not involve denying (a priori) the *possibility* of such pure practical reasons, nor even that it is possible that reasons of this kind might be able to motivate rational agents. What I do deny, however, is that Korsgaard's argument provides any convincing explanation for how reasons of this kind would motivate.[44]

Lund University/University of British Columbia.

References

Bennett C. (2019) Russell on naturalism and practical reason. Ethical Theory Moral Practice [Online]
Dennett D (1984) Elbow room: varieties of free will worth wanting. Oxford University Press, New York
Dennett D (2003) Freedom Evolves. Viking Press, New York
Davidson D (1980) Essays on actions and events. Clarendon Press, Oxford
Fischer J, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. Cambridge University Press, Cambridge
Hart HLA (1968) Punishment and responsibility: essays in the philosophy of law. Clarendon Press, Oxford
Hume D (1739/2000) A Treatise of human nature. Norton DF, Norton MJ (eds) Oxford University Press, Oxford
McKenna M (2012) Conversation and responsibility. Oxford University Press, New York
Pereboom D (2014) Free will, agency, and the meaning of life. Oxford University Press, New York
Pereboom D (2007) Hard Incompatibilism. In FischerJM, Kane R, Pereboom D, and Vargas M. Four Views on Free Will. Blackwell, Oxford

[42] This was a problem that Kant recognized but failed to solve in any convincing way (on this see my remarks in *Limits*, 129n9). Put in Kantian terms, the puzzle is how the *causality* of pure reason relates to the world of phenomena, which is governed by law-like regularities?

[43] See, for example, Wallace 1994, 12–3,86,157,159-66,219–22. Wallace describes his view as a "broadly Kantian approach". For another account, that relies on a similar distinction, see Fischer & Ravizza 1998, 41–6.

[44] I am grateful to Ezio Di Nucci and Andras Szigeti for their helpful comments and suggestions on this paper.

Russell P (1995) Freedom and moral sentiment: hume's way of naturalizing responsibility. Oxford University Press, New York
Shoemaker D (2015) Responsibility at the margins. Oxford University Press, New York
Skinner BF (1971) Beyond freedom and dignity. Knopf, New York
Smith A (1759/1976) The Theory of Moral Sentiments, Raphael DD & Macfie AL eds. Clarendon Press, Oxford
Smith A. (2020) Responsibility, reactive attitudes, and 'the morality system'. Ethical Theory Moral Practice 653 [Online]
Strawson PF (1962) Freedom and resentment. Reprinted in the philosophy of free will. Russell P & Deery O, (eds) New York: Oxford University Press, 2013 (63–83)
Wallace RJ (1994) Responsibility and the moral sentiments. Harvard University Press, Cambridge, MA
Wallace R. (2019) The tension in critical Compatibilism. Ethical Theory Moral Practice [Online]
Williams B (1985) Ethics and the limits of philosophy. With a commentary byMoore AW and a foreword by Lear J. Routledge, London & New York, 2006
Williams B (1993) Shame and Necessity. University of California Press, Berkeley, CA