

The Tension in Critical Compatibilism

Robert H. Wallace
University of Arizona

Abstract: Paul Russell's *The Limits of Free Will* is more than the sum of its parts. Among other things, *Limits* offers readers a comprehensive look at Russell's attack on the problematically idealized assumptions of the contemporary free will debate. This idealization, he argues, distorts the reality of our human predicament. Herein I pose a dilemma for Russell's position, critical compatibilism. The dilemma illuminates the tension between Russell's *critical* and *compatibilist* commitments. The problem is not obviously insurmountable, and as a compatibilist who is sympathetic to the view, my aim is to spark further discussion.

1. Introduction

In the introduction to *The Limits of Free Will*, Russell writes that “the collection, considered as a whole is, I believe, much more than simply the sum of its parts” (2017: xvi). I agree. The most exciting aspect of Russell's work as exemplified in *Limits* is his attack on the methodological assumptions of the contemporary free will debate.

What characterizes this attack? Russell (2018) quotes Bernard Williams as saying of his overarching philosophical aim, that it was: “to make *some* sense of the ethical as opposed to throwing out the whole thing because you can't have an idealized version' (Williams's emphasis).” Likewise with Russell. His overarching aim is to shed light on some idealized assumptions about freedom and responsibility. This idealization, he argues, tempts us to throw the whole thing—our being free and responsible—out. Worse than this, this idealization distorts the reality of our human predicament.

In what follows, I will not focus on a specific essay or argument in *Limits*. Instead, I will offer a brief overview of Russell's critical compatibilism. I'll then argue that there is a tension between the *critical* and the *compatibilist* components of the view. A dilemma problem illuminates this tension. As a compatibilist who is sympathetic to Russell's project, I see this dilemma as a way to further explore critical compatibilism. The problem is not insurmountable, but Russell should say more. I conclude by suggesting that one horn of the dilemma, which emphasizes compatibilism at the cost of pessimism about free will, is attractive.

2. Critical Compatibilism

The free will debate typically proceeds in light of a skeptical problem: we might have to throw the whole thing out—our being free and responsible—if determinism is true. Russell's critical compatibilism suggests that there is a set of compatibilist-friendly conditions such that, when met, an agent's actions are free and responsible. But Russell denies that, for any action that meets these conditions, it cannot be the member of the set of actions that are subject to fate and luck. Thus, he accepts a version of compatibilism that involves pessimism about the kind of free will which is invulnerable to fate and luck (2017: 262-263). So, in very broad strokes, Russell's non-idealized theory

has two components: a compatibilist theory of responsibility and a critical take on metaphysically demanding forms of agency.

Let's start with the theory of responsibility. P.F. Strawson, on whose work Russell builds in several essays in *Limits*, (in)famously argued that determinism would not be grounds for excusing or exempting from moral responsibility. Why? Moral responsibility consists in being the appropriate target of moral emotions, and in particular, reactive attitudes like resentment and indignation. He argues that these emotions are not made inappropriate by determinism, as they are made inappropriate when, for instance, you find out someone hit you on accident. Some have worried that Strawsonian compatibilism fails because we can find token instances of pleas which do generalize if determinism were true. So, we should either throw the whole thing out or accept libertarian free will!¹

This is a powerful skeptical challenge to the kind of naturalism about freedom and responsibility both Strawson and Russell endorse. First, let's consider Russell's diagnosis of what went wrong in Strawson. Russell calls some of Strawson's claims token-naturalist. In places, Strawson seems to think that determinism would not be grounds for dispelling any token instance of a reactive attitude. This is implausibly strong, as the objection shows (see 2017, chap. 3). But Russell thinks that Strawson also developed a *type*-naturalist strategy, which focuses on our proneness to such attitudes. Here, the argument is that the skeptical challenges would not dispel reactive attitudes at the type-level. It is this sort of strategy that Russell develops against the skeptic.

Russell admits that this move does not offer an easy answer to skepticism internal to our moral system. But type-naturalism does allow one to ward off a kind of rationalist skepticism. For if we have an innate proneness to such attitudes, we can deny that we require some external rational justification for them, and so for our practices of responsibility.

To get a grip on why this response is compelling, let's consider Russell's criticisms (2017, chap. 6) of R. Jay Wallace's (1994) amendments to Strawson. Wallace restricts the Strawsonian strategy to the moral emotions tied to deontic morality. So, Wallace denies that there are positive reactive attitudes (such as gratitude). He moreover evaluation of character (and so admiration) from our moral responsibility practices. For Wallace, moral responsibility is a matter of moral expectations and the emotions which are reactions to a failure to meet them, e.g., resentment.

Russell argues that Wallace's construal of moral responsibility is parochial and suggests that this is highly problematic. For one, by only focusing on negative reactive attitudes, he misses the constructive role praise and admiration play in supporting worthy actions and character (2017: 104) Wallace's view moreover seems to entail that moral systems which lack deontic moral notions lack a sense of morally responsible agency. But this can't be right. Consider, for instance, a set of moral responsibly practices focused around shame and character—perhaps as in ancient Greek ethical thought— as opposed to guilt and action (2017: 105-106).

Most importantly, going narrowly deontic robs the Strawsonian of the type-naturalist line of reasoning. If Wallace is right that moral responsibility is constituted by the moral emotions tied to deontic morality, and if there are systems of ethical thought which lack these features, then it seems like there is no set of reactive attitudes to which all persons, universally, are liable. No common set of reactive attitudes serves as “a natural foundation for all recognizably human forms of ethical life” (Russell 2017: 105). Such a natural and universal foundation is needed. Why? “Confrontations with other cultures and forms of ethical life will place us in the position of needing external rational justification for the entire framework of moral responsibility so conceived.” (2017: 105). Skepticism thus opens up at the type-level. We can be asked to justify whole classes of attitudes—the one's which constitute our parochial conception of moral responsibility— rather than specific instances of them.

¹ In addition to Russell's articulation of this problem in (2017: 34-45, 76, 99, 193), see also Wiggins (1973), Bennett (1980), and Watson (1987). Russell (1995: 179) suggests that Hume's account of responsibility faces a similar objection.

So, what kind of type-naturalism should we endorse? Russell offers the following view. First, we should deny that all moral responsibility systems employ narrow deontic notions like obligation. Instead, we should adopt the view that the pertinent moral emotions at play in responsibility systems are *ethical* reactive attitudes. These are “emotional responses to the weight and value given by an agent or person to ethical considerations widely conceived”, involving “thick” language terms like “being creepy” or a “cad” (2017: 109; cf. Williams 1985: 129-130). This set of attitudes is wide enough to accommodate gratitude, admiration, shame, and so on. Russell claims that these responses are “recognizably contiguous with our own conception of moral responsibility.” (2017: 112). Even if excuses and exemptions from moral responsibility *narrowly* construed might generalize in each token instance given determinism, this would not mean we would not have a propensity towards ethical reactive attitudes, and so to responsibility broadly conceived. We would remain, in spite of this token disengagement, “committed to ethical reactive attitudes at the type level.” (2017: 115).

So, we shouldn’t tie responsibility judgements to what Williams (1985) pejoratively called *the morality system*, whose demands preclude fate and luck. Rather, responsibility should be tied the wider range of human values by which we might assess agents. This, in turn, helps us make sense of the strange and disturbing cases where we feel that an agent is both free and responsible *and* subject to fate and luck.²

Notice that is a local form of free will skepticism, skepticism with regards to one sort of freedom which agents might have (cf. 2017: 262-264). And here, we move on to the second component of Russell’s critical compatibilism. What the critical compatibilist accepts, like the classic free will skeptic, is that there are no actions that are free from luck and fate. Indeed, Russell’s accepts that “fate and luck are infused into the very exercise and operation of our moral capacities as agents” (2017: 259). We lack freedom at the “horizon”, involving ultimate control over the factors that settle our conduct and character (2017, chap 11).³

Nevertheless, Russell denies, *pace* orthodox compatibilism, that fate and luck rule out responsibility. Indeed, he worries that orthodox compatibilism involves unsatisfactory evasions on the lucky origins of our powers of control. A such, we compatibilists should adopt *local free will skepticism*, the denial of the sort of freedom and responsibility encouraged by the morality system. Russell argues, however, that *global free will skepticism* is a form of bad faith because it denies the importance of the luck-infused agentic capacities that human persons clearly have and can exercise (2017: 263). These capacities do in fact give rise to meaningful ethical evaluation in terms of naturalism-friendly ethical reactive attitudes. Skeptical philosophers are too quick to ignore these capacities in light of the overly-idealized (and so overly-demanding) morality system.

What do these capacities look like? Russell offers a picture of “middle-distance” forms of control. He writes that “the responsible agent must be able to feel and understand moral sentiments and reactive attitudes” (2017: 55). He sees this sentimental capacity as being connected with other compatibilist-friendly accounts of the pertinent capacities, like reasons-responsiveness (2017: 46-66, cf. Fischer and Ravizza 1998) or theories cast in terms of the “mesh” between the elements of an agent’s psychological architecture (2017: 194, cf. Watson 1975).

Nevertheless, Russell’s view is that we should accept free will pessimism. It is disconcerting to realize that one does not settle the nature of one’s conduct and character. But the boon gained is open-eyed reflection on the contingency of the human life. The classic free will problem is, as he puts it, “not a *problem* to be solved but a *predicament* waiting to be recognized and acknowledged” (2017: 273). The proper response to this is not “Pascalian despair” but rather “a more modest sense of being *disconcerted*” (2017: 238).

² For a famous case, see Watson (1987).

³ We can even imagine a “cosmic” level of freedom involving absolute, God-like self-creation (2017: 239).

3. The Convergence Between Critical Compatibilism and Hard Incompatibilism

With this brief overview in place, here is the tension in critical compatibilism. In adopting local free will skepticism, there is convergence between critical compatibilism and forms of incompatibilism which argue that both determinism and indeterminism rule out free will. I'll explain in a moment. Given this, critical compatibilism faces a problem. The more the critical compatibilist distinguishes their position from this kind of incompatibilism, the less it seems like they ought to endorse free will pessimism. Why? The more robust our conception of freedom and responsibility is, the less we have to be pessimistic about. On the other hand, the more they push the case for free will pessimism, the more the view risks collapsing into incompatibilism.

Here's how the two views converge. Following Russell, local free will skepticism involves a denial of the kind of free will required for basic desert (2017: 269-270, ft. 41). For an agent to deserve praise or blame in this basic sense is for her to deserve it *just because* she knowingly acted in such-and-such a way, rather than deserving praise or blame on the basis of some set of consequentialists or contractualist principles (cf. Feinberg 1970, Pereboom 2014). A local free will skeptic denies we have free will understood as the kind of control over our actions that would entail our basically deserving anything. Hard incompatibilism, as defended by Pereboom (2014: 3-4) is the position that physical determinism is incompatible with free will, and that either physical determinism obtains or a form of indeterminism incompatible with free will obtains. In either case, no one has free will and so no one is morally responsible. This is a form of local free will skepticism, as Pereboom argues that the kind of free will which we lack is the sort required for basic desert.

So, both views endorse local free will skepticism. Interestingly, Russell additionally suggests that a hard incompatibilist can accept the non-skeptical features of critical compatibilism. To a great degree the two views can "accommodate" each other. They disagree with respect to what kind of attitude we should have towards this skepticism. A hard incompatibilist is typically optimistic about the implications of skepticism, seeing it as a way out of the cruel vindictiveness of retributive morality (2017: 92-93, 269-270, ft. 41). A critical compatibilist can accept some of this optimistic story but will insist that some degree of free will pessimism is warranted (2017: 269-270, ft. 41).

What are these non-skeptical features of critical compatibilism that hard incompatibilists can accept? They must be forms of freedom and responsibility that are consistent with fate and luck and do not entail desert in the basic sense.

Let's start with the plausible thought that different kinds of moral evaluations have different inaptness conditions. Moreover, one may plausibly think that some of these inaptness conditions generalize if (for instance) determinism is true. Resentment and indignation are emotions tied to practices of sanction and blame, which are in turn tied to notions of control which skeptics like Pereboom (2014: 128-129) have argued preclude determinism. But there are, of course, kinds of emotions for which it is highly implausible to imagine there being universal inaptness. Consider admiration. Plausibly, even if no one had any control over the development of his or her virtue, we still might see the value of said virtue and admire it.

Indeed, when we admire someone for their courage, we admire them for the *manifestation* of their courage. We can admire this manifestation, it seems, regardless of whether or not it was up to the agent that she so manifested courage. Now, there is some control or agency on the part of a courageous agent. She brings it about that she does the courageous thing by retaining motivation to act in spite of fear, say. But whatever control is implicated in her bringing this about is compatible with it not being up to her that she brings it about.

There is no obvious reason why a hard incompatibilist cannot accept such a story. Indeed, Pereboom (2014, *inter alia*) labors mightily to show just how much of the moral and deliberative aspects

of human life are compatible with determinism. Indeed, Pereboom (2017) has articulated a conception of moral responsibility without basic desert in terms of communicated moral protest, and a conception of wrongness without moral obligation or demand. It is, in effect, an axiological, consequentialist replacement for the morality system. On this conception, a skeptic like Pereboom can admit that there is a set of conditions on action, which do not preclude fate and luck, such that when met, agents can be held morally responsible (in this axiological-only sense) for those actions. He just denies the kind of free will—the kind of control—required for basic desert.

If a hard incompatibilist can accept the forms of freedom and responsibility which are also endorsed by the critical compatibilist, then the orthodox disputants of the free will debate should wonder whether critical compatibilism is really a form of compatibilism after all. Sure, critical compatibilists and hard incompatibilists have different metaphysical attitudes. However, on one way of construing the free will debate, one that I find plausible, we are concerned with the features of moral agency which constitute morally *accountable* agency, the sort of agency that gets you “on the hook” for your actions (Watson 1996). This kind of agency involves desert, obligation, and the appropriateness of hard feelings like resentment and indignation. If the critical compatibilist denies that we lack the kind of control needed for this kind of responsibility, then in what sense is the view compatibilist rather than incompatibilist in terms of the pertinent question? The critical compatibilist needs to say more about what distinguishes the two views.

4. Saying More about Moral Responsibility

I'll sketch two options for how a critical compatibilist might say more about moral responsibility and desert in order to show that their view is really a form of a compatibilism. Both strike me as appealing, but each would need to be further developed in light of objections. Each, however, seems to diminish the warrant of free will pessimism.

There was an obvious difference in the foregoing discussion that may help distinguish critical compatibilism from hard incompatibilism. Russell's way of describing the ethical reactive attitudes involves thick concepts. Pereboom's alternative moral responsibility system is cast in terms of care, harm, and axiology. I suspect many would say that this system involves rather thin conceptual resources. In fact, I think anyone who sympathizes with P.F. Strawson's (1962) attack on the utilitarian conception of (what I take to be) accountability responsibility—being on the “hook” for one's actions—should be suspicious of this axiology-only approach. Likewise, anyone who sympathizes with Williams should be wary too. He calls utilitarianism a marginal form of the morality system, insofar as its “obligations” are inescapable (1985: 178). So, there is a serious and obvious difference between Pereboom-style hard incompatibilism and Russell's critical compatibilism. The example of evaluation discussed above—calling someone courageous—is a far cry from the free will skeptic's standard fare.

It is not obvious, though, that the thickness or thinness of our evaluative responsibility terms settles the matter of what form of desert is at play in our responsibility practices. Skeptics like Pereboom understand accountability in terms of basic desert. Perhaps an argument could be given that accountability needn't require basic desert. Perhaps a richer, non-consequentialist picture of moral responsibility of the sort Russell gestures at could support a non-basic notion of desert and so a compatibilist-friendly notion accountability. Perhaps it could license resentment and indignation in token instances, *pace* the incompatibilist's concerns about these emotions. If so, then perhaps critical compatibilism is rightly identified as a compatibilist view.

There is conceptual space for such a view. R. Jay Wallace (1994), for instance, grounds the appropriateness of resentment and indignation, and so accountability for actions, in considerations of fairness. Recently, some compatibilists have suggested that the traditional free will debate—the debate about whether or not we have the control needed to be accountable for our actions if determinism is

true—can be recast in terms of non-basic desert by way of contractualist resources (e.g., Lenman 2006 and McKenna 2019). Pereboom (2019), disagrees. Indeed, Pereboom has worried that if we allow views to be called “compatibilist” just because they maintain that some non-basic version of desert for actions is compatible with determinism, then “virtually everyone in the debate stands to be a compatibilist”, rendering the debate insubstantial (2014: 2). How Russell fits into this debate is not yet clear.

Perhaps instead Russell will want to say instead that accountability responsibility is just a product of the problematically idealized morality system and that it has basic desert notions baked into it. Thus, we should abandon it in favor of a different conception of responsibility. Gary Watson (1996) famously suggested that there are two kinds of moral responsibility. Accountability, being on the hook for one’s actions, and attributability, the appropriateness of attributing a character trait to someone, i.e., aretaic evaluation. There is even a third option: an answerability conception of responsibility, where responsibility is understood in terms of what things—actions and attitudes—a person can “answer” for with reasons.⁴

Both of these alternative conceptions are, it seems, compatibilist-friendly on their own terms, perhaps in part because they needn’t involve basic desert claims about actions. Indeed, since attributability responsibility is a matter of properly attributing some quality to an agent, it seems especially suitable for Russell’s purposes. It sounds like Russell’s discussion of ethical reactive attitudes as involving the perception of an agent in a certain ethical light (2017: 109).⁵ Beyond this, Russell argues that “true” desert in the artistic and athletic realms does not require libertarian free will (2017: 141-145). These domains seem highly amenable to an attributability treatment—the attribution of manifest skill in the work of the artist and the performance of the athlete. As the discussion of courage and control above shows, there is a strong *prima facie* case to think we can appropriately attribute courage to some agent even when determinism is true. Perhaps this is just a way of holding that agent responsible in some sense distinct from putting her “on the hook” for what she has done. This way of proceeding gives weight to the thought that views like Pereboom’s hard incompatibilism are actually very modest forms of compatibilism.

Abandoning accountability responsibility is not without cost. Again, plausibly, when we theorize about moral responsibility in the context of the free will debate, we are aiming for a theory of morally *accountable* agency. So, it is not yet clear that abandoning an accountability framework constitutes a form of compatibilism in the pertinent sense. Moreover, the desert claims of these alternatives to accountability responsibility are non-basic. Again, if this is sufficient for a compatibilism about a substantial kind of free will and moral responsibility, then we risk adopting a view on which virtually everyone will count as a compatibilist.

The success of either of these options will paint the critical compatibilist as comparatively more *optimistic* than they first appear. How so? Well, either (1) the kind of control at issue in the traditional free will debate does not entail a commitment to basic desert, in which case the metaphysical aspirations of the traditional free will debate are metaphysically modest, or (2) substantive notions of control and responsibility lie outside the contours of the traditional debate, in which case the metaphysical aspirations of the free will debate cast in terms of accountability seem moot. Either option seems to significantly dampen the need to accept free will pessimism, the view either our predicament with respect to fate and luck is or “should be significant to us” (2017: 259).

⁴ Scanlon (2008) and Smith (2012) offer accounts of the latter. See Shoemaker (2015) for a pluralistic account of responsibility.

⁵ Compare with how Williams (1993: 78) describes the basic experience of shame: “The basic experience connected with shame is that of being seen, inappropriately, by the wrong people, in the wrong condition.”

5. Pessimism About Free Agency

Could a commitment to free will pessimism persist in light of a successful naturalistic and compatibilist account of freedom and responsibility, one that distinguishes critical compatibilism from incompatibilist versions of local free will skepticism? Perhaps. I worry, though, that by plumping for free will pessimism will run the risk of plumping for incompatibilism too.

Let's start with descriptive side of the question at hand. Why would someone embedded in the sort of naturalistic system of moral responsibility critical compatibilism recommends be pessimistic about the kinds of free will ruled out by fate and luck? Relative to the standard of the morality system, I can understand pessimism. Internal to the morality system, one might *expect* oneself to have wide-reaching powers of control immune to the whims of chance and fate. But why would someone brought up in an ethical system which accepts fate and luck be moved by worries about the metaphysical limits of her powers? Relative to what such an ethical system would say about *her*, what could ground pessimism about herself or her situation?⁶

It is natural here to think about ancient Greek portrayals of tragic heroes, persons who accept liability—shame, and perhaps even guilt and the indignation of others towards them—even though they *regret* things that are out of their direct control, like their character, or their unintentional actions.⁷ Perhaps this regret is an expression of pessimism about our metaphysical predicament. And perhaps it shows that pessimism can coexist with robust notions of responsibility and control.

Indeed, Bernard Williams devotes an entire chapter of his *Shame and Necessity* (1993) on to “the Greeks' recognition that in their world one's whole life, all the ways in which one was treated, one's ethical identity, might depend on a chance.” (106). I will use his discussion as a guide. Williams puts the point poignantly in terms of the tragic life of Oedipus: “The whole of *Oedipus Tyrannus*, that dreadful machine, moves to the discovery of just one thing, that *he did it*” (1993: 69). We don't dread Oedipus's unintentional actions because we also believe in magic blood-guilt, but rather because we too understand “that in the story of one's life there's an authority exercised by what one has done, and not merely by what one has done intentionally” (1993: 69). These unintentional horrors *call out for response* on the part of our hero. Attitudes like shame and regret, for instance. Of course, Oedipus didn't have to blind himself. And yet doing nothing seems unfitting, a failure to acknowledge that *he did it*.

There is, then, conceptual room for a notion of responsibility in a shame culture like the Greeks. Perhaps, as Williams (1993: 91-95) suggests, accountability-like notions of guilt and obligation could be nested within a larger agent-centered conception of the ethical life. Nevertheless, it is a further question whether or not the regret seen in Greek tragedy is rooted in pessimism. Regret that we are just play-things of the Gods, or bound by the thread of fate, is not the same thing as being “disconcerted” about the sources of the self (Russell 2017: 238).

By my lights, it is the startling *composure*, not the lack of it, which characterizes the Greek heroes. It is surprising to moderns (well, at least to this modern!) that Ajax does not lose his composure after Athena tricks him into embarrassingly slaughtering animals. Sure, he laments. But he also quickly resolves himself to suicide. His *ethos* demands that he cannot live in such shame (Williams 1993: 72). This does not strike me as the acknowledgement of a predicament, but rather, a response to the lived reality of his finite condition.

Beyond this, there is the altogether further matter of what *our* metaphysical expectations would be, if *we* were raised up in a successful, naturalist conception of responsibility. We can look to the

⁶ An anecdote: a virtue ethicist once told me not to worry about the problem of free will and determinism since Aristotle didn't worry about it.

⁷ Thanks to an anonymous reviewer for this helpful suggestion.

Greeks for inspiration in the constructive project of reimagining our ethical worldview. We cannot go back, though. Maybe my intuitions are wrong. Nevertheless, I suspect that if we had the sort of naturalist conception of responsibility Russell recommends *all along*, we would not be so upset about our limits.

It might be the case that given our historical moment, we do in fact have metaphysical aspirations which cannot be met. This would naturally lead to pessimism. And moreover, it might be that someone who recognizes that they do not settle the matter of *who they are* in fundamental ethical respects could also feel guilty for their actions, and so perhaps hold themselves accountable. Whether or not pessimism is *warranted* in spite of this recognition is a matter of whether or not aspirations to settle these facts about oneself are reasonable.

Perhaps, as Russell puts it, free will pessimism shows “appropriate sensitivity to features of human life that should be recognized as troubling or difficult” (2017: 259). We have now shifted to the normative question: *should* someone guided by a luck-and-fate infused system of ethics be troubled by the limits of control and the contingency of her life? Russell (2017: 197) has a helpful discussion of a passage in Spinoza that gets to the heart of this normative question. Spinoza gives voice to a deep resistance to naturalism. We conceive of ourselves as having sovereignty, as ordering nature rather than being ordered by it (2017:197-198). Since we see ourselves in this way, we see ourselves as having dignity. And it does seem like we should worry about whether or not we have dignity.

Now, though, one worries that pessimism starts to sound like hard incompatibilism. If self-shaping is the standard for when we can rightly see ourselves as dignified, then perhaps a good portion of our responsibility practices, even when we adopt a wider set of ethical notions than those permitted by the narrow confines of the morality system, will be inapt if determinism is true. Strawson (1962: 22) says the central commonplace on which he insists is that we very much care about the attitudes that others have towards us and have a natural set of reactions to beliefs about those attitudes. This commonplace can be captured (if oversimplified) in terms of language about “human self-respect and its connection with the individual’s dignity” (1962: 22). Critical compatibilism, then, starts to look once again like hard incompatibilism, if it denies us this dignity. It was precisely a dignity-less, axiological utilitarianism which Strawson was so concerned with, and that we who admire Strawson might see a shadow of in Pereboom’s revisionary project.

Thus, we have traversed both horns of the dilemma. Russell needs to distinguish his view from hard incompatibilism. Doing so requires a more fully worked out view of the kind of moral responsibility system that critical compatibilism endorses. But on two plausible ways of doing so, the grounds of adopting free will pessimism seem significantly diminished. On the other hand, if free will pessimism is warranted in spite of this, then it starts to look like hard incompatibilism is too.

6. Realism as a Metaphysical Attitude

Here is one way to respond to this dilemma. Perhaps the critical compatibilist needn’t have a uniformly pessimistic or optimistic response to the problem at hand. Maybe there are some reasons to be hopeful about abandoning the morality system and some reasons to be disconcerted. In fact, Russell suggests this (2017: 257). I’m not sure this resolves the dilemma problem, however. The dilemma problem indicates that our reasons to be optimistic (at least in part) preclude our reasons to be pessimistic, and vice versa. This suggests that a uniform response is preferable, if not altogether needed.

So, how else could a critical compatibilist respond to this dilemma problem? Russell himself says that the pessimistic effect of abandoning the morality system “is likely to recede or dissipate over time as *we* (moderns, Westerners) withdraw from the morality system” (2017: 258). It seems, then, that a critical compatibilist’s commitment to compatibilist naturalism about responsibility could be deeper or more basic than their commitment to free will pessimism.

I suspect that there is something *unrealistic* about the Spinozistic set up which licenses any kind of pessimistic response. If so, then perhaps pessimism isn't the warranted attitude in light of a critical take on the classic free will problem. And this might ease the tension in critical compatibilism.

Compare with Williams' conclusion in "How Free Does the Will Need to Be?", that "if we bring our ethical ideas nearer to reality, then assuredly we shall find that they are consistent with naturalistic explanations of our choices and actions" (1995: 19). Indeed, Williams notes in *Shame and Necessity* that Greek ethical thought had a realism to it insofar as it involved a broad conception of responsibility (1993: 95). We seem to have lost this sense of realism.

Whence our unrealistic expectations, then? One might say something here, picking up on Bennett's (1980) critique of Strawson, about competing Christian-Kantian and Spinozist conceptions of the self and the origins of the free will problem. Indeed, Strawson himself (1992) characterizes the pertinent conflict as between our sense of ourselves as free against the Spinozistic thought that this sense is an illusion. Russell too (2017:104) characterizes the aspirations of the morality system as being a product of our modern Western Christianized culture. That our aspirations have this origin is not an uncommon thought. Russell (2017: 256) cites Strawson (1994), Wallace (1994), and Dennett (1984) as each offering remarks about the Judeo-Christian sources of our metaphysical aspirations.

But the label "Judeo-Christian" (or "Judeo-Christian-Kantian") can apply to a diverse body of views. (Perhaps so diverse as to make these unhelpful categories altogether). And I suspect our unrealistic expectations are the product of more recent historical, philosophical, and theological developments. For instance: should someone raised up in a Thomist worldview be pessimistic about the limits of her agency? Internal to that view, limited human agency is the norm. Human agency must be enhanced by God's grace. Our achieving true virtue is not up to us, at least not fully (Stump 2012). Yet the view seems to endorse a modest form of libertarianism about free will: "a decision is free only if it is not the outcome of a causal chain that originates outside of the agent" (Stump 2006: 302). So, the view endorses a form of control that is more than middle-distance yet is crucially circumscribed at the horizon. Internal to the ontology of the view, it would be unrealistic (arrogant, perhaps, by the views own standards) to expect of human agency that it achieves horizon level self-making. Pessimism doesn't seem warranted. On the other hand, optimism seems unrealistic too. As Williams (1985: 195) notes, orthodox Christianity offers no "calculable road from merit to salvation". In this respect it is both unlike and in tension with the aspirations the morality system. (Of course, this agential limitation is squarely supernatural, and so different than our predicament with respect to physical determinism).

Is there a way we could be realistic naturalists about our current metaphysical situation? By my lights, *Strawsonian* compatibilism has the best shot at doing so. On some recent interpretations of Strawson's responsibility naturalism (e.g., Watson 2014, Beglin 2018), our reactive attitudes and responsibility practices are expressions of our basic human concerns. One could argue that, given the content of these basic concerns, we don't really have need for great metaphysical expectations. As Strawson (1985: 41) notes, "there can only be a *lack* where there is a *need*." Or perhaps a realistic attitude could be nurtured by way of a "therapeutic" philosophical project. We could aim to show how our apparent metaphysical need comes from an incoherent demand. Perhaps there is enough and as good left over when we embrace more modest senses of "ultimate" agency (cf. McKenna 2008).

I suggest this realistic attitude with some trepidation. This attitude does not obviously accommodate notions of sovereignty and dignity. These notions are really worth fretting over. (What would a modest conception of dignity look like, anyway?) I nevertheless submit that in recognizing the limits of our agency, we do not have to choose between optimism and pessimism. Within a compatibilist-friendly system of moral responsibility underpinned by ethical reactive attitudes and thick moral notions, we might come to a realistic sense of self in the world.

7. Conclusion

Although I have offered a dilemma for Russell's critical compatibilism, it is by no means a decisive one. I have only gestured at one line of response in the foregoing. And I have offered this dilemma in the spirit of admiration and with a desire to prompt discussion. The free will debate is notoriously deadlocked, and Russell's work thoughtfully opens up new areas of philosophical exploration. But more importantly, Russell's philosophical facing-up to our limits as natural beings embodies intellectual humility and humanity in equal measure. As a young scholar, I am especially grateful to have encountered his work.

References:

- Beglin, D. 2018. "Responsibility, Libertarians, and the "Facts as We Know Them": A Concern-Based Construal of Strawson's Reversal *." *Ethics* 128.3: 612-25.
- Bennett, J. 1980. "Accountability." In Zak van Straaten, ed. 1980. *Philosophical Subjects: Essays Presented to P.F. Strawson*. Oxford: Clarendon: 14-47.
- Dennett, D. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge, Mass: MIT Press.
- Feinberg, J. 1970. *Doing & Deserving: Essays in the Theory of Responsibility*. Princeton, N.J.: Princeton University Press.
- Fischer, J.M. and M. Ravizza. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press
- Lenman, J. 2006. "Compatibilism and Contractualism: The Possibility of Moral Responsibility *". *Ethics*, 117(1): 7-31.
- McKenna, M. 2019. "The Free Will Debate and Basic Desert." *The Journal of Ethics*, 23(3): 241-255.
2008. "Ultimacy & Sweet Jane." In Nick Trakakis and Daniel Cohen, eds. *Essays on Free Will and Moral Responsibility*. Cambridge Scholars Publishing: 186-208.
- _____. 1998. "The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism." *Journal of Ethics* Vol. 2, No. 2: 123-42.
- Pereboom, D. 2019. "What Makes the Free Will Debate Substantive?" *The Journal of Ethics*, 23(3): 257-264.
- _____. 2017. "Responsibility, Regret, and Protest," *Oxford Studies in Agency and Responsibility* 4, David Shoemaker, ed., Oxford, UK: Oxford University Press, 2017: 121-40.
- _____. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Russell, P. 2018. "Bernard Williams: Ethics from a Human Point of View". *Times Literary Supplement*. <https://www.the-tls.co.uk/articles/public/bernard-williams-ethics-human-point-view/>
- _____. 2017. *The Limits of Free Will*. Oxford, UK: Oxford University Press.
- _____. 1995. *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility*. New York, NY: Oxford University Press.
- Scanlon, T.M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, Mass: Belknap Harvard Press.

- Shoemaker, D. 2015. *Responsibility from the Margins*. Oxford, UK: Oxford University Press.
- Smith, A. 2012. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122: 575-589.
- Strawson, G. 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75:5-24.
- Strawson, P. F. 1992. "Freedom and Necessity." Chapter 10, *Analysis and Metaphysics: An Introduction to Philosophy*. Oxford, UK: Oxford University Press.
- _____. 1985. *Skepticism and Naturalism: Some Varieties*. New York: Columbia University Press.
- _____. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 187-211.
- Stump, E. 2012. The Non-Aristotelian Character of Aquinas's Ethics: Aquinas on the Passions. In *Faith, Rationality, and the Passions* (pp. 91-106). John Wiley and Sons.
- _____. 2003. New York, NY: *Aquinas*. Routledge
- Wallace, R.J. 1994. *Responsibility and The Moral Sentiments*, Harvard University Press, Cambridge, Mass.
- Watson, G. 2014. "Peter Strawson on Responsibility and Sociality." In David Shoemaker and Neal A. Tognazzini, eds., 2014, *Oxford Studies in Agency and Responsibility, vol. 2: 'Freedom and Resentment' at 50*. Oxford: Oxford University Press: 15-32.
- _____. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 227-48.
- _____. 1987. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In Ferdinand Schoeman, ed. 1987. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press: 256-86.
- _____. 1975. "Free Agency". *Journal of Philosophy* 72: 205-20.
- Wiggins, D. 1973. "Towards a Reasonable Libertarianism." In Ted Honderich, ed., 1973. *Essays on Freedom and Action*. London: Routledge & Kegan Paul: 31-62
- Williams, B. 1995. "How Free Does the Will Need to Be?" Chapter 1, *Making Sense of Humanity and Other Philosophical Papers*. Cambridge, UK: Cambridge University Press.
- _____. 1993. *Shame and Necessity*. Berkeley, Calif.: University of California Press.
- _____. 1985. *Ethics and the Limits of Philosophy*. Cambridge, Mass.: Harvard University Press.