

Contracting Responsibility

James Lenman

(This is my own final version of a paper published in Ton van den Beld (ed.): *Moral Responsibility and Ontology* (Dordrecht: Kluwer Academic Publishers, 2000). © Kluwer Academic Publishers 2000. It differs in points of layout, formatting and pagination from the published version. Please refer to latter if citing or quoting.)

I

Most of us most of the time hold most people responsible for most of their actions. We allow that there are certain conditions in which the character of an action may excuse or the character of an agent exempt the agent from being so held responsible. To hold responsible for some action someone who adequately satisfies some exempting or excusing condition would, we believe, be unfair.

Understand by *determinism* the familiar claim that the laws of nature together with the facts about the past wholly determine the future and, in particular, our future actions. Incompatibilists argue that, if determinism is true, the conditions that make it unfair to hold certain people responsible for certain actions hold universally. So that it is *never* fair to hold people responsible. This is what Jay Wallace calls the generalization strategy¹ and the most serious challenge to compatibilism.

II

Wallace defends this explicitly normative way of understanding the issue between the compatibilist and the incompatibilist in terms of *fairness*.² Such an interpretation has the merit of capturing the normative character of the issue while also respecting our dissatisfaction at many attempts to defend compatibilism on pragmatic grounds.

The simplest form of compatibilism seeks to justify practices such as punishment along consequentialist lines by appeal to their deterrent and reformatory effects. A more sophisticated pragmatic style of defence of compatibilism was brought to the debate by Strawson.³ Strawson emphasized the way in which the stance of holding others responsible is bound up with what he called the *reactive attitudes* among which he included resentment, gratitude, anger, forgiveness and certain kinds of love. To understand this, he suggested, was to understand two further things: first, that we do not seriously have the option at all of giving up this stance; secondly, that even if we did the abandonment of the stance and with it of the reactive attitudes would involve a drastic impoverishment of human life. This gives us a reason to maintain the practices and attitudes constitutive of this stance, a reason that would be compelling even given the truth of determinism.

¹ R. Jay Wallace: *Responsibility and the Moral Sentiments* (Cambridge, Ma.: Harvard University Press, 1994), pp. 16-17.

² *Responsibility and the Moral Sentiments*, chapter 4.

³ P. F. Strawson, "Freedom and Resentment" in his *Freedom and Resentment and Other Essays* (London: Methuen, 1974).

Susan Wolf in “The Importance of Free Will”⁴ has granted that such pragmatic justifications of our practices may be entirely successful but insists that they do not address the concerns of the incompatibilist. For what they, in effect, show is that we may be justified in living a lie. We might then, albeit justifiably, treat things that do not matter as if they did matter, punish and reward, praise and blame people where they in no sense merit such reactions. While it would then be rational to treat each other as free and responsible beings we could not rule it out that “as a matter of metaphysical fact we might not be free and responsible beings.”⁵

Wallace too observes that such pragmatic considerations even in their richer Strawsonian form may leave the incompatibilist legitimately dissatisfied. For although pragmatic considerations of “the gains and losses to human life”⁶ may justify our persisting in our attitudes and practices it may nonetheless be the case that “the truth of determinism would make it unfair to hold people morally responsible”.⁷ Even though the pragmatic considerations might be compelling enough for us to think we must tolerate the risk of such unfairness that would still leave us with an unsatisfying form of compatibilism and leave the issue of determinism alive as a source of legitimate philosophical disquiet.

This is a convincing explanation of why such pragmatically motivated versions of compatibilism can seem evasive. To switch to a question about fairness is to continue to address the issue in normative terms but leaves little purchase for a like charge. Suppose we had established that holding people responsible was fair. Could the incompatibilist now urge that we are evading the issue insofar as while there might indeed be considerations favouring the fairness of our practices this could at best justify them in the teeth of the possibility that they nonetheless failed to conform to the “metaphysical facts”?

At this point it seems that Strawson's charge of “overintellectualizing the facts”⁸ becomes a plausible one. For the facts that plausibly matter here are precisely those that bear on the issue of how fair it is to hold people responsible for their actions - that bear, in particular, on determining the standard exempting and excusing conditions which it is the incompatibilist's strategy to generalize. By framing the issue in these terms we allow the relevance of just those facts that matter in this way so that there seems little scope for a residual Wolfian disquiet.

III

Being held responsible for one's actions can be onerous - there are times when it will cost you something. But this is a coin with two sides - being held responsible for our actions is, on the whole, something we welcome. There is a sense, very hard to make precise but surely impossible to deny, in which to hold someone responsible for her actions is to accord her a kind of respect and to fail to do this somehow to insult her dignity.

Being held responsible for one's actions has, then, its good points. The times when these good points are least conspicuous are of course those occasions when we are caught with our fingers in the till. This happens to all of us, albeit in modest ways, from time to time. When it does happen we may seek to disown the actions in question, to plead that they were, in some sense or other, things we could not help. This costs us something in terms of self-respect but it is when things go wrong in this way that the desirability of being held responsible is most compromised.

⁴ *Mind* 90, 1981. Also in John Fischer and Mark Ravizza (eds.): *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press, 1993).

⁵ *Ibid.*, p. 118 (of Fischer and Ravizza reprint).

⁶ “Freedom and Resentment” pp. 13.

⁷ *Responsibility and the Moral Sentiments*, p. 102.

⁸ “Freedom and Resentment”, p. 23.

Whether or not things go wrong is *up to you*. This is obviously significant. For if we want to establish the fairness of our practices it is surely just what we are looking for. This is plausibly just what makes it fair to hold responsible the people we do when we do. And many of the conditions we recognize as exempting or excusing are precisely the conditions where things go wrong in ways that are not up to the agent.

IV

Here's what I'm going to do. I'm going to spin you a tale about a quite imaginary social order. For want of a better word let's call this fictitious polity *Freedonia*. In spinning this tale I will avail myself of the license of the fabulist and stipulate that in the possible world where I have set it determinism is true. Determinism is true and everybody knows determinism is true. I do that to make it hard for myself and beg no questions.

Here's what I do to make it easy for myself. Rather than think about the fairness of our own practices I will consider this issue of fairness with respect to the practices of a society where people are held responsible for their actions but where the institutions and practices that operate are somewhat different from our own. I will claim then to have described a society where it is perfectly fair to hold people responsible even though determinism is true. If I succeed in this I do not of course succeed in showing that our own society's practices are fair in this respect. For our society is different from Freedonia and the differences may render it unfair for us to hold people responsible for their actions. But if there is this difference it will be simply because the two societies have *different kinds of institutions*. It will not be for any deep *metaphysical* reason. In particular it will not be because I have described a possible world where determinism is false. *Ex hypothesi* I have not. So that if our practices of holding each other responsible are not fair that is not because some metaphysical theory is true but because our institutions are of the wrong kind. And that diagnosis is bad news for incompatibilists.

Now let me tell you some interesting things about Freedonia. First of all I will suppose the Freedonian social order governed by *norms* of many kinds and at many levels. I will secondly suppose that these norms are public norms - that their content is no secret but something to which all Freedonians have ready access.

As we will be asking the question - Is this society *fair* in holding responsible just those people whom it does? - it is worth being clear about a third supposition I will make - that the norms of the society are not patently unfair *in other ways*. For example, punishment and resentment should be incurred only by wrong actions (wrong by *our* lights) not by right or indifferent ones and they should be proportionate to the wrongness of those actions. I should not be punished for writing poems praising your apples. Nor should I be hanged for stealing them.

I will suppose fourthly that most of these norms are internalized, accepted, respected and observed by most of the people most of the time.

I will fifthly suppose that the Freedonians - or the great majority of them - possess a high degree of what, following Wallace, I will call *reflective self-control*.⁹ I don't mean exactly the same by this term as Wallace does - what I do mean is:

1. that they understand the public norms of their society.
 2. that they have the ability to govern their behaviour in the light of reasons furnished by norms they accept.
- and
3. that they have, insofar as this is possible, the ability to account for their actions when called on to

⁹ *Responsibility and the Moral Sentiments*, pp. 157-161.

do so in the light of reasons furnished by the public norms of their society.

I'll be filling this account out a bit in due course. But for now let's just notice that there is nothing metaphysically contentious about *abilities* as I am speaking of them here. Most Freedomians have the abilities I referred to here in the straightforward sense in which (the example is from Wallace¹⁰) Maria Callas had and I do not the ability to sing.

Let us further suppose that a prominent feature of Freedomian life is a certain *rite of passage*, a rite that marks the passage from childhood to the status of full adult citizenship. This has, and is understood by all to have, *a certain contractual significance*. In undergoing this rite the subject voluntarily and publicly assumes the status of a fully responsible adult citizen: she *agrees to accept responsibility* for her future actions with all that that entails.

What *does* that entail? That's a hard question- but plausibly at least by "all that that entails" we may understand a certain set of attitudes and a certain set of practices. The attitudes are the reactive attitudes, negative and positive. The practices include those involved in punishment insofar as this is expressive of such attitudes and also perhaps less unwelcome engagements between society and citizen including arguably aspects of what goes to make up the status of citizenship itself - insofar as viewing people as full responsible citizens contrasts with the way we view, say, children or people otherwise unfitted for whatever trust and respect that status brings. This is vaguer than might be wished if only because it is desperately hard to be precise about what exactly we buy into when we buy into the business of imputing responsibility. Strawson with his contrasting of the objective and participant stance gestures at a clear account but it is not easy to bring much of what he says into sharper focus.

The intuitive contrast is maybe clearest when we consider the different ways we ordinarily regard adults and children. We may love and respect children but we do not ordinarily love and respect them as we may adults. We punish them but their punishment typically serves a purely forward-looking corrective and deterrent function. We may resent them when they harm us but not, unless we are being foolish, very deeply or for very long. In our dealings with them our central value is concern for their welfare and we are far readier than with adults to let this override a concern for their autonomy. Nor do we trust them with much of a voice in the political processes in which we determine our collective futures. As adults we will regard them otherwise. We grant them a new status, a status where certain forms of unwelcome behaviour will bring them in the way of blame and resentment. It is also however a revocable status, a status which in certain circumstances an adult may lose just when we despair of her capacity for reflective self-control. In our society this status comes to us with adulthood whether we like it or not. And the central respect in which Freedomia is different is that there they get a choice. They are in effect at liberty to retain something like the status of a child if they so choose.

When *we* reach adulthood this status changes. That change is in fact a gradual and subtle one and we have no choice about whether we undergo it. In both these two respects Freedomia is different. For Freedomians the change is a stark one signalled in a rite of passage taking place at a determinate time - my motivation for this part of the story is just simplicity and nothing essential hangs on it. But for Freedomians the change is also an optional one. An agent can forego the rite of passage and thereby opt out of the enhanced status of an adult. But once they have made this commitment they may expect (barring the usual exempting and excusing conditions) to be held to it.

V

Most Freedomians, let us finally suppose, accept the bargain they are offered. Some of them

¹⁰ Ibid. p. 182.

subsequently act wrongly, incurring resentment and perhaps punishment. Is this *fair*?

What seems crucial, we saw, is that whether I act wrongly, incurring such sanctions, is up to me. It will help at this point to consider a development of this notion: the notion of what is *up to me from my perspective at a certain time*. Let us say that something is up to me from my present perspective if it is *under my control* from that perspective. That is, whether it happens depends on things about me now: my character, my dispositions, desires, values and commitments.

This sort of control is of course not enough to satisfy the incompatibilist - and for a familiar reason. These facts about myself - my character and dispositions - may not themselves be entirely up to me but depend on facts about, say, my genetic makeup, early environment and upbringing that are not - and never were - up to me. And, if we accept determinism, these facts in turn depend on facts about what happened, say, in the seventeenth century that were still more incontrovertibly not up to me.

How a Freedonian will make out is not, in this unconditional sense, up to her. But it is largely up to her from her perspective on what we will call *the Big Day* - the day she makes the decision to accept responsibility for her future actions. I think it can be made plausible that this, given her consent to the rite of passage in question, suffices for the fairness of her being held to that consent.

On the Big Day she is a certain kind of person - a person with a particular sort of character, a particular set of projects and aspirations and a particular set of values. She will take this psychological dispositional luggage with her into her incipient adult life and it will inform the way she lives that life. She is *not* regarded as in any way responsible for having this particular set of luggage and not some other set. But she has it and we are supposing she has a high degree of reflective self-control.

Here I want to flesh this notion out by suggesting it to involve and furnish grounds for the following three things: a degree of self-knowledge, a degree of self-acceptance and a degree of self-confidence.

The agent has a degree of self-knowledge insofar as she knows that this is the luggage she carries.

She has a degree of self-acceptance insofar as there is a great chunk of this dispositional luggage she's happy about. She couldn't altogether help the fact that she grew up to carry this luggage with her into adulthood but, given that here she is carrying it, that is, on the whole, fine by her. She has, at a minimum, certain values and aspirations that she wouldn't be disposed to swap for others. (There is a clear sense in which someone who is disposed to swap value set X for value set Y has already done so.) Even if there were much that she repudiates in her character, her desires and dispositions she must, so to repudiate these, possess certain values in the light of which the repudiation is carried out and which themselves she does not repudiate.

She has a degree of self-confidence insofar as she trusts her better self to more or less run the show. She believes she is able to live her life and shape her future character in ways informed primarily by those of her dispositions she accepts and identifies with. So this is not an agent who is hopelessly akratic, not one alienated from her character, desires and dispositions. So that given her self-acceptance at the level of highest order values and desires and the grip these have on her other dispositions she may be assumed to have a more general self-acceptance.

So the agent may accept a status that will license others to hold certain expectations of her if at the outset she has, more or less confidently, certain expectations of herself. For her to have this confidence and be warranted in having it, certain assumptions are required about her reflective self-control. But we are assuming that in my fictitious society people in general have levels of reflective self-control sufficient for such assumptions to be warranted most of the time.

The people of Freedonia will naturally want to have certain such safeguards in place before they will contract into responsibility in the way I have described. They will naturally want to see the recognition of certain *excusing and exempting conditions*.

The aim of these safeguards can be viewed as the aim of eliminating from the domain of one's responsibility things which involve an undue element of *luck* such that their coming about is *not under one's control from one's own perspective on the Big Day*. All kinds of luck need not be excluded but *only* those that involve factors out of one's control from one's perspective on the Big Day.

All kinds of luck need not be excluded because in Freedonia the acceptance of responsibility remains a *gamble*. An agent with reflective self control has a degree of self-knowledge and a degree of self confidence but both of these are fallible. She may not make the moral grade, incurring resentment and punishment where a more cautious decision on the Big Day would have got her off the hook. But she understood this and took the risk. The point of the exempting and excusing conditions is then to make the risk an *acceptable* one.

In particular one kind of luck that the Freedonians are *not* concerned to eliminate from their openness to imputations of responsibility is *constitutive luck*, the luck that determines that on the Big Day they have the character and dispositions that they do. On the Big Day, facing the big decision, the agent will be as she is largely on account of facts about her genetic makeup, her early environment and upbringing that she did not control. But if she chooses to accept responsibility for her future actions she is taken to accept these facts about herself - or at a minimum to accept her values and aspirations as these facts have shaped them. This acceptance is made reasonable simply by the fact that she is supposed to enjoy reflective self-control. This involves her having some capacity to choose her dispositions and desires - she can reject desires she does not like and seek to wean herself from the habits they inform while cultivating those she values. Assuming such a degree of self-control that still leaves her with the brute contingency of her happening to have the values that she does.¹¹ But there are two things to be said about this contingency. Firstly, I cannot see how the thought that it is unjust to hold someone responsible for actions she has contracted in advance so to be held responsible for can begin to be plausibly motivated by reference to it. Secondly, I do not see how the falsity of determinism could be supposed to liberate anybody from it.

This game of accepting responsibility is a gamble and a risky one. In accepting future responsibility these incipient adults take a chance. If they are going to be good citizens leading responsible and useful lives they have everything to gain. But if they are going to screw up somewhere they stand to lose a lot. For in effect by refusing to accept future responsibility they prepare for themselves a plea of non-responsibility should they ever require it. If they think the risk too high they can forego the initiation into full adulthood and go through life as children, or in any case as something less than fully responsible adult citizens.

Taking such a risk would be a matter of the sorts of self-knowledge, self-acceptance and self-trust already mentioned in characterizing reflective self-control. A Freedonian can tell whether the risk is worth taking to a large extent because she has a degree of self-knowledge and self-trust. She doesn't have perfect knowledge of herself and can't predict with confidence her dispositions and actions many years hence - so there is nonetheless a risk. But, faced with such a choice and endowed with appropriate self-knowledge, self-acceptance and self-trust, it can hardly fail to impress her that whether she does find herself in this sort of trouble is, after all, *up to her*. And the sense in which it is up to her is just the conditional sense outlined above, a sense that does not seem to be hostage to the metaphysical facts about determinism.

It is crucial to the question of fairness that the Freedonians have *consented* to be held responsible - but there is a natural worry here - *are they responsible for this original act of consent?* Inside the normative world of Freedonia this question has an odd ring to it - responsibility is *constitutive* here of a certain status that is enjoyed in virtue of this act of consent. The central issue is

¹¹ I have discussed how far we should worry about such contingency in my "Michael Smith and the Daleks: Reason, Morality and Contingency", *Utilitas*, forthcoming.

whether it is fair to hold Freedomians to what they so consent to. Here what seems important is that the contracting parties *understand* the bargain they are striking - that they meet the requirements of reflective self-control and that they are not subject to any undue *constraint* (in the traditional compatibilist's understanding of "constraint") in their choice beyond of course the costs and benefits built into the nature of the outcome.

Perhaps in Freedomia not everyone meets these conditions. Those that do not, we may suppose, are not invited to accept responsibility for their future actions, a status they cannot be expected to live up to and may not understand. But we will suppose that there are not many such people.

To sum up then, the stance of holding responsible is plausibly fair only when the conditions that make it appropriate and the exempting and excusing conditions that make its suspension appropriate are such as to make the acceptance of responsibility for one's own future actions a gamble that a reasonable person could accept. And when we construe these conditions in this way we no longer need fear the generalization strategy.

VI

Two reminders. Firstly I have assumed that in Freedomia there is extensive normative convergence but I have not assumed full convergence. Secondly I have assumed that most people enjoy reflective self control where this entails that they understand the *public norms of their society* and that they have the ability to control their behaviour in the light of reasons furnished by *norms they accept*. But what if the norms they accept are not the public norms of their society? For while I have understood reflective self-control in a way that requires agents so characterized to enjoy a degree of rationality, single-mindedness and self-mastery nothing I have said rules out their being just plain bad.

Such people have character traits, values and dispositions that they accept but that the wider society does not. In agreeing on the Big Day to be held responsible for their future actions they understand that the norms they will be asked to take responsibility for observing are not norms they themselves accept. They may nonetheless accept the terms of this offer because they think the status of responsible agent is desirable enough to be worth the sacrifice of observing norms they do not accept. Or they may not think this but still accept the offer not expecting to observe society's norms because they make a calculation *about what they can get away with*. Or, of course, like everyone else, they do not need to accept at all.

It is a matter of controversy how we should regard people who while rational do not accept the basic moral norms most of us share, who simply do not care about, for example, the harm they do to others. Is it fair to hold such people fully responsible for their actions or should we abandon such attitudes and policies in favour of something more managerial and therapeutic, something more like Strawson's objective stance. Thus we might try to contain and improve them, as we would other, more impersonal, dangerous phenomena and not think to blame them or, as a matter of *justice*, to punish them. In Freedomia either of these two ways of treating such people is possible. Which is *up to them*. But we don't give them the choice when they stand in the dock accused of serious crimes. We give it them on their Big Day, at the threshold of their adult life. And if they choose then to accept full responsibility for their future actions we hold them to that choice - and that seems eminently fair.

VII

When we think about Freedomia I suggest we can begin to see how there could be a possible social order where imputations of responsibility were fair and where this fairness was not hostage to deterministic metaphysics. That, I have suggested, is all we need to do to defeat the incompatibilist

insofar as, whatever differences there may be between this society and ours, they are not the sort of *metaphysical* differences in which the incompatibilist is interested. There are nonetheless differences about which we might properly raise questions. Perhaps, in particular, while holding people responsible is fair in Freedonia the differences mean that it is not fair in our own society.

One worry here would concern my background assumptions. In Freedonia the public norms are generally just and subject to a high degree of convergence. And most Freedonians enjoy a high degree of reflective self control. Certainly this is an idealization. Indeed if we hold a pessimistic picture of modern society or of human nature we may think the idealization grotesque, believing ourselves to be a morally derailed and fragmented society of weak and normatively confused losers. That would certainly be a depressing and problematic state of affairs but it suffices here to suggest that the falsity of determinism would hardly serve to make it any less so.

A rather different worry is that perhaps to make our own practices fair we should aim to make our own society more like Freedonia by making the status of responsible agent an optional one. That is not an altogether attractive thought. It is unattractive because in our own societies we regard the status of responsible agent as *inalienable*. Not even the agent himself can negotiate this status away.

Part of our disquiet about the optional status of Freedonian responsibility springs from tough-mindedness - we don't want to let the bad guys off the hook - but it is also a form of democratic generosity - for, as I have suggested, there is a way in which to accord someone the status of responsible agent is one of the most important ways in which we can *respect* him. I have found it extremely hard to imagine clearly what a society would be like in which all people could and some did opt out of this status but perhaps it would be a society in which there was something not unlike a class of natural slaves - a class of people whose nature unfitted them for the status of full adult citizenship where the relevant disqualifying aspect of their nature was just that they preferred to lack this status. And we may not much want a society like that even where the degraded status in question is a matter of consent.

If we reject the idea that we should be more like Freedonia, that need not mean we cannot apply my fantasy to the defence of our own practices and attitudes. What my fantasy does is to make dramatically a point that can be made without it. The status of responsible agent carries certain risks and certain costs and they come as a package. We can't have our cake and eat it here. We may want to wimp out of this status when we find ourselves in the dock but *generally* we do not. Plausibly a rational agent *would* not choose to opt out of this status if that were possible - if it were a choice that had to be made at the outset of adult life and stuck to however his life turned out. He would not be tempted by this because, from his own perspective at that outset, it is not outwith his control how his life goes. Hypothetical consent is not consent but may nonetheless do real work in normative ethical theory. Thus it has famously been thought highly significant that a certain sort of practice is one to which a reasonable person could be expected to agree. And it might plausibly be hoped that the story I have offered about the imagined Freedonia has highlighted what might prove to be reasons for saying this about our own practices of holding people responsible. That would bring the present thoughts into broad conformity with the mainstream of contemporary liberal analytic political philosophy, a place I'd be not at all sorry to find myself.

All this is to look ahead to ways in which these thoughts might be developed beyond my present objective of examining a way of making sense of the claims of compatibilism. The further developments sketched might make a fruitful project for another occasion. Suffice it for now to note that in carrying out such a project we would be seeking further to illuminate the concept of responsibility by simply *doing political philosophy*. And, in so doing we would have left what Strawson calls the "obscure and panicky metaphysics of libertarianism" securely behind us.¹²

¹² I am indebted to many people for feedback on this paper, notably Robert Audi, Paul Brownsey, Anthony Duff, Brad Hooker, Peter van Inwagen, Dudley Knowles, Maggie Little. Angus McKay,

Paul Russell, Angela Smith, Michael Smith, Elizabeth Telfer and Jay Wallace. Many thanks to all.

