JUSTICE AND THE GREY BOX OF RESPONSIBILITY[*]

Abstract: Even where an act appears to be responsible, and satisfies all the conditions for responsibility laid down by society, the response to it may be unjust where that appearance is false, and where those conditions are insufficient. This paper argues that those who want to place considerations of responsibility at the centre of distributive and criminal justice ought to take this concern seriously. The common strategy of relying on what Susan Hurley describes as a 'black box of responsibility' has the advantage of not taking responsibility considerations to be irrelevant merely because some specific account of responsibility is mistaken. It can, furthermore, cope perfectly well with an absence of responsibility, even of the global sort implied by hard determinism and other strongly sceptical accounts. Problems for the black box view come in where responsibility is present, but in a form that is curtailed in one significant regard or another. The trick, then, is to open the box of responsibility just enough that its contents can be the basis for judgments of justice. I identify three 'moderately sceptical' forms of compatibilism that cannot ground judgments of justice, and are therefore expunged by the strongest 'grey box' view.

I. Introduction

Attributions of responsibility are central to practices of punishing in the criminal law. During the past two decades several political theorists have suggested that responsibility should also be a key consideration in the allocation of benefits and burdens among citizens. Both types of practice can be construed as being based on

[*] This is a post-peer-review, pre-copyedited version of an article published in *Theoria*. The definitive publisher-authenticated version, 'Justice and the Grey Box of Responsibility', *Theoria*, 57 (2010), 86-112, is available online at: http://www.ingentaconnect.com/content/berghahn/theoria/2010/00000057/00000124/art00004.

what I will call the *responsibilitarian principle,* which states that those who are responsible for bringing about some thing ought to be rewarded or punished according to the (moral or prudential) goodness or badness of that thing.[1] The application of this principle to a matter of criminal justice is commonly known as *retributivism.* The relative of the general principle most discussed in political philosophy goes by the name *luck egalitarianism.*

The contrast between retributivism and luck egalitarianism extends beyond the fields in which they are applied. Those inclined to invoke the simple left-right political spectrum might note that, while retributivism is nowadays usually taken to be a conservative position on criminal justice, luck egalitarianism is, at least according to its advocates, a view firmly located on the liberal left. Luck egalitarianism's critics might view that as a useful weapon in their arsenal. It is necessary, then, to observe that the egalitarianism of luck egalitarianism is at least a little more than a misnomer. For luck egalitarians combine the responsibilitarian principle with the *equality default principle,* which holds that all departures from equality must be justified on responsibility grounds.[2] Retributivism appears to encompass a counterpart principle stating that in the absence of bad responsible acts, there should be no punishment.[3]

Retributivism and luck egalitarianism have been the subjects of much criticism, and there are similarities in the kinds of criticism the two views have faced. The most common objection to retributivism is that offenders' responsibility for their crimes does not by itself say anything about how the law ought to respond.[4] The claim

[1] This combines elements of H. L. A. Hart's (1968, 231) principles of responsibility and proportionality; the third principle he lists, that of just requital, is specific to retributivism.

[2] Hurley 2003, 153-4, 172.

[3] Hart 1968, 9. Throughout this article 'acts' should be taken to include omissions.

[4] Wootton 1963.

is that the retributivist simply asserts that punitive action is the appropriate response to crime, when other responses – therapeutic intervention, for instance – are more appropriate, especially given the modern orientation towards preventing crime and away from punishing evil. Similar objections have been raised to luck egalitarianism. Why exactly does someone's responsibility for an act that harms themselves or others make it legitimate for them to be disadvantaged relative to those whose acts have prudentially or morally good effects?[5] The shift here reflects the differing scope of the two views under consideration: whereas retributivism is concerned only with punishing wrongs, luck egalitarianism is also concerned with rewarding action that has good effects. But in both cases it is objected that differential exercises of responsibility cannot justify differences in individuals' benefits and burdens. We might call this the *arbitrariness of reaction objection.*

Some of the critics of luck egalitarianism and retributivism also object to the notion that it is possible in theory and/or practice to reward and punish on the basis of any deep sense of responsibility. While it is of course entirely possible to use social institutions to ensure that certain acts are considered to be responsible ones, in the sense that they are *treated as* appropriate bases of reward and punishment, it is quite another to assert that such political and legal arrangements are sensitive to what people are actually responsible for in a metaphysically and/or morally significant way. Even if an act appears to be responsible, and satisfies all the conditions for responsibility laid down by society, the response to it may be unjust where that appearance is false, and where those conditions are insufficient. In other words,

[5] One specific form of this complaint is the 'harshness objection' (or 'abandonment objection') put forward by Elizabeth Anderson (1999) and other critics of luck egalitarianism. For discussion of this objection see Voigt 2007; Segall 2007; Knight 2009a, ch. 4, 6.

socially prescribed *accountability* is insufficient for responsibility (and the rewards and punishments that might justify). This is the *arbitrariness of attribution objection*. It is independent of, and in a way more fundamental than, the arbitrariness of reaction objection. An adherent might accept that there is nothing objectionable about reacting, even punitively, to genuinely responsible acts, but deny that such acts exist in general or in the cases under consideration. The most frequently cited general threat to such genuine responsibility comes from *causal determinism* – the view that all events and states of affairs, including human action, are causally determined by earlier events – and particularly from its combination with the view that responsibility is incompatible with such determinism, which combination is called *hard determinism.* But this is far from being the only way in which the connection between responsibility and retributive or luck egalitarian applications of justice might be undermined.

In this article I argue that an appealing response to the arbitrariness of attribution objection, while successfully protecting retributivism and luck egalitarianism from the allegation of basing distributions on the mere façade of responsibility, only serves to expose responsibilitarianism to the arbitrariness of reaction objection and to conceptual confusion. I believe that retributivists and luck egalitarians alike should be concerned with a deep sense of responsibility, rather than with some more conventional notion that might be better described as accountability. But it turns out that the strategy of placing responsibility in a 'black box', and reacting to responsibility however that is best construed, results both in morally unjustified reactions and reactions that are not truly responsibilitarian.

To demonstrate this, I consider several accounts of responsibility, paying special attention in each case to responsibility's appropriateness as a fundamental ground for rewarding and punishing. There is a focus on this topic in legal

philosophy, but political philosophy has barely considered it. This is unsurprising given that luck egalitarianism has dragged responsibility into the centre of discussions about distributive justice only in the last twenty years, whereas responsibility has been central to criminal justice for centuries. My arguments will therefore have more significance in relation to luck egalitarianism – that is, in showing how it can resist the arbitrariness of attribution objection while reacting to attributions of responsibility in a morally justified and conceptually clear way. A further objective is to bring out some basic similarities in the kind of responsibility required by retributivism and luck egalitarianism, a topic that is neglected in both legal and political philosophy.

I begin by favourably comparing 'black box responsibilitarianism' – the form of responsibilitarianism that defines responsibility thinly, basing distributive and punitive responses on whichever account of responsibility appears to be most metaphysically and morally convincing – with the alternative 'white box responsibilitarianism', which defines responsibility thickly, responding only to a specific pre-defined account of responsibility. I show that the truth of hard determinist or regressive accounts of responsibility would not undermine the theoretical claims of black box responsibilitarianism, although it would severely harm the prospects of retributivism and luck egalitarianism as distinctive answers to practical problems (section II). I then address three versions of *compatibilism* – understood as the view that responsibility is compatible with determinism – that do indeed appear to undercut black box responsibilitarianism at the theoretical level in that they render its reactions either morally arbitrary or conceptually unsound. The first of these versions of compatibilism is characterized by being non-distributive and non-punitive in its implications (section III), the second is defined by its concern with forward-looking social regulation (section IV), while the third is a restricted version of the social

regulation view (section V). I respond to the problems posed by these accounts by defending 'grey box responsibilitarianism' – that is, a position that rules certain accounts of responsibility inadmissible on the ground that, even if they were correct, they could not justify punishment or reward on responsibilitarian grounds. The specific form of grey box responsibilitarianism that I defend responds only to exercises of responsibility that satisfy three conditions: they must have punitive or distributive consequences according to the account of responsibility in question, they must take the form of acts in the past, and they must reflect autonomous agency.

II. Black box responsibilitarianism

Susan Hurley has noted that, '[w]hile the luck-neutralizing [i.e. responsibilitarian] account gives responsibility a central role in distributive justice, it has not focused much analysis on responsibility itself or on how its structure or character might constrain its role within distributive justice'.[6] The strategy of writers such as Richard Arneson, G. A. Cohen, and Larry Temkin has rather been to bracket off questions concerning the correct characterization of responsibility. As Cohen notes, this 'subordinates political philosophy to metaphysical questions that may be impossible to answer'.[7] The intention is to allow discussion about distributive justice to proceed fruitfully, with the apparently interminable debates about free will, determinism, and related matters consigned to a black box marked 'responsibility'.

On my view the black box approach to responsibility is not that far from being the correct one for the responsibilitarian to take. The main consideration in its favour is obvious enough, especially when we compare such a strategy to the alternative

[6] Hurley 2003, 1.

[7] Cohen 1989, 934; see also Arneson 1989; Temkin 1993.

'white box' strategy.[8] This latter strategy is defined by its commitment to a certain very specific account of responsibility. A view endorsing this strategy might, for instance, combine responsibilitarianism with a particular mechanism-based reason view.[9] It seems clear enough that, at least given the present state of debates about responsibility, one endorsing the white box strategy is severely *overcommitted.* Should that very specific account of responsibility – the claim about the contents of the box – turn out to be mistaken, the overall view falls with it. The overall view would be stuck making arbitrary *attributions* of responsibility – attributions which are based on a particular view of responsibility which has been shown to be wrong. But the rightness or wrongness of responsibilitarianism does not turn upon the rightness or wrongness of any particular mechanism-based reason view, nor with mechanism-based reason views as a whole, nor with reason views as a whole, nor even with compatibilism. To reject retributivism and luck egalitarianism for such specific reasons would be to throw the baby out with the bath water. The black box responsibilitarian, by contrast, rejects all such potentially damaging commitments.[10]

[8] In physics, electronics, and computing black boxes are systems or devices which are considered in terms of their inputs and outputs, while white boxes (sometimes also known as 'glass boxes' or 'clear boxes') are systems or devices whose inner workings are visible and of interest. Similar ideas are also sometimes used in philosophy and social science, especially in discussions of behaviourism.

[9] See, e.g., Fischer and Ravizza 1998.

[10] A referee suggests that there may be disadvantages to this lack of commitment: 'the justice claims are really meaningless or inadequate for a practical purpose (and we are in the realm of practical philosophy here) unless at least some account of responsibility is forthcoming and correct: staying agnostic about which account of responsibility is correct simply renders justice claims based upon responsibility opaque as the very condition for such justice claims being made may not be realized'. In reply I would point out that, while black box responsibilitarianism does not specify an account of

Hurley suggests that '[t]he black box of responsibility that luck-neutralizing accounts build into the heart of egalitarian justice turns out, when opened, to be something of a Pandora's box'.[11] On a superficial reading this might appear to be supportive of the black box responsibilitarian's position: after all, things would have been better had Pandora not opened her box. But Hurley's point is rather that the contents of the box, whether opened or not, do not serve the responsibilitarian as well as has been supposed. Responsibility has implications for political and (by extension) legal philosophy that are quite incompatible with luck egalitarianism and (by extension) retributivism.

The 'Pandora's box' claim is not much developed in Hurley's book.[12] We might nevertheless ask what these implications so problematic for responsibilitarians might be. The following possible claims present themselves:

> (1) Libertarianism is true – that is, free will and responsibility are possible and incompatible with causal determinism – but since the conditions for libertarian responsibility (e.g. the presence of 'contra-causal freedom'[13]) are rarely satisfied, responsible acts are very unusual.

---

responsibility in advance, that does not mean that it lacks an account of responsibility at any given time. It simply endorses whichever account of responsibility is presently best justified. At any given moment it therefore has just as much content as white box responsibilitarianism, and can give just as much distributive or retributive guidance in practice. (Elsewhere I have suggested that a 'responsibility committee' of professional philosophers might provide their best guess about responsibility to a luck egalitarian government's distributive arm; see Knight 2006.)

[11] Hurley 2003, 3.

[12] See Watson 2007.

[13] Campbell 1951.

(2) Compatibilism is true – in the sense that free will and responsibility are possible and compatible with causal determinism – but since the conditions for compatibilist responsibility (e.g. the presence of strong reasons responsiveness) are rarely satisfied, responsible acts are very unusual.

(3) Libertarianism is false, so responsibility is impossible.

(4) Hard determinism is true – that is, causal determinism is true and incompatible with free will and responsibility – so responsibility is impossible.

(5) Regressive control – control not only of one's act, but also of the causes of one's act – is required for responsibility, and since regressive control is impossible, responsibility is impossible.

The non-empirical components of (1) and (2) do not appear to be at odds with black box responsibilitarianism. Indeed, if we take the metaphysical and moral claims of (1) and (2) in isolation from the empirical claims, we have support for that position. Do the empirical claims undermine it? Surely not. The black box responsibilitarian claims only that where responsibility is present – where the output of the black box has a certain character – a certain institutional response is appropriate. It says nothing about how often responsibility actually is present – nothing, that is, about the frequency of that certain kind of output.

That is not to say that implications (1) and (2) would be irrelevant to practical problems, were they to hold. They might be used to ground a *specific* form of the

arbitrariness of attribution objection. It might be maintained that, while it is possible for someone to be responsible for a criminal act, the kind of cases in which persons are routinely held responsible for criminal acts are not actually cases involving responsible action. At its most specific, this sort of objection might consist simply in pointing out that the real world facts about responsibility are not as the retributivist claims them to be. A slightly more generalized form of the specific objection would deploy a genuinely responsibility-sensitive form of retributivism against a less demanding, more conventional form of retributivism. It could be pointed out, not that what the opposing retributivist says is present is not present, but rather that this protagonist is concerned with the presence or absence of the wrong kind of thing – with the mere appearance of responsibility. The present point is that both kinds of objection would find traction where claims (1) and (2) – or weaker claims, suggesting, for example, that satisfaction of the conditions for responsibility is less common than social institutions appear to suppose – are true, since their truth might undermine justifications of retributivist practices. Of course, parallel points could be made about luck egalitarian practices, were they present.

Only a *general* arbitrariness of attribution objection can undermine black box responsibilitarianism. (3) appears to be sufficiently general. Although such a stark and contentious claim is not often articulated nowadays, it has sometimes been suggested that a refutation of libertarianism would be sufficient refutation of responsibilitarianism, on the ground that libertarianism is required to maintain the key responsibilitarian distinction between those things for which we are responsible and those things for which we are not responsible.[14] But while maintaining that

[14] The legendary American trial lawyer Clarence Darrow often criticized retributive laws from a determinist perspective, asserting that 'every scientific man knows … that the whole current conception

libertarianism is false is a perfectly consistent position, it does not at all imply that free will and (hence) responsibility are impossible, since free will and responsibility may be compatible with determinism, as compatibilism holds.[15] It is more generous, then, to read such critics of responsibilitarianism as really endorsing (4) or (5).[16]

The first of these views is the traditional one for sceptics about responsibility to take. If hard determinism is true, it follows that reward and punishment (and, for that matter, praise and blame) cannot be based on responsibility. By contrast, (5) is a form of scepticism about free will and responsibility that has particularly come to prominence in the wake of 'Frankfurt examples' and other apparent demonstrations of the dissociability of responsibility and the ability to do otherwise.[17] Since the ability to do otherwise is often thought to be ruled out by causal determinism, this dissociability appears to undermine the threat that determinism poses to responsibility. But it does nothing to undermine the thought that control of the ultimate causes of one's acts is required for responsibility. Consequently, some philosophers have suggested that the gravest threat to responsibility is now the

---

of the individual and his responsibility is a gross error' (Darrow 2004, 20). In the mid-twentieth century similarly sweeping deterministic arguments against legal responsibility were advanced by writers of a scientific background (Slater 1954; MacDonald 1955), and the law's rejection of determinism continues to trouble some (Hill 1988; Coffey 1993). For nuanced criticism of luck egalitarianism along broadly the same lines see Scheffler 2003; 2005.

[15] Causal indeterminism is often thought to be as much of a barrier to libertarian free will as determinism. No violence would be done to the argument of the paper were the definitions of libertarianism and compatibilism revised to refer to free will and responsibility's (in)compatibility with *both* determinism and indeterminism.

[16] Or, indeed, (6). See the next section.

[17] Frankfurt 1969.

regression requirement, for responsibility is impossible where its necessary conditions include the agent being responsible for causes temporally located prior to her birth.[18]

While proof of (4), (5), or other strongly sceptical claims about responsibility[19] would have wide reaching *practical* implications for black box responsibilitarianism, it would not undermine the position, since that position holds only that *where* responsibility is present, appropriate reward and punishment ought to follow. Knowing that responsibility is never present would tell us how to apply retributivism or luck egalitarianism – specifically, by telling us not to punish or reward apparently responsible acts. Such knowledge might, then, give us grounds for criticizing existing practices done in the name of responsibility. But it would do nothing to undermine responsibilitarianism as a theory, since there would be no dissonance between this knowledge and the position. 'Reward or punish those who satisfy W', where W denotes whatever substantive conditions for responsibility are specified in the box, is just as plausible as a moral, legal, and political principle where no prima facie responsible acts are in fact responsible as it is where many or all such acts are responsible.

It might be thought that it would hardly be worth defending black box responsibilitarianism where strongly sceptical claims about responsibility are proven beyond a shadow of a doubt, as it has the same prescriptions in such circumstances as views that do not refer to responsibility. The point, however, is that black box responsibilitarianism is no less plausible than responsibility-insensitive views even

[18] For discussion see Hurley 2003, pt. I.

[19] These include 'hard incompatibilism' claims to the effect that responsibility is impossible if determinism is true, and (given the available evidence) impossible if indeterminism is true; see G. Strawson 1986; Pereboom 2001.

under such extreme conditions, and if we return to more likely circumstances, where strongly sceptical claims are far from proven, it is at a significant advantage relative to responsibility-insensitive positions, which can take no account of the moral significance of responsibility. Black box responsibilitarianism does not make attributions of responsibility arbitrarily, as white box responsibilitarianism does, and it does not arbitrarily (that is, without reference to the metaphysics and morals of responsibility) refuse to make attributions of responsibility, as responsibility-insensitive views appear to.

III. Opening the box, part 1: non-distributive, non-punitive compatibilism

There are further possibilities that show that the black box responsibilitarian's lack of commitment is not always an advantage. The first is represented by this further claim:

> (6) Compatibilism is true, and responsibility only justifies non-distributive, non-punitive reactive attitudes of praise and blame.

Some writers have argued that a luck egalitarianism grounded in compatibilism might be mistaken not because compatibilism is wrong, but rather because responsibility grounded in compatibilism stops short of justifying inequalities in holdings.[20] This limitation would be equally problematic for retributivism. If one's responsibility for some criminal act is only appropriate grounds for censure, rather than, say, incarceration or financial penalty, retributivism would have to undergo a thorough overhaul or be abandoned altogether.

[20] Fleurbaey 1995; cf. Scanlon 1998, ch. 6.

One might try to maintain a form of retributivism by replacing the usual punishments with public humiliation and other more subtle tools. This might in part be achieved by bringing existing collateral consequences of criminal convictions within the ambit of sentencing (an objective already being pursued in some jurisdictions for quite different reasons). Similarly, one might try to maintain distinctively luck egalitarian policies by using praise and blame as tools. This would be most promising as a strategy for welfarist luck egalitarians. It is possible, then, that the truth of (6)-type claims would not be fatal for black box responsibilitarian accounts. Even so, it is clear that it would have a significant impact on such accounts. In what follows in this section I assume that responsibilitarianism is not made non-punitive and non-distributive, as I am sure that most retributivists and luck egalitarians would not want to see their positions so severely restricted.

The most pressing point for the present argument is that the possibility of claim (6) shows that, just as white box responsibilitarianism is overcommitted, the black box view is *undercommitted.* 'Reward or punish those who satisfy W', where W denotes whatever substantive conditions for responsibility are specified in the black box, is an insufficiently discriminating formula, since responsibility may sometimes be present, but reward or punishment, at least by the usual means, quite out of order. If an account is simply concerned with the presence or absence of responsibility in particular individual acts, with no regard to how exactly those acts are responsible, it is open to the danger that its distributive or punitive response is inappropriate for *that kind* of responsibility. In other words, the attribution of responsibility may be sound, but the reaction arbitrary from a moral point of view.

The relationship between this question of *appropriateness* and the question of *proportionality* is an interesting one. Any form of responsibilitarianism has to grapple

with the question of what a proportionate response to a given responsible act is. It may provide a metric by which acts can be measured, probably taking into account (a) the extent of individuals' responsibility for those acts and (b) the extent of the good and bad things the acts bring into being. This metric will then be paired with a list indicating which benefits and burdens are appropriate for each act, with the benefit or burden being greater, the greater the goodness or badness of the consequences of the responsible act. But the response that follows from such an approach to proportionality will fail to take into account the prohibition on reward and punishment implied by (6). (6) suggests that rewards and punishments of certain familiar sorts may be inappropriate responses to responsible acts even where reward or punishment have been designed to be in proportion with the goodness or badness of the acts' consequences. This is because (6) says that praise and blame are the only appropriate responses however good or bad a responsible act's consequences may be.

Non-distributive, non-punitive compatibilism need not be identified as such by its proponents in order for responsibilitarian responses to it to be inappropriate. Consider, for instance, the Strawsonian view, according to which reactive attitudes and practices are themselves constitutive of responsibility.[21] While it may be true, as Peter Strawson holds, that we cannot (rationally) divest ourselves of reactive attitudes no matter which theoretical considerations are brought to bear, we can (rationally) alter our reactive practices such that they roughly comport with theoretical considerations, at least within a certain broad range of such considerations. In other words, the Strawsonian view of responsibility may be correct as regards reactive attitudes, but at the same time be an inappropriate basis for significant reward or penalty where theoretical considerations (both metaphysical and moral) militate

[21] P. Strawson 1974.

against our attitudes. Whether or not such a limitation is acknowledged by any particular Strawsonian – or, indeed, by Strawson himself[22] – does not affect the fact that the responsibilitarian should pay heed.[23] The point has general application: any compatibilism that establishes a form of responsibility that only has non-distributive, non-punitive implications cannot be a sound basis for luck egalitarian or retributive responses.[24]

The above discussion suggests that the responsibilitarian requires a position that says more about the relevant kind of responsibility than the black box view, but less about it than the white box view. I will call this position *grey box responsibilitarianism.* For the moment let us stipulate that this position simply adds to black box responsibilitarianism the condition that the kind of responsibility required by luck egalitarianism or retributivism is the kind that has distributive or punitive responses. It modifies the responsibilitarian formula so it reads 'reward or punish

[22] Paul Russell observes that, '[a]ccording to Strawson, our reactive attitudes and retributive practices are intimately (i.e. naturally or 'humanly') connected. In ["Freedom and Resentment"], however, Strawson has very little to say about the problem of punishment as it arises within the framework of his naturalistic account of responsibility' (Russell 1992, 297-8 n. 11; see also P. Strawson 1974, 22). Strawson is similarly unforthcoming on questions of distributive justice.

[23] The limitation may, however, be unnecessary on a variant of Strawsonian theory which 'holds that morally responsible agents are not just those who, as a matter of practice, are recipients of the reactive attitudes; rather agents are morally responsible if and only if they are *appropriate* recipients – that is, they are rationally accessible to the reactive attitudes' (Fischer and Ravizza 1993, 18, original emphasis; see also Fischer 1986; Fischer and Ravizza 1998, 5-8).

[24] The implications of this may be very wide-ranging indeed given some writers' 'worry … that even if the compatibilist is successful, the kind of responsibility that can be rescued is not the kind that is needed; it is not responsibility in the fully-fledged sense we need to be confident when the consequences for someone (perhaps ourselves) might be very grave' (Matravers 2007, 63).

those who satisfy X*'*, where X denotes whatever substantive conditions for *punitive or distributive* responsibility are specified in the black box. X-type conditions are on the whole more demanding than W-type conditions, where W just denotes whatever substantive conditions for responsibility are specified in the 'black box'. X-type conditions will include W-type conditions, but they may also include further conditions not required by one concerned only with praising and blaming. The move to grey box responsibilitarianism constitutes a partial opening (or 'greying') of the box of responsibility, since we are now concerned not only with whether there is responsibility, but also with whether it is of a certain sort – admittedly, still a very broad sort.

Although the move to X-type conditions might seem an obvious one to make, it is not one familiar from the literature on luck egalitarianism. This is most likely because luck egalitarians have not considered the possibility that (some kinds of) responsibility might only justify praise and blame, or alternatively because they have considered that possibility but rejected it as implausible. I will not in this paper directly address this latter claim.[25] That task seems unnecessary, especially since there are additional cases in which both black box responsibilitarianism and grey box responsibilitarianism, as stated thus far, appear to fall down.

IV. Opening the box, part 2: socially-regulative compatibilism

A stock compatibilist position makes something like the following claim:

[25] I argue in favour of it in Knight 2006. There I also suggest that even if (6)-type claims were true, they would not undermine luck egalitarianism, but I do so without explicitly distinguishing between black box and grey box strategies as I do here.

(7) Compatibilism is true, and it is a sufficient condition for compatibilist responsibility that good effects issue from the ascription of responsibility.

Several writers have endorsed a 'social regulation', 'economy of threats', 'effect compatibilist', or 'hard compatibilist' view of responsibility.[26] As the latter names suggests, this view endorses determinism, and denies that this is any impediment to responsibility; but it also (typically) denies that free will is possible. It is further distinguished from other compatibilist views by its hard-nosed focus on the consequences of holding persons responsible. J. J. C. Smart, for instance, maintains that, while some of our commonplace ascriptions of responsibility do presuppose libertarian freedom, many ordinary notions of responsibility can be salvaged if we drop such metaphysical pretensions and instead justify them on pragmatic grounds. If a schoolboy fails to do his homework, whether we ascribe responsibility to him or not ought to depend, as intuitively it should, on the reason for his failure (laziness or stupidity, say). But this is for no reason other than that it can indicate whether blame and punishment can make his compliance more likely in the future.

Socially-regulative compatibilism should be distinguished from views which make comparable prescriptions, but for different reasons. Barbara Wootton is well known for her view that criminal justice should be orientated towards social protection rather than punishing the wicked, and that this is best realized by extending strict criminal liability and considering mens rea only after conviction in order to

[26] The terms are from Watson 2004a, Arneson 2003, Smilansky 2000, and Wallace 1994, ch. 3, respectively. Hobbes, Mill, and Sidgwick all subscribed to a version of this view. Modern statements of it include Smart 1961; Schlick 1966; Hobart 1966; Dennett 1984.

establish the best way of correcting harmful behaviour.[27] But while the policy here is similar to that of a socially regulative compatibilist such as Smart, it is not premised on any view about the truth of causal determinism and its compatibility with practices of holding people responsible. Indeed, Wootton is explicitly uncommitted on the truth of causal determinism,[28] and rebuked H. L. A. Hart for his continued emphasis on punishment, in spite of his endorsement of the goal of social protection.[29] Some have read the core of John Rawls' position that there is no prejusticial notion of desert or responsibility as being, like Wootton's, purely moral and quite detachable from metaphysical claims.[30] Even the alternative interpretation of Rawls as basing his 'justice as fairness' on determinism does not paint him as a socially regulative compatibilist, but rather as a hard determinist.[31] I do not, then, believe that either Wootton or Rawls provide or even intend to provide a full account of how people can be responsible, so the present question – which accounts of responsibility, when combined with responsibilitarianism, result in arbitrary distributive and punitive reactions? – does not arise.

The social regulation view may appear to be promising for the grey box responsibilitarian in that it offers the prospect of reconciling responsibility practices with a naturalistic understanding of the world. In other ways, however, it seems quite inhospitable. Indeed, responsibilitarianism and socially-regulative compatibilism are

[27] Wootton 1963, 46-51.

[28] Wootton 2003, 119-28.

[29] Wootton 1963, 45-6; See Hart 1968.

[30] See Matravers 2010, sec. 3. Of course, in later work (1993), Rawls has proclaimed his revised position to be 'political not metaphysical', but the present suggestion is that even in *A Theory of Justice* Rawls does not have metaphysical commitments specifically regarding free will and determinism.

[31] Craig 1975, 72; Smilansky 2003; cf. Rawls 1999.

commonly viewed as incompatible. Moritz Schlick's especially forthright defence of his version of the view describes the conflict with retributivism quite clearly. According to the social regulation view, 'punishment is concerned only with the institution of causes, of *motives* of conduct, and this alone is its meaning'.[32] On this view, punishment is purely an 'educative measure', and '[t]he question of who is responsible is the question concerning the *correct point of application of the motive'*.[33] While Schlick notes that punishment is often viewed as 'a natural *retaliation* for a past wrong', he dismisses the idea that 'an increase in sorrow can be "made good again" by further sorrow' as 'altogether barbarous'. Richard Arneson, a prominent luck egalitarian, is more measured, though equally uncompromising on the question of compatibility:

> The [social regulation] theory of responsibility rejects the idea that individuals are ever truly virtuous or vicious. Nor can they be responsible or irresponsible for what they do in any deep sense … This being so, it is not intrinsically morally valuable that the virtuous and responsible enjoy good fortune to a greater extent than the vicious and irresponsible. … If there were virtue or truly responsible choice or true desert, it might be morally a good idea to reward it, but since these concepts have no application, in effect the virtue and responsibility and desert rankings of any person are always the same as anyone else's, so moral principles that specify that the good and bad fortune that people get should vary with the moral quality of their choices have no application.[34]

[32] Schlick 1966, 60.

[33] Schlick 1966, 61.

[34] Arneson 2003, 250.

In my view there is no conceptual problem with applying grey box responsibilitarianism as stated thus far (or, for that matter, black box responsibilitarianism) to a situation in which the social regulation view is (perceived to be) correct. Although Arneson is right to state that the latter view rejects notions of true virtue and viciousness, it does not reject the notion of responsibility; rather, it seeks to explain that notion, albeit in unusual terms. The operative sense of responsibility is not metaphysically deep, but it is intended to be morally significant. While virtue and desert rankings must be identical for all persons, they are irrelevant to responsibilitarianism (except, perhaps, insofar as they overlap with or are indicators of responsibility rankings). Even if every person's responsibility rankings were identical, responsibilitarianism would still be applicable. That application would yield equalities of outcome, but for responsibilitarian reasons.[35] Consequently, responsibilitarianism can be applied, and in such a way that rewards and punishments vary interpersonally on responsibility grounds. Where rewarding or punishing a person will affect their and/or others' future behaviour beneficially, the social regulation view says both that they can be rewarded, punished, praised or blamed on that basis, and that that basis is a matter of responsibility. Rewarding or punishing a different person may have different effects, in which case different ascriptions of responsibility and (on the grounds of the resulting interpersonal responsibility differential) assignments of rewards or punishments will be appropriate. The social regulation view is, then, a possible bedfellow of responsibilitarianism, though it would be a strange one.

[35] To my knowledge Arneson was in fact the first to recognize this important point, albeit in the different hypothetical context of hard determinism being true; see Arneson 1989.

It may be noted in passing that, while the social regulation view and responsibilitarianism appear to be consistent with each other, they are also doubly dissociable. That responsibilitarianism is not committed to the social regulation view is evident enough. That the social regulation view is not committed to responsibilitarianism is less obvious. It is illustrated by the possibility of one endorsing the position that the social regulation view of responsibility is correct, but that it should not have distributive or punitive consequences. In other words, one might hold that ascriptions of responsibility depend on the positive effects (especially incentive effects) such ascriptions would have, but that these ascriptions of responsibility only justify reactive attitudes of praise or blame. Naturally, the restricted implications of responsibility here will affect the ascriptions of responsibility themselves, since (mere) praising and blaming is likely to have effects other than those of rewarding and punishing. This non-distributive, non-punitive version of the social regulation view is rarely, if ever, considered in the literature, probably because the social regulation view is often assumed to be bound up with utilitarianism or other consequentialist positions that are resistant to prohibitions on distributive responses.

Socially-regulative compatibilism and grey box responsibilitarianism, as they have been stated thus far, are, it seems, doubly dissociable but harmonious. However, there remains a strong case for modifying the latter such that it precludes the former. Arneson's claim that responsibility rankings do not vary under socially-regulative compatibilism is both mistaken and, strictly speaking, irrelevant. Rewarding or punishing those who satisfy the bare, W-type responsibility conditions *or* the more discriminating X-type conditions, where these are set on grounds of social regulation, just means rewarding or punishing where this has the best consequences. This is not a

mistake for *moral* reasons, as black box responsibilitarian reward or punishment on the basis of claim (6) would be. The mistake is more a conceptual one. Application of the social regulation view reveals that neither black box nor grey box responsibilitarianism in fact captures the intuitive rationale for responsibilitarianism – that persons' treatment ought to be decided by reference to what they have (or have not) actually done. It might be right to ascribe responsibility, and allocate the associated rewards and punishments, solely for the end of bringing about better actions (ones more conducive to total utility levels, say). But that can only be the case if responsibilitarianism is wrong; that is, if the goodness or badness of the things people *have* brought about is irrelevant to their treatment. If responsibilitarianism allows responsibility to be defined in social regulatory fashion, it risks becoming a mere derivative of utilitarianism or whichever other forward-looking consequentialist view guides social regulation. The reference to responsibility would add nothing that was not already present in the foundational consequentialism; it would be a mere linguistic remnant, and since it would be purely forward looking, one somewhat at odds with ordinary language at that.

The situation here would be quite different to that where (4) or (5) are true. In those cases the black box and grey box approaches recognize that responsibility can play no retributive or distributive role. By contrast, if the social regulation view is held to be one possible account of responsibility for retributive and distributive purposes, these approaches allow the pretence of rewarding or punishing on the basis of responsibility, even though the moral load is in fact carried elsewhere – by reference to consequences. This is a recipe for conceptual confusion.

To illustrate, suppose that I have just composed a particularly inspiring orchestral piece. If attributed to me the work would likely start me on a reasonably

lucrative musical career. However, my march does have a certain resemblance to the compositions of Edward Elgar, and all things considered social consequences would be better were it accepted as his great long lost work. For one thing, its uplifting tones will receive a wider hearing if they are attributed to a more famous composer than I. It seems uncontroversial, then, that the social regulation view holds that I am not responsible for the work, and so should not be congratulated or rewarded for it. Furthermore, it holds that Elgar is responsible for the work, even though he was dead when its creation was initiated; posthumous praise and awards may even be in order, if the piece is sufficiently meritorious. Would responsibilitarianism really agree with such judgments if it turned out that socially-regulative compatibilism was our best chance of salvaging responsibility? On some combination of metaphysical and moral facts it may be the case that I am not responsible for bringing this music into being, and responsibilitarianism may adjust my entitlement accordingly. But responsibilitarianism can not hold that Elgar is entitled to anything – even so much as a good word around the dinner table – on account of the creation of this march.

As the above example shows, on the social regulation view someone's performance of an act is itself an irrelevance as regards responsibility. For this reason, the social regulation view does not in any fundamental way recognize the notion of being responsible *for some thing*, and it is hence incapable of filling out the responsibilitarian formula 'those who are responsible for some thing ought to be rewarded or punished according to the goodness or badness of that thing'. As a guide to policy it is forward looking, and is therefore quite unable to shed light on backwards-looking positions such as retributivism and luck egalitarianism.[36] For these

[36] One could deny this apparent conflict between backward- and forward-looking views by drawing attention to Rawls' suggestion that 'one must distinguish between justifying a practice as a system of

reasons it is not enough for responsibilitarianism to state that the kind of responsibility in question justifies legal or political responses. We need to further open up the box of responsibility or, to mix metaphors, make it a lighter shade of grey. The appropriate version of grey box responsibilitarianism requires that we 'reward or punish those who satisfy Y', where Y denotes whatever substantive conditions for punitive or distributive *past act-responsibility* are specified in the box. It is not enough just that any conditions for punitive or distributive responsibility have been satisfied; those conditions must also relate to the characteristics of a past act, rather than to the characteristics of future acts. Just as a person can be punished on retributive grounds only where she had the appropriate actus reus and mens rea at some temporally prior point, so too can a person be rewarded on luck egalitarian grounded only where she previously acted in a certain mental state.

V. Opening the box, part 3: restricted socially-regulative compatibilism

The social regulation view often fails to coincide with our ordinary practices and understandings of responsibility.[37] We have seen that one type of case in particular gives rise to significantly counterintuitive responses. Where an individual has not performed an act, but good effects will follow from holding her responsible, the social regulation theorist appears to recommend such a response. More emotive examples than the Elgar example are available: a socially regulatory response seems at its most

---

rules to be applied and enforced, and justifying a particular action which falls under these rules; utilitarian arguments are appropriate with regard to questions about practices, while retributive arguments fit the application of particular rules to particular cases' (Rawls 1955, p. 5). While it is true that this response resolves the conflict, it does so simply by conceding that forward-looking utilitarianism is the appropriate guide to policy. This is precisely what responsibilitarianism denies.

[37] For further demonstration see Bennett 1980; Scanlon 1988; Wallace 1994, 54-9.

unacceptable where it constitutes punishment of the indisputably innocent.[38] This seems evidently at odds with retributivism. However, Arneson suggests that socially-regulative compatibilism can resist such cases on two grounds: first, by incorporating individual rights precluding punishment of the innocent; and second, by requiring that any individual being held responsible has in fact performed an act.[39]

Arneson's first line of resistance essentially grants lexical priority to rights – effects only have an influence insofar as this is consistent with rights. To address the Elgar example one must extend the line of resistance to encompass rights to the recognition of authorship of one's good works. It may then be observed that serious wrongdoers appear to be escaping without punishment in some cases, and that this also needs rectification with further rights. In this way the forward-looking character of the social regulation view is eroded; indeed, we may define a complete set of rights and obligations such that no good or bad acts go without praise, blame, punishment, or reward, and effects rarely, if ever, come into the picture, lest rights be violated. When socially-regulative compatibilism is restricted so thoroughly it has really changed beyond all recognition: our first response to any act will be ordinary questions about who performed the act. Furthermore, it has really just become a very naïve compatibilist view: it just says that a person having performed a past act is

[38] Smilansky 2000, 28-30.

[39] Arneson 2003. There is some similarity here with Hart's position on punishment. He says that '[t]here are values quite distinct from those of retributive punishment which the system of responsibility does maintain, and which remain of great importance even if our aims in punishing are the forward looking aims of social protection' (Hart 1968, 180). Thus on Hart's view, even if we are concerned with social regulation, we should still, contra Barbara Wootton (1963), be concerned with punishing voluntary acts.

sufficient for holding them responsible for that act. A responsibilitarianism based on such a view would be subject to the arbitrariness of attribution suggestion.

As we have seen, the most acceptable (though, for our purposes, ultimately fruitless) extension of the first line of resistance holds that it is only ever appropriate to hold someone responsible when they performed the act in question, *and that holding such persons responsible is required in all such cases.* The second line of resistance rejects the italicized clause. It requires for responsibility that an act has been performed, but also that beneficial results issue from holding the actor responsible. This restricted socially-regulative compatibilism can be represented by this final set of claims:

> (8) Compatibilism is true, and the necessary and jointly sufficient conditions for compatibilist responsibility are:
>
> > (a) that a person has performed an act; and
> >
> > (b) that good effects issue from holding that person responsible for that act.

According to (7), identifying someone as responsible is just a matter of identifying them as someone whose motives and, hence, future actions can be favourably modified through praise, blame, reward, or punishment. (8) gets around the obvious shortfall of this position by adding a prior-act condition. This is a common restriction on socially-regulative compatibilism; certainly Smart's example assumes that the lazy schoolboy has in fact acted prior to the ascription of responsibility and resultant blame and punishment.

Restricted socially-regulative compatibilism appears to satisfy grey box responsibilitarianism as it has been stated so far. It does specify substantive conditions for past-act responsibility – that holding someone responsible for their past act will have good consequences – and so it does satisfy the grey box responsibilitarian requirement of rewarding or punishing those who satisfy Y, where Y denotes whatever substantive conditions for punitive or distributive past act-responsibility are specified in the box. However, it still seems unsatisfactory. Consider again my classical composition. The past-act restriction ensures that the obviously counter-responsibilitarian idea of holding Elgar responsible for the composition is prohibited. But attribution of the piece to me remains conditional. If social consequences would be better if it was widely held that the composer is unknown – perhaps the added intrigue will draw listeners – then I have no right to recognition. Again, responsibilitarianism accepts the possibility that I am not responsible for my composition (hard determinism might be true, regressive control might be required, my autonomy may be undermined locally, etc.), but it does not accept the possibility, on which the restricted social regulation view insists, that (for its purposes) my responsibility might be undermined by a favourable public response to doubt over the composer's identity. It responds to the goodness or badness of what someone has done, not the goodness or badness of the effects of someone being held responsible for what they have done. If more ambitious accounts of responsibility fail, and social regulation (restricted or not restricted) is all we can hope to achieve with the concept, then it is time for responsibilitarianism to stop referring to it. Conceptual clarity requires that responsibilitarianism does not amount to quasi-utilitarianism.

I suggest, then, this final formulation of grey box responsibilitarianism: that we 'reward or punish those who satisfy *Z',* where Z denotes whatever substantive

agency conditions for punitive or distributive past act-responsibility are specified in the box. This precludes substantive conditions that are not concerned with agency (such as ones concerning effects). It reflects the fact that grey box responsibilitarianism requires agent responsibility, or attributive responsibility. Both lines of resistance open to the social regulation view include limitations to past-acts, but they are not sufficiently discriminatory about the kinds of acts at hand. Agent responsibility requires that the act is appropriately connected to an exercise of autonomous agency. I hope to have given two good reasons why such a specification is necessary. First, it is unacceptable for a responsibilitarian that good consequences be a necessary condition for a response, even if the presence of a past-act is also a necessary condition. Second, that an act was performed in the past by an agent is not sufficient to show that it is in any way autonomous, as a responsibilitarian response requires.

The 'agency conditions' stipulation of Z may, like the limitation to punitive or distributive responsibility, seem like an obvious stipulation to make. But just as in the earlier case, it is not a stipulation widely made in the luck egalitarian literature. To some, the past-act stipulation may have seemed sufficient. Ronald Dworkin and John Roemer, two theorists often identified as leading luck egalitarians, appear to allow that acts only need to have been performed in order for them to be eligible for distributive responses. Autonomy only comes into the picture where it is undermined by 'cravings' (Dworkin) or politically recognized obstacles (Roemer).[40] Even writers such as Arneson, Cohen, and Temkin, who take care to specify that their views only respond to a philosophical sense of responsibility, tend not to specify that they are

[40] See Dworkin 1981 and Roemer 1993 respectively.

talking about agent responsibility.[41] Arneson is quite explicit that luck egalitarianism might respond to non-agent responsibility. I hope to have shown some of the difficulties associated with leaving responsibilitarianism – especially luck egalitarianism – underspecified in this way.[42]

VI. Conclusion

I have argued that a white box responsibilitarian strategy which precisely defines responsibility is overcommitted, for responsibilitarianism does not stand or fall with any particular conception of responsibility (section II). It is therefore subject to the arbitrariness of attribution objection. By contrast, responsibilitarian views such as black box responsibilitarianism which largely leave the issue of defining responsibility to metaphysicians and moral philosophers are undercommitted as some conceptions of responsibility do not provide appropriate grounds for reward or punishment, leaving responsibilitarianism open to the arbitrariness of reaction objection (section III), while other conceptions of responsibility do not provide appropriate grounds for specifically *responsibilitarian* reward or punishment (sections IV and V).

The account defended here, grey box responsibilitarianism, steers a path between the two by defining some of the substance of the conceptions of responsibility that may serve as grounds for reward and punishment, but not so much as to give a hostage to the fortune of any particular conception of responsibility. The

[41] See, e.g., Arneson 1989; Cohen 1989; Temkin 1993.

[42] It is also possible, of course, to over-specify luck egalitarianism. In later work, Cohen arguably does just this, by building conditions into his 'equal access to advantage' that should, if they are to be accommodated at all, be built into the account of responsibility to which luck egalitarianism responds; see Cohen 2004; Knight 2009b.

relevant conceptions of responsibility are both non-arbitrary grounds for distributive or punitive reaction and within the conceptual parameters of responsibilitarianism. Specifically, grey box responsibilitarianism requires that we only reward or punish on the basis of distributively or punitively relevant past-act agent responsibility.[43]

I have explained why some accounts of responsibility that fail to satisfy these requirements cannot serve responsibilitarianism. The problematic accounts appear to be those that are *moderately sceptical:* sceptical, in that they doubt whether persons are genuinely meritorious or demeritorious in the way required by traditional views of responsibility, but moderate in that they do not take this as grounds for doubting that persons can be (held) responsible. The explanation for this is quite simple. Responsibilitarianism is largely motivated by the idea that persons can be responsible in the sense of being genuinely meritorious or demeritorious, but as we have seen it has no difficulty responding appropriately to an absence of responsibility (be it local or global). Problems arise where the disconnection between the motivation – which appeals to a particular class of conceptions of responsibility – and the barebones definition of responsibilitarianism – which refers to responsibility simpliciter – is exposed. Where there is responsibility of the shallower variety endorsed by the moderate sceptic, responsibilitarian distribution lacks motivation and is no longer true to the guiding ideals of responsibilitarianism.

Within the confines of this paper I have not of course been able to survey those accounts of compatibilism and libertarianism that *do* satisfy the requirements. In

[43] The 'past-act' component of the 'past-act agent responsibility' formulation may seem superfluous to some. However, the bare idea of agent responsibility is not explicitly tied to past acts. Some views of responsibility may fill out the idea of agent responsibility with a connection between hypothetical acts and exercises of autonomous agency. Since responsibilitarianism does not respond to such responsibility, my formulation seems appropriate.

general, these can be expected to be those that retain something like the traditional conception of responsibility, and reject even the limited scepticism of the moderate sceptic. There may also be additional views that are inimical to responsibilitarianism. Some of these views may be filtered out by the grey box account described here, but others may necessitate further conditions for the kind of responsibility of interest to responsibilitarians. For instance, some accounts of responsibility may appeal to a form of autonomous agency that is inadequate for responsibilitarian purposes. The box of responsibility may need to be opened a little more.

I have also not explained how responsibilitarians assess the distributive or punitive value of responsibility. Sometimes the agent may be responsible for something which does not require a distributive or punitive response. In that case the 'reward or punishment' has a value of zero (zero units of resources, zero years in prison, etc.). In other cases the act is good or bad in a way that requires a distributive or punitive response. The issue of identifying such cases, and establishing the class and magnitude of the reward or punishment, is distinct from the question addressed here – that is, what we might be responsible for in responsibilitarian terms – but of equal importance.[44]

We can take the grey box position without insisting that non-distributive, non-punitive compatibilism or socially-regulative compatibilism of either form are wrong. Maybe it is never right to punish or reward on grounds of responsibility, though one may praise or blame on such grounds. Alternatively, maybe the right thing to do, all things considered, is to treat Elgar as the composer of the piece I have written if that

[44] The distinction drawn here is similar to those between appraisability and liability in Zimmerman 1988, and attributability and accountability in Watson 2004b. For discussion of assigning consequences to responsible choices see Dekker 2009; Knight 2009a, ch. 5.

has the best social effects. If the only existent kind of responsibility was the non-distributive, non-punitive sort or the socially regulatory sort, the responsibilitarian would simply say that the kind of responsibility with which they are (or rather, would be) concerned does not exist. In both cases the situation for the grey box responsibilitarian is, to all intents and purposes, the same as that where hard determinism is true: reward or punishment on the basis of responsibility is always inappropriate. But we would still have been right to hold that, *insofar as responsibilitarianism bases distributions and punishments on responsibility*, it responds only to full-blown, retributive and/or distributive responsibility, and only to such responsibility grounded in past acts, not future effects. Grey box responsibilitarianism recognizes legal and political implications of different conceptions of responsibility in a way that black box responsibilitarianism and non-responsibilitarian views such as utilitarianism or outcome egalitarianism do not.

The foregoing arguments will seem very abstract, and quite removed from legal and political realities. But the acceptance of grey box responsibilitarianism could have decisive impacts on people's lives in the event that retributivism or luck egalitarianism were accepted as guiding principles for major institutions of society, as their proponents would like them to be. For instance, socially-regulative compatibilism allows that people may be imprisoned, or for that matter have the welfare state's protections pulled from under them in the event of unemployment, ill health or disability, merely because such measures serve the wider interests of society. If socially-regulative compatibilism is a good conception of responsibility, black box responsibilitarianism does not have the means to deny that these measures further retributive or luck egalitarian goals. But grey box responsibilitarianism can deny that, and so alert us to the fact that such harsh treatment may be criticized not only from

the perspective of fairness but also from that of individual responsibility. In an age when political discourse increasingly urges that people are accountable for how they live their lives, it is more important than ever that we insist on demanding conditions for responsibility before we reward or punish on its basis: without such conditions the disadvantaged are likely to be disadvantaged further, and through no fault of their own.[45]

References

Arneson, Richard J. 1989. Equality and Equal Opportunity for Welfare. *Philosophical Studies,* 56, 77-93.

Arneson, Richard J. 2003. The Smart Theory of Moral Responsibility and Desert. In Serena Olsaretti (ed.), *Desert and Justice.* Oxford: Oxford University Press.

Bennett, Jonathan. 1980. Accountability. In Zak van Staaten (ed.), *Philosophical Subjects.* Oxford: Oxford University Press.

Campbell, C. A. 1951. Is 'Freewill' A Pseudo-Problem? *Mind,* 60, 441-65.

Coffey, Maureen P. 1993. The Genetic Defense: Excuse or Explanation. *William and Mary Law Review,* 35, 353.

Cohen, G. A. 1989. On the Currency of Egalitarian Justice. *Ethics,* 99, 906-44.

Cohen, G. A. 2004. Expensive Taste Rides Again. In Justine Burley (ed), *Dworkin and His Critics.* Oxford: Blackwell.

Craig, Leon H. 1975. Contra Contract: A Brief Against John Rawls' 'Theory of Justice'. *Canadian Journal of Political Science,* 8, 63-81.

[45] An earlier version of this article was presented to the Historical, International, Normative Theory (HINT) group at the University of Glasgow. I am grateful to the participants on that occasion, and also to three anonymous referees for their helpful comments.

Darrow, Clarence. 2004. *Crime: Its Cause and Treatment.* Whitefish, MT: Kessinger.

Dekker, Teun J. 2009. Choices, Consequences, and Desert. *Inquiry,* 52, 109-26.

Dennett, Daniel. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting.* Cambridge, MA: MIT Press.

Dworkin, Ronald. 1981. What is Equality? Part Two: Equality of Resources. *Philosophy and Public Affairs,* 10, 283-345.

Fischer, John M. (ed.). 1986. *Moral Responsibility.* Ithaca, NY: Cornell University Press.

Fischer, John M. and Mark Ravizza (eds.). 1993. *Perspectives on Moral Responsibility.* Ithaca, NY: Cornell University Press.

Fischer, John M. and Mark Ravizza. 1998. *Responsibility and Control.* Cambridge: Cambridge University Press.

Fleurbaey, Marc. 1995. Equal Opportunity or Equal Social Outcome? *Economics and Philosophy,* 11, 25-55.

Frankfurt, Harry G. 1969. Alternate Possibilities and Moral Responsibility. *Journal of Philosophy,* 66, 829-839.

Hart, H. L. A. 1968. *Punishment and Responsibility.* Oxford: Oxford University Press.

Hill, John L. 1988. Freedom, Determinism, and the Externalization of Responsibility in the Law: A Philosophical Analysis. *Georgia Law Journal,* 76, 2045.

Hobart, R. E. 1966. Free Will as Involving Determinism and Inconceivable Without It. In Bernard Berofsky (ed.), *Free Will and Determinism*. New York: Harper and Row.

Hurley, S. L. 2003. *Justice, Luck, and Knowledge.* Cambridge, MA: Harvard University Press.

Knight, Carl. 2006. The Metaphysical Case for Luck Egalitarianism. *Social Theory and Practice*, 32, 173-89.

Knight, Carl. 2009a. *Luck Egalitarianism: Equality, Responsibility, and Justice.* Edinbrugh: Edinburgh University Press.

Knight, Carl. 2009b. Egalitarian Justice and Valuational Judgment. *Journal of Moral Philosophy,* 6, 482-98.

MacDonald, J. E. 1955. The Concept of Responsibility. *Journal of Mental Science*, 101, 704-717.

Matravers, Matt. 2007. *Responsibility and Justice.* Cambridge: Polity.

Matravers, Matt. 2010. Mad, Bad, or Faulty? Desert in Retributive and Distributive Justice. In Carl Knight and Zofia Stemplowska (eds.), *Responsibility and Distributive Justice.* Oxford: Oxford University Press, forthcoming.

Pereboom, Derk. 2001. *Living without Free Will.* Cambridge: Cambridge University Press.

Rawls, John. 1955. Two Concepts of Rules. *Philosophical Review*, 64, 3-32.

Rawls, John. 1993. *Political Liberalism.* New York: Columbia University Press.

Rawls, John. 1999. *A Theory of Justice*, 2nd edition. Oxford: Oxford University Press.

Roemer, John. 1993. A Pragmatic Theory of Responsibility for the Egalitarian Planner. *Philosophy and Public Affairs*, 22, 146-66.

Russell, Paul. 1992. Strawson's Way of Naturalizing Responsibility. *Ethics*, 102, 287-302.

Scanlon, Thomas. 1988. The Significance of Choice. In Sterling M. Murray (ed.), *The Tanner Lectures on Human Values,* vol. VIII. Salt Lake City, UT: University of Utah Press.

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scheffler, Samuel. 2003. What is Egalitarianism? *Philosophy and Public Affairs,* 31, 5-39.

Scheffler, Samuel. 2005. Choice, Circumstance and the Value of Equality. *Politics, Philosophy and Economics,* 4, 5-28.

Schlick, Moritz. 1966. When Is a Man Responsible? In Bernard Berofsky (ed.), *Free Will and Determinism*. New York: Harper and Row.

Segall, Shlomi. 2007. In Solidarity with the Imprudent: A Defense of Luck Egalitarianism. *Social Theory and Practice,* 33, 177-98.

Slater, Eliot. 1954. The M'Naghten Rules and Modern Concepts of Responsibility. *British Medical Journal,* 2 (4890), 713-18.

Smart, J. J. C. 1961. Free-Will, Praise, and Blame. *Mind,* 70, 291-306.

Smilansky, Saul. 2000. *Free Will and Illusion.* Oxford: Oxford University Press.

Smilansky, Saul. 2003. Free Will, Egalitarianism, and Rawls. *Philosophia,* 31, 127-38.

Strawson, Peter F. 1974. Freedom and Resentment. In *Freedom and Resentment and Other Essays.* London: Methuen.

Strawson, Galen. 1986. *Freedom and Belief.* Oxford: Oxford University Press.

Temkin, Larry. 1993. *Inequality.* Oxford: Oxford University Press.

Voigt, Kristin. 2007. The Harshness Objection: Is Luck Egalitarianism Too Harsh on the Victims of Option Luck? *Ethical Theory and Moral Practice,* 4, 389-407.

Wallace, R. J. 1994. *Responsibility and the Moral Sentiments.* Cambridge, MA: Harvard University Press.

Watson, Gary. 2004a. Responsibility and the Limits of Evil. In *Agency and Answerability*. Oxford: Oxford University Press.

Watson, Gary. 2004b. Two Faces of Responsibility. In *Agency and Answerability*. Oxford: Oxford University Press.

Watson, Gary. 2006. The Problematic Role of Responsibility in Distributive Justice Contexts. *Philosophy and Phenomenological Research,* 72, 425-32.

Wootton, Barbara. 1963. *Crime and the Criminal Law: Reflections of a Magistrate and Social Scientist*. London: Stevens and Sons.

Wootton, Barbara. 2003. *Lament for Economics.* London: Routledge.

Zimmerman, Michael J. 1988. *An Essay on Moral Responsibility.* Totowa, NJ: Rowman & Littlefield.