A Companion to the Philosophy of Action

Edited by

Timothy O'Connor and
Constantine Sandis

WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

39

Responsibility and Autonomy

JOHN MARTIN FISCHER

The Concept of Responsibility

We use the term 'responsibility' to pick out various different kinds of responsibility, including causal responsibility, role-responsibility, and moral responsibility. One is causally responsible for some upshot insofar as one is part of the causal chain leading to that upshot, quite apart from whether one is morally accountable for it (or morally accountable at all). So, for example, the lightning bolt might be causally responsible for the fire, the earthquake for the crack in the roof, and an individual who sneezed loudly might be causally responsible for waking up the baby. There might also be cases in which an agent is morally but not causally responsible for an upshot, but this is a bit more contentious. Imagine, for example, that one does not initiate or contribute 'positively' to a causal sequence issuing in a certain upshot, but that it is one's duty to prevent the upshot, and one intentionally fails to prevent it. This is arguably a case of moral responsibility for the upshot without causal responsibility, although it might be argued that one has caused the upshot through one's *omission* to act.

Role-responsibility is a matter of the duties associated with a specific role. So, for instance, the chair of the philosophy department is responsible for arranging for certain courses to be taught every year; the mayor is responsible for making sure that the city's officials are paid; and a parent is responsible for making sure that his or her child goes to school on time. Role-responsibility typically involves moral responsibility, because the relevant roles are usually assumed voluntarily. But the two kinds of responsibility can pull apart, especially if a certain role is 'thrust upon one.' Suppose, for instance, that one is forced by the dean to be department chair at gunpoint. (Whereas this is obviously fanciful, it is perhaps not wildly implausible!) In such a case, if the chair is really *compelled* to assume the role (and not simply cajoled or importuned), then she would presumably not be morally blameworthy for not fulfilling the chair's role-responsibilities (in case she does not fulfill them).

We might call an individual 'responsible,' or say that she is 'very responsible,' meaning that she fulfills her role-responsibilities well. At a wedding, I was once asked by a relative what I write about, and I answered that I was working on a book on moral responsibility. She replied, "Good, we need more of that." I interpreted her as saying

that it would be desirable if people took their role-responsibilities more seriously. Perhaps there are role-responsibilities associated with being a human being; of course, these responsibilities are not undertaken voluntarily, but perhaps they are included in the relevant notion of responsibility, which we might dub 'substantive responsibility.'

Here I shall be concerned with a more 'abstract' notion of moral responsibility, which I shall simply call 'moral responsibility.' To be morally responsible in this more abstract sense need not entail that one is morally praiseworthy (or even blameworthy); it is a matter, very roughly speaking, of being *accessible to* or *an appropriate target for* certain distinctively normative responses.

We say that individuals or groups are morally responsible, and we also say that (for example) individuals are morally responsible *for particular items*: choices, actions, omissions, consequences, and even traits of character. Presumably an individual is a morally responsible agent insofar as he is morally responsible for at least some item. Further, it is often helpful to distinguish the *item* for which an agent is morally responsible, where this is a specification of what the agent is morally responsible for, from questions about the appropriate *degree* of praiseworthiness or blameworthiness of the agent, if indeed the agent is praiseworthy or blameworthy. It may be that an agent is not morally responsible for one item, but rather for another; and the specific nature of our response to the agent – including its 'degree' or 'intensity' – is a separate (although related) matter.

With respect to the abstract notion of moral responsibility, it is important to distinguish between 'being responsible' and 'holding responsible' (and the related notion of 'being held responsible'). The specific relationship between these notions is unclear. The standard view would be that being morally responsible is analytically prior, and thus it is appropriate to hold someone morally responsible only if he is indeed morally responsible (where the status of being morally responsible is independently established). Jay Wallace, however, has argued that the order of explanation is quite the opposite (Wallace 1994). He claims that one's being responsible is to be analyzed in terms of its being fair to hold one responsible. (For a critical discussion, see Smith 2007). Although I would agree with Wallace that there is a fundamental connection between 'being responsible' and 'holding responsible,' I do not think that Wallace has adequately captured this connection; after all, one could be an appropriate candidate for certain distinctively normative responses in the sense that it would not be a category mistake to target one with those responses, even though it might not be *fair* to target one in this way. Here I would contend that the agent is morally responsible (insofar as he is an appropriate target for the responses constitutive of holding responsible), even though it would not be fair to hold him responsible.

We have already seen that 'responsible' is multiply ambiguous. It should also become increasingly clear that even 'moral responsibility,' construed abstractly, is multiply ambiguous. Making the distinction between 'being responsible' and 'holding responsible' is just the beginning of uncovering the complexity; indeed, philosophers frequently 'talk past each other' precisely because they fail to recognize that they are operating with distinct notions of moral responsibility (Fischer and Tognazzini, forthcoming).

For an agent to be morally responsible for an item is, in my view, for that item to be attributable to the agent in a way that would make it in principle justifiable to react to the agent in certain distinctive ways (on the basis of the item in question). Being respon-

sible, on this view, is a kind of "attributability" (Watson 1996). More specifically, one might distinguish two sorts of attributability. For an agent to be responsible for an item might be for the item to be attributable to the agent in a way that would make the agent a sensible or appropriate target for what Watson calls "aretaic appraisals"; these concern the agent's moral virtues and vices as manifested in thought and action (ibid.). On a slightly broader version of this first kind of attributability, for an agent to be responsible for an item would be for the item to be attributable to the agent in a way that would make the agent a sensible or appropriate target for distinctively moral judgments on the basis of the item (where these judgments could encompass more than virtues and vices).

We can distinguish the first sort of attributability from a second. For an agent to be responsible for an item, on the second view, would be for the item to be attributable to the agent in a way that would make the agent an appropriate or sensible target for what Peter Strawson called the "reactive attitudes" (and related practices such as punishment; Strawson 1962). Strawson's reactive attitudes include gratitude, indignation, resentment, hatred, love, and forgiveness. That someone is an appropriate or sensible target for such an attitude implies that it would not be a category mistake to have such an attitude toward that individual; in contrast, it would seem to be a category mistake to be resentful of a goldfish, or to be grateful to a rodent. Importantly, when one is an appropriate or sensible target for such attitudes, it is in principle justifiable that one should be the target of the attitudes in question; but of course it does not follow that in any given context anyone is actually justified in targeting one with the attitudes. This point helps to explain the difference between 'being responsible' and 'holding responsible' (and the twin notion of 'being held responsible').

Being responsible, then, is a matter of attributability – either 'aretaic attributability' (or, more broadly, 'normative attributability') or 'reactive attributability.' *Holding* someone responsible, on the other hand, is a matter of actually targeting the individual with the relevant attitude or judgment. Again, the judgments and attitudes in question might be either 'aretaic,' 'normative in general,' or the 'reactive attitudes.' If one judges someone to be blameworthy, this could be considered a form of holding that individual responsible – quite apart from any outward expressions which flow from that inner judgment, any reactive attitudes or public expressions of such attitudes, or any sort of harsh treatment or condemnation. Similarly, having indignation or resentment toward someone is to hold that individual responsible, quite apart from any outward or public expression of these attitudes, or any further condemnation or harsh treatment. Of course, the outward or public expression of reactive attitudes would be an additional way of holding someone responsible; I simply wish to point out that these are not *essential* to holding responsible. Similarly, harsh treatment or condemnation would be forms of holding an individual responsible; but, again, they are not *essential* to holding responsible.

Moral responsibility, then, is a complex notion, involving various different aspects (or, in Watson's word, "faces"). Here I have focused (in a sketchy way) on the concept of moral responsibility or, alternatively, on the 'essence' of moral responsibility (what it is to be morally responsible). This of course leaves open the conditions under which the concept applies (or the conditions in which moral responsibility is actually present). Perhaps the most salient and contentious issue here is whether moral responsibility is

311

compatible with causal determinism (or universal causation of a deterministic sort). There are also questions about the relationship between moral responsibility and causal indeterminism, and between moral responsibility and various kinds of freedom or control.

Discussion of these issues is beyond the scope of the present chapter.

Responsibility and Autonomy

It is striking that there are 'parallel literatures' in contemporary philosophy about the similar – but different – notions of responsibility and autonomy. Indeed, similar analyses have been suggested, similar objections made, and so forth. Often there has been a perplexing lack of communication or contact between the two literatures; it is as if they were parallel philosophical universes. I seek to bring them together, at least in a preliminary way, here.

In my view, moral responsibility is a necessary but not sufficient condition for autonomy. (We have seen that responsibility is a complex notion; there is a similar complexity in the notion of autonomy.) In order to be an autonomous agent (or to act autonomously in a given context), one must be a morally responsible agent (or act in such a way as to be morally responsible). But some additional features must also be present; one can be morally responsible without being autonomous. Put metaphorically, the crucial additional ingredient is: 'listening to one's own voice' or 'being guided internally.' Of course, it is difficult to make these metaphors more precise, but the idea is that one can meet the conditions for moral responsibility without meeting the additional conditions for autonomy because one is not, in the relevant sense, being guided internally.

To exhibit the parallel nature of the contemporary discussions of moral responsibility and autonomy, and to bring out an important difference between the two notions, I shall begin by discussing the 'hierarchical analysis' of 'acting freely.' In the context of moral responsibility, this analysis has been developed by Harry Frankfurt (1971); in the context of autonomy, this sort of analysis has been developed by Gerald Dworkin (1970 and 1988).

Frankfurt distinguishes "first-order preferences" – preferences for states of affairs or actions – from "second-order preferences" – preferences about one's first-order preferences. He contends that we, unlike mere non-human animals, can step back from our first-order preferences and form preferences about them. A subset of our second-order preferences is the set of "second-order volitions," which, according to Frankfurt, are the second-order preferences that specify which of our first-order desires we wish to move us to action. Frankfurt gives us a "mesh" theory of acting freely; that is, on his approach, an agent acts freely insofar as there is a mesh between his second-order volition and the first-order preference which actually does move him to action. On Frankfurt's "hierarchical" approach, one fails to act freely insofar as the mesh does not exist, that is, insofar as there is not a harmonious match between one's second-order volition and the first-order preference that moves one to action. Although the details differ slightly, Gerald Dworkin gives a similar hierarchical account of "acting freely" (in the context of his discussion of autonomy).

A crucial point for Frankfurt is that the second-order volition need *not* be based on any sort of rational reflection or normative considerations. On the view suggested by Frankfurt (1971), an agent "identifies" with the selected first-order preference by forming a second-order volition (to act in accordance with it). That is, the view suggested by Frankfurt in his classic 1971 paper is that forming a second-order volition constitutes *identification*, and identification plus acting on the selected first-order preference (the first-order preference identified with) constitutes acting freely. On this view, it does not matter what the *basis* of the identification is, or whether it has any basis at all. It is also important to note that Frankfurt was seeking to characterize an element of an account of acting freely which would fit into a theory of moral responsibility. In his early work he seeks to give an account of "acting freely," which, he contends, is the freedom component of the conditions for moral responsibility. And his account of "identification" is supposed to be the ingredient that, when added to action on the selected first-order desire, gives us the relevant notion of freedom – the notion that plays a role in the analysis of moral responsibility.

Now various problems have been raised for the analysis suggested by Frankfurt (1971). Perhaps the classic presentation of the main worries is Gary Watson's 1975 paper. Here Watson raises the "regress" objection. The worry is that, if mere action in accordance with a first-order preference is not enough to confer the status of "acting freely," why exactly does it help to add a second level? That is, why exactly does it help that one has in place a mesh between a second-order volition and the preference on which one acts? Why does adding a second level help, when one can presumably ask the same questions about the provenance and basis of the preferences (volitions) at the second level as one can about those at the first? It is as if one were told that the earth is in place because it is standing on a giant tortoise, and, when one asked what keeps that tortoise in place, one were told that the first tortoise is standing on another giant tortoise. Perhaps it is not conceptually impossible for there to be tortoises all the way down; but is it really plausible that, whenever we act freely, we have an infinite number of levels of preferences, all in harmony?

Much of Frankfurt's subsequent work on these topics, especially the central notion of "identification," can be understood as attempts to answer Watson's regress objection (Frankfurt 1988). Dworkin also gives considerable attention to parallel worries in the context of his own development of the hierarchical approach to acting freely (Dworkin 1988). Whereas various theorists have insisted that we must add some ingredient which states that the *basis* of the second-order volition must include 'rational reflection' or some sort of normative notion, Frankfurt himself has resisted this move. He has consistently adhered to the contention that the "identification" in question need not be based on rational reflection or have any particular substantive content.

I think this point is important, and I believe that it helps to clarify the difference between moral responsibility and autonomy. Recall that Frankfurt's initial suggestion of an account of identification in terms of a mesh between elements of the first and second levels in our motivational economy was made in the context of seeking an analysis of moral responsibility, *not* autonomy. I claim that Frankfurt's minimalist 'Humean' notion of identification fits better into an account of moral responsibility than into an account of the more substantive notion of autonomy. Further, I believe that there has been a subtle migration of the crucial notion of autonomy in Frankfurt's own

thinking about these issues over the years – a migration that has lead to what might be called 'mission creep.' That is, Frankfurt initially discussed "identification" in the context of seeking to specify the crucial additional element in an account of moral responsibility; in his later work, he speaks of "identification" as helping to specify the "true self" or "real self." Other theorists have also framed their discussions in terms of such notions and of the related idea of 'agential authority.' But it is unclear what exactly the relationship is between the notions of 'real self,' 'true self,' and agential authority on the one hand, and (mere) moral responsibility on the other.

To understand my view here, recall that various theorists have urged Frankfurt to *add* a value component to his hierarchical account of identification (and of acting freely). One such theorist is Eleonore Stump (1988). Others, including Gary Watson, have rejected the hierarchical approach entirely, but have nevertheless insisted on a "normative" or "value" component to their analysis of acting freely (Watson 1975 and 2004). Let us call any approach to giving an account of acting freely which contains a significant requirement that the agent should act in accordance with what he takes to be rationally or normatively defensible a 'value-added' approach. (I owe the term to Andrew Eshleman.) Clearly, there are hierarchical and non-hierarchical value-added models of acting freely.

The problem for *any* value-added approach to acting freely is weakness of the will. That is, it certainly seems that we can act freely but *against* what we take to be normatively defensible (in any of the senses given by the various value-added models). Given this, we cannot define acting freely in terms of acting in accordance with what one finds normatively defensible; similarly, we cannot define acting freely in terms of what we do or would reflectively endorse. Just as Frankfurt has spent much of his career seeking to respond adequately to the regress objection, Watson has given considerable attention to the problem of weakness of will. In my view, neither Watson nor any of the value-added theorists (sometimes called 'normative theorists') has addressed this problem successfully, insofar as the notion of acting freely is the notion relevant to *moral responsibility*. (Note that it does *not* seem straightforward that it is possible to act freely, in the sense required for moral responsibility, when one acts against a Frankfurtian second-order volition, where that is *not* necessarily based on rationality or normative defensibility. In my view, then, the problem of weakness of will does not threaten Frankfurt's 'minimalist' hierarchical approach to giving an account of the sort of freedom implicated in moral responsibility.)

But suppose that we make it explicit that the target of our analysis is the notion of acting freely in a sense which is relevant to autonomy. Now it is not so clear that the value-added model leads us astray. My claim is that we can act freely, in the sense relevant to moral responsibility, and nevertheless exhibit weakness of will; thus our acting freely in the sense relevant to moral responsibility must be consistent with acting against what we take to be normatively defensible, and the value-added approaches are problematic here. Insofar as the initial intention of Frankfurt was to give an account of acting freely that would play the required role in a theory of moral responsibility, he has been correct to resist the move to adding a 'rational' or 'normative' component to the basis of the second-order volition. But things are different if the target of our analysis is the notion of acting freely that is relevant to *autonomy*. It is not clear to me that one is acting *autonomously* if one is acting against what one takes to be normatively

314

defensible. It is not implausible, then, that the value-added model (in contrast to Frankfurt's more minimalist approach) gives a promising account of acting freely, or at least a necessary element of such an account, insofar as this is the notion of freedom relevant to autonomy.

Think about it in terms of the admittedly metaphorical and vague terms, 'real self' or 'true self.' There is a notion of 'self-governance' that can be understood as governance by the real or true self. Further, if one identifies the real or true self with the value module, then it would turn out that self-governance (of the relevant sort) would be governance by the value module (that is, governance in accordance with what one takes to be normatively defensible). This sort of self-governance seems to me to be something more than mere moral responsibility; it appears to come close to the more substantive notion of autonomy.

If this is correct, then it suggests the following (no doubt oversimplified) diagnosis. Frankfurt originally suggested that the pertinent mesh (the harmony between the second-order volition and the first-order preference on which one acts) is a sufficient condition for identification, which is the crucial missing ingredient in the account of acting freely. But acting freely plays a role in an analysis of moral responsibility and also of the related but more demanding notion autonomy. The initial context in which Frankfurt's discussion took place was an attempt to give an analysis of moral responsibility. But there has been a 'slippage' or 'mission creep' over time, both in the work of Frankfurt and in that of others who have participated in these discussions, so that there has been a conflation of considerations relevant to moral responsibility and considerations relevant to autonomy. In particular, although the value-added approaches are inadequate as accounts of the freedom involved in moral responsibility, they are not thereby to be ruled out as accounts of the freedom involved in autonomy.

Autonomy, as is suggested by etymology, might be regarded as 'self-governance.' I have suggested above that the value-added model might be interpreted as giving an account of the relevant self-module, as it were, and thus of the notion of self-governance. I have further suggested that mere moral responsibility does not demand this sort of self-governance or inner guidance. This suggests that the value-added model of acting freedom gives a necessary condition for autonomy, but nor for moral responsibility. It is unclear, however, whether it also gives a *sufficient* condition for the kind of freedom involved in autonomy. This is because it might be that an individual's normative orientation is such that she takes it that it is normatively defensible for her to take her cues from others in certain ways. For example, a 'deferential housewife' might take it to be normatively defensible to be deferential to her husband in certain ways. If this is so, then it is not clear whether we should say that acting in accordance with such normative views renders the agent free in the sense implicated by autonomy. This is a complicated question, which I cannot take up here. The value-added model does seem to me to offer at least a necessary condition for the notion of freedom required for autonomy; after all, what would we make of an individual who acts against what she reflectively endorses, by failing to be deferential? That is, I do not think we would be confident in thinking that someone is acting autonomously in standing up for her own interests, when she does *not* think that this is normatively defensible. This problem illustrates a tension between a *structural* and a *content-based* account of autonomy.

See also: HABITUAL ACTIONS (10); VOLITION AND THE WILL (13); MENTAL ACTS (27); AGENT-CAUSATION (28); MOTIVATIONAL STRENGTH (33); ADDICTION AND COMPULSION (34); AKRASIA AND IRRATIONALITY (35); FREE WILL AND DETERMINISM (38); ACTION AND CRIMINAL RESPONSIBILITY (42); INTENTION IN LAW (43); SCIENTIFIC CHALLENGES TO FREE WILL (44).

References

Dworkin, G. (1970). Acting freely. *Nous*, 4, 367–383.

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Fischer, J. M., and Tognazzini, T. (forthcoming). The physiognomy of responsibility. *Philosophy and Phenomenological Research.*

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68, 5–20.

Frankfurt, H. G. (1988). *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Smith, A. (2007). On being responsible and holding responsible. *Journal of Ethics*, 11, 465–484.

Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 187–211.

Stump, E. (1988). Sanctification, hardening of the heart, and Frankfurt's concept of free will. *Journal of Philosophy*, 85, 395–420.

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

Watson, G. (1975). Free agency. *Journal of Philosophy*, 72, 205–220.

Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, 24, 227–248.

Watson, G. (2004). *Agency and Answerability*. Oxford: Clarendon Press.

Further reading

Feinberg, J. (1987). *Harm to Others*. New York: Oxford University Press.

Feinberg, J. (1988). *Offense to Others*. New York: Oxford University Press.

Feinberg, J. (1989). *Harm to Self*. New York: Oxford University Press.

Feinberg, J. (1990). *Harmless Wrongdoing*. New York: Oxford University Press.

Fischer, J. M. (1994). *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell.

Fischer, J. M, (ed.) (2005). *Critical Concepts in Philosophy: Free Will*, Vols 1–4. London: Routledge.

Fischer, J. M. (2006). *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.

Fischer, J. M., and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66, 829–839.

Oshana, M. (2006). *Personal Autonomy in Society*. Aldershot: Ashgate.

Pereboom, D. (2001). *Living without Free Will*. Cambridge: Cambridge University Press.

Taylor, J. S. (ed.) (2005). *Personal Autonomy: New Essays in Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press.