

RELATING TO RESPONSIBILITY

ESSAYS FOR TONY HONORÉ ON HIS
EIGHTIETH BIRTHDAY

Edited by
Peter Cane
and
John Gardner



OXFORD – PORTLAND OREGON
2001

THE CAPACITY TO HAVE DONE OTHERWISE: AN AGENT-CENTRED VIEW

*Philip Pettit**

INTRODUCTION

Whenever we hold someone responsible for a given action we assume that in some sense they could have done otherwise. Perhaps, unbeknownst to them, they would have been forced to perform that action had they not chosen to perform it voluntarily; perhaps they would have been induced by neuro-scientific or hypnotic or strong-arm tactics to act in that same way (Frankfurt, 1988). But even in such a case they could presumably have tried to do something else instead (Otsuka, 1998; Fischer, 1999). And were that not so then it is hard to see how we could continue to hold them responsible for what they did.

Not only does the fact of holding someone responsible commit us to believing in this sense that they could have done otherwise. Believing that they could have done otherwise also involves us in believing that they were responsible, at least in some measure, for what they did; they were responsible in the sense that there is nothing wrong about praising them or blaming them for what happened. If the agent could have done otherwise, as we naturally understand that condition, then there is no ground on which they can avoid having the action put down to their credit or discredit.¹

I shall say no more here to defend the assumption that responsibility for an action and the capacity to have done otherwise go together. Assuming the linkage alleged, the paper is concerned with the question of how we should understand the capacity to have done otherwise. The standard line is that it

* My thanks to Victoria McGeer, François Schroeter and Michael Smith, and to Peter Cane and John Gardner, for useful comments on an earlier draft.

1. The linkage between being responsible for an action and having the capacity to have done otherwise is quite consistent, notice, with responsibility for other matters—say, responsibility for negligent inaction—not involving any parallel capacity. And it is quite consistent, of course, with something less than responsibility proper—say, strict liability under a regime of law—not involving such a capacity. Those are separate topics and require separate treatment.

must consist in a special feature of the way the action was produced. I think that this act-centred line has proved unproductive and I want to explore instead an approach that would equate it with a more general, agent-centred capacity. The approach is in the spirit of Tony Honoré's (1999: 11) claim that "in morals and law capacity means a person's general capacity to perform successfully"; it draws on a point of view that I have tried to develop more fully in a recent book (Pettit, 2001; see, too, Pettit and Smith, 1996).

This paper is in three main sections. In the first I look at the problems that bedevil approaches that give the capacity to have done otherwise an act-centred reading, and argue that they all fail at least one of two intuitive constraints, one naturalistic, the other normative. In the second section I introduce the alternative approach, according to which the capacity is agent-centred in character. And then in the third section I argue that this agent-centred approach promises to satisfy both of the constraints introduced earlier. The general line defended is summed up in a brief conclusion.

1. THE ACT-CENTRED APPROACH: A CRITIQUE

The standard approach to explicating the capacity to have done otherwise assumes that the agent displays this capacity in virtue of the way the action arose. The idea is that whatever was true of the agent in other respects, the process leading to the action satisfied a certain distinctive sort of condition and that the satisfaction of this condition makes it the case that he or she could have done otherwise. All versions of the standard approach agree on that central assumption. They diverge from one another only when it comes to addressing the further question as to exactly what act-centred condition makes it the case that the agent could have done otherwise.

The big divide that appears as soon as that question is asked is the division between approaches that are compatible and approaches that are incompatible with the truth of determinism (Watson, 1982). Incompatibilists argue that for it to be the case that the agent could have done otherwise, it must be that the laws which govern the natural world, together with the history of the world up to that point, left open the possibility that the action-generating process should have had a different output. There must be a gap in the causal order of the world such that nothing that happened before the action can have made it inevitable that the process led to precisely that upshot (Van Inwagen, 1983).

Compatibilists hold, by contrast, that this is too strict and that all that is required is something weaker: viz., that the sorts of antecedents that we spontaneously treat as hostile to agency—antecedents like natural obstructions,

psychological compulsions, *idées fixes* and the like—should not have ensured, all on their own, that the process led to a particular result. The action will have been fully determined by its antecedents taken as a whole—antecedents that at the psychological level will include the agent's normal beliefs and desires—but it will not have been fully determined by hostile antecedents alone. The agent will not have been robbed of his or her decision-making capacity by the intrusion of such "constraining causes" (Ayer, 1982: 21).

There are two constraints that any account of the capacity to have done otherwise should satisfy and all versions of the standard approach, compatibilist and incompatibilist, fail to meet one or the other of these. So, at any rate, I shall argue.

The first constraint is that the account should be consistent with what we may describe as a naturalistic picture of the universe. A picture of the universe will count as naturalistic just so far as none of the entities or forces that it posits are add-ons to those entities and forces—whatever they are—that are recognised in the scientific image of the world.² None of them is of a kind with the *res cogitans* postulated in Descartes' theory of mind, for example, or with the *vis vitalis* posited in early theories of life. The constraint holds that an account of the capacity to have done otherwise should be naturalistic in this sense; it should not force us to reject the image of the world projected in natural science.

The second, normative constraint is that the account should also make sense of why someone who could have done otherwise is thereby taken as fit to be held responsible for it in some measure, being subject to praise or blame for what was done. The capacity to have done otherwise is associated with the agent's being fit to be held responsible, as we saw earlier, and the account must therefore explain why it is permissible to praise or blame someone who possesses the capacity for doing what they did. It will be permissible to praise or blame them in this sense so far as it is permissible, not just to make an assessment of what they did—this could be quite detached—but to react to them in the positive mode of gratitude or appreciation or in the negative mode of resentment or indignation (Strawson, 1982).³

2. How to demarcate the entities and forces that can be recognised in the scientific image? I equate them with the entities and forces that are guaranteed to figure in the world by virtue of the way things are in the microphysical realm postulated by physics (Pettit, 1993b). But other lines are also possible on that demarcation problem—see Jackson (1998)—and we do not need to judge between them here.

3. In the case where an agent acts on behalf of a principal, in particular on behalf of a group, it may be unclear who should be the target of such a reaction, the individual or the collective. My own view, defended in Pettit (2001), is that even if the collective is the entity that ought to be praised or blamed for what was done through its agent, still that agent will be properly subject to praise or blame—assuming that nothing like duress or coercion was involved—for having done what the group required.

Let us consider incompatibilist accounts in the light of these constraints. One such account is naturalistic in character, arguing that natural laws may leave it to objective chance whether one or another event occurs at a given time; the laws may ensure that A does not occur, while leaving it possible that either of B or C occur. Such an indeterministic picture might make it true, for some actions, that the agent could have done otherwise: it might allow that though the agent performed an action, B, the laws of nature and the history of the world up to that point were equally consistent with the action-generating process leading to C instead. But while this incompatibilist account would satisfy the naturalistic constraint, it would clearly fail the normative one. For why would the fact that an action was underdetermined by the laws and history of the natural world make it permissible to praise or blame the agent for performing it? All that the story implies is that the action was the product of objective chance and that scarcely permits us to lay it at the door of the agent.

What incompatibilists clearly need to do is not just to deny that the laws of nature uniquely determined what was done by a responsible agent but also to postulate that it was determined within the agent in a way that implies responsibility: the action was produced and owned by the agent in a distinctive manner. The problem for this version of indeterminism, however, is how to give a naturalistic account of such “agent-causation” (Chisholm, 1982). Under any naturalistic story, the only causes in the natural world will be causes that operate in virtue of the natural laws that are operative there, and the causation whereby an agent is held to close the gaps left by indeterministic, natural laws cannot itself belong in that naturalistic category. This variety of incompatibilism will violate the naturalistic constraint.

Compatibilist accounts of the responsible agent’s ability to have done otherwise, where this is also understood as a special, act-centred capacity, run into parallel difficulties. Such accounts are well-fitted to satisfy the naturalistic constraint but none of them succeeds in meeting the second, normative constraint. They identify the capacity to have done otherwise with a naturalistically intelligible, act-centred condition but none of the conditions canvassed makes sense of why it is permissible to praise or blame the agent for an action performed when it obtains.

The usual compatibilist account is hypothetical in character. It says that an agent could have done otherwise at a given time just in case he or she would have done otherwise if a certain condition had been fulfilled. One line takes the condition to be that had the agent chosen, then he or she would have done otherwise (Moore, 1911; Ayer, 1982). But this is no good. Choosing or willing or any such cognate is an action, so that there will always be a question as

to whether it itself satisfies that condition, and this question will open up an indefinite regress. The now more standard line takes the condition to be that had the agent not desired to act in that way, then they would not have done so (Davidson, 1980). This proposal is not subject to the same difficulty as the one that invokes choice, since desiring is not an action and the question does not arise as to whether it itself is a free action.

This standard proposal, however, is problematic in another respect. For all the condition about desire stipulates, it may be that in order for the agent’s desires to have gone the other way, the conditioning and drilling to which the agent was subject as a child would have had to be different from what it actually was (Chisholm, 1982). Perhaps the agent could have done otherwise, in the sense that he or she would have done otherwise in the event of having a different desire. But it may be that the agent could not have desired otherwise; it may be that the desire that actually produced the action was more or less hard-wired into the agent’s make-up.

The problem involved is, of course, that under this compatibilist account of the capacity to have done otherwise, it is not intuitively permissible to praise or blame the agent for an action. If an agent does something under the influence of a desire that is more or less hard-wired into their make-up—and this, through no fault or virtue of their own—then they can hardly be praised or blamed for what is done. If a person is subject to such an unavoidable motive, then by ordinary criteria it will not be permissible to praise or blame them for acting on that motive.

Can this problem be resolved by requiring that the desire from which the action issues be one that the agent desires to be moved by, as distinct from being a desire that operates willy-nilly (see Frankfurt, 1988)? I do not think so. For just as the first-order desire manifested in the action may have been hardwired into place, so the same may be true of the higher-order desire to be moved by that desire at the first-order level (Pettit, 2001).

This short review of variations on the standard approach to our problem should be sufficient to explain why I am pessimistic about finding any that will answer to the naturalistic and normative constraints. No naturalistic account is likely to satisfy the normative constraint and no account that satisfies the normative constraint is likely to be naturalistic. Or so at least it seems.

2. INTRODUCING THE AGENT-CENTRED APPROACH

The alternative approach to explicating an agent’s capacity to have done otherwise starts from quite a different reading of the remark: “X could have

done otherwise". Rather than assuming that that remark is meant to direct us just to something about the way the action was generated within the person, it suggests that the intended interpretation bears on the sort of agent that X more generally is.

A good way to introduce this alternative reading may be to consider the import of parallel remarks in other contexts. Think of a mechanic who reflects on how two cars did in a particular race and says that one performed up to its limits while the other was capable of achieving a better time. Or imagine a horse trainer who considers the performance of two animals in a dressage event and says that while the first gave of its best, the second could have done better. Or think of an engineer who says that while two missiles did equally well in homing in on a target, one of them—a smart missile, let us suppose—could have achieved a greater level of accuracy.

The brunt of such remarks is to comment on the cars or horses or missiles in general terms, not to reflect on something that holds just in virtue of how the performance was generated. What we are told in each case is that one of the pair in question has a general capacity that is lacking in the other and that the similar performances of the members of each pair do not reflect this difference in capacity. To say that one car or horse or missile could have done better is to say that it is a better car or horse or missile and that if this was not reflected in actual performance, then that was due to an accident of circumstance. It is to suggest that we should not judge the entities in question—in particular, we should not judge their general capacities—on the basis of this performance. That performance was not typical; it was not one in which the car or horse or missile performed to type.

The agent-centred approach to explicating the responsible agent's capacity to have done otherwise suggests that something similar is true in this case. When we say of an agent that he or she could have done otherwise, so the idea goes, we are presupposing the relevance of certain background standards and, in the case where the agent fails to meet those standards, we are saying that this failure was not typical. The agent could have done otherwise, we remark, intending to convey the thought that he or she is capable of better. The actual choice they made may have fallen away from those standards but this should not be taken as indicative of the sort of agent they are; it should be put down to the influence of a more or less incidental factor.

It need not be the case under this reading that there was any point in the process leading to the action where we can see the presence of something we might describe as volition. For all that is implied, the process leading to the action performed may have made it absolutely inevitable that the agent should have acted as they did. The remark that the agent could have done

otherwise, interpreted on the proposed lines, means only that an incidental feature of the circumstances played an important role in the process leading to action and that the behaviour would have been different—in particular, it would have been more typical of the agent's character—had that feature been absent. The incidental feature that played this role may be an event that distracted the agent's attention, or the presence of passion or fatigue or boredom, or just a glitch in the way the agent's memory worked. The possibilities are endless, though how we enumerate them—what we count as perturbers that interfere with the exercise of a capacity, without undermining its existence—will depend on our background view of agents in general and that agent in particular.⁴

This may explain what is meant by saying in the case of failure to meet certain standards that an agent could have done otherwise. But what of the case of success? Suppose that the agent does act to type and satisfies the relevant standards. What does it mean to say in that event that the agent could have done otherwise? We might say it means that they would or might have acted otherwise had an incidental feature of some kind thrown them off form. But while that would mirror the story told for the case where the agent fails, it would not give us an intuitive account of the content of the remark. When we say that someone who does well could have done otherwise we surely mean to convey something positive about the agent, not merely the negative message that they might have been put off their stroke by this or that perturbation. Is it possible to vindicate that intuition within the agent-centred approach?

I think that this is possible. There are two interpretations of what it means for agents to be disposed to act so as to satisfy certain standards—to track certain standards—and we have only been allowing for one. Under a first, weaker interpretation it means that the agents are disposed to act in those ways that, as it happens, are in line with the standards. Under a second, stronger interpretation it means that they are disposed to act in line with the standards, whatever the standards should happen to require: they are actually disposed to act in the ways that, as it happens, are in line with the standards, but had the standards required different modes of action then they would have been disposed in that counterfactual event to act in those different ways. The agents in the first case are cued to the behaviours that happen to satisfy the

4. Some may even think that the required capacity can remain in place consistent with more or less continuing failure. They may think that the type to which we assign the agent is not to be determined by empirical performance, or not just by empirical performance, but by an independently sourced sense of counterfactual possibility.

standards, the agents in the second case are cued to the standards themselves: they are aware of them as standards that they should meet; they are reliably disposed to enact whatever they take the standards to require in any context; and they are reliably disposed to interpret the demands of the standards correctly.

If we stipulate that agents have the general capacity to meet certain standards in the second, stronger sense, then it turns out that we can get over the difficulty raised. Suppose that an agent does well on a particular occasion, manifesting the general capacity envisaged. If we think that he or she has that capacity in the strong sense, then the remark that they could have done otherwise need not mean that they might have been thrown off form by an incidental factor. It can mean that their satisfying the standards is no accident. They would have acted so as to satisfy the standards even if the standards had required a different form of behaviour. They are explicitly focused on the demands of the standards and they are reliably disposed to meet those perceived demands.

The picture emerging, then, is this: if an agent does badly and we say that he or she could have done otherwise, then we mean to suggest that they have a general capacity to do better and that the failure should be treated as an accident, not as something typical. If an agent does well and we say that he or she could have done otherwise, then we suggest in similar vein that they have a general capacity to do well—they would have done well even if the standards had required a different response—and that the success should not be treated as an accident but as something typical of them. To say that an agent could have done otherwise is always to speak about the agent involved in the act, not just about the act itself, and it is always to make a positive comment about that agent.

3. ENDORSING THE AGENT-CENTRED APPROACH

How is such an agent-centred account likely to fare with the naturalistic and the normative constraints? Does it postulate a capacity that we can imagine a regular, naturalistically unmysterious creature possessing? And does it postulate a capacity such that any action done in the presence of the capacity is one for which it is permissible to hold the agent responsible: to praise the agent if the capacity present is actually exercised, to blame them if it is present but unexercised? We turn to those questions in this final section.

The naturalistic constraint does not raise any distinctive problems for the agent-centred approach. There is no particular difficulty in giving a naturalistic account of a capacity or disposition, since nature is rich in propensities

of that kind. True, the approach postulates a strong capacity to satisfy standards, which in turn requires an explicit awareness of standards and a disposition to try to meet their perceived demands. And true, there is no agreed account of what makes such normative sensitivity possible. But the challenge thereby raised for naturalistic approaches is not specific to the area of our discussion and need not concern us here.

What, however, are the prospects of the agent-centred account meeting the normative constraint? Suppose that an agent is disposed to track relevant standards and so has the capacity, in the agent-centred sense, always to have done otherwise than he or she did. Does this make it permissible to praise or blame the agent for what they do on any occasion? Does it make it permissible to react to them with gratitude or resentment, appreciation or indignation?

Some may say that this will be permissible, so far as praising or blaming someone serves to shape their behaviour; and that it will serve in this way, so far as people care about being praised or blamed for meeting relevant standards: that is, pursue the good opinion expressed in praise and flee the bad opinion expressed in blame. The idea is that the susceptibility to praise and blame of the person who tracks certain standards—the person who has the capacity, in the agent-centred sense, always to have done otherwise—will make it useful and therefore permissible to subject them to the influence of praise and blame.

This response makes the activity of holding someone responsible counter-intuitively strategic and manipulative (Strawson, 1982). The posture of holding someone responsible for action will be of a kind with our disposition towards the dog—or perhaps the young child—when we expose it to rewards and penalties that are designed to shape its behaviour. If we praised or blamed an agent only because of hoping to reinforce or alter their behaviour in such a manner—and were this a matter of common knowledge, as it inevitably would be—then praising or blaming someone would be a highly disrespectful act and would be a reasonable ground for their resenting us. Praising or blaming a person is intuitively respectful in character, involving an acknowledgement of their agency and autonomy, and whatever makes the activity permissible, it cannot just be people's susceptibility to the shaping effects of praise and blame. The basis of permissibility must be more subject-friendly than that.

This observation provides the cue for what I think is the right answer to the question before us. For the uniquely subject-friendly basis on which it might be permissible to praise or blame an agent is that the agent gave his or her permission for this to happen. And it turns out that we can identify such a basis of permissibility under an intuitive development of the agent-centred account.

Suppose that agents represent themselves to others as tracking certain standards, where this representation is overt: it is a matter of common knowledge among the parties involved that this is what is happening. If agents do this, then it is equally going to be a matter of common knowledge that others will expect them to meet those standards and may act out of reliance on their doing so; we would not give credence to the agents' representation of themselves as tracking standards, after all, if they expressed surprise at others' reacting in this way. But if agents overtly represent themselves to others as tracking certain standards, despite its being a matter of common knowledge that others will therefore form and act on corresponding expectations, then they presumably acquiesce—again, as a matter of common knowledge—in others holding them to those expectations; this must be a matter of presumption so long as they do not reject reciprocity and community with the others involved. They license others to feel aggrieved about any failure on their part to satisfy the standards, and they license them to feel gratified by any success. They give others permission to blame them for failure and to praise them for success; they invite those responses by the way in which they represent themselves.

None of this should be surprising. In overtly representing themselves as tracking the standards involved, the agents will not just have reported their attachment to those standards, as if this were a matter of merely idle interest. They will have avowed and committed themselves to those standards as criteria by reference to which others are permitted to assess them and are permitted to react, as appropriate, with gratitude or complaint, praise or blame (Bilgrami, 1998). In putting themselves forward as committed to the standards in question, they will have accepted the right of others to judge them on the basis of those standards; to react negatively or positively, depending on the quality of their performance, and indeed to expect this reaction to have some effect.

The agent-centred account of the capacity to have done otherwise equates it with an ability to track certain standards. That ability is consistent with the naturalistic constraint on a satisfactory account, as we saw earlier, and it now begins to seem that it is consistent also with the normative constraint. The ability of agents to track certain standards will make it permissible to praise or blame them for what they do in the presence of that ability, provided that they avow those standards that they track. If this proviso is fulfilled then it will be permissible to praise or blame agents for what they do in the presence of that ability because the agents will themselves have licensed or permitted that reaction.

But will the avowal proviso have to be fulfilled in those cases where we think it is permissible to praise or blame an agent, in particular an agent who

could have done otherwise in the agent-centred sense? Or will it be possible for the agent to escape responsibility just by cancelling that avowal? I argue that there are standards relevant in all these cases such that it will not be possible for the agent to cancel the avowal.

We would not think it permissible to hold agents responsible unless we saw them as potential, conversable interlocutors: unless we saw them as persons with whom we could in principle reason about how things stand and about what should be done. Agents who lay beyond the reach of reason and discourse in this sense would be at best like mute animals and would not present themselves as the sorts of creatures it makes sense to hold responsible. It is striking in this connection, after all, that we do not praise or blame mute animals in any serious sense: we do not react to them, or at least not if we are being sensible, with attitudes of resentment or gratitude. But if we regard an agent as conversable in the required sense then we must believe that they are disposed to track certain standards and indeed to avow those standards. So at any rate I argue.

The standards that agents must be disposed to track, on pain of not really counting as conversable subjects, are those standards that people must fulfil under intuitively favourable circumstances, if they are to count as having intentional states like perceptions, beliefs and desires and if they are to count as expressing and enacting those states in the manner of a creature with whom we can reason. Other things being equal, conversable subjects must be disposed to form the belief that p on perceiving that p ; to form the belief that q on coming to believe that p , where they already believe that if p then q ; to form and act on the intention to X on coming to believe that by X -ing they can bring it about that r , where they already desire that r ; and so on. In particular they must be disposed to respond in such ways by virtue of recognising that those responses are supported by relevant reasons, since otherwise they would not count as creatures with whom we could reason (Pettit, 1993a: ch. 2; McGeer and Pettit, 2001). Not only are they disposed to form the belief that p on perceiving that p , as a mute animal might do; not only are they disposed to be triggered into believing that p by the perception that p . They are disposed to recognise that the perceptual evidence supports the belief that p and, other things being equal, to be led by that recognition to form the belief that p . And so on in the other cases (McDowell, 1996).

So much by way of illustrating the standards that people must be disposed to track, on pain of not really counting as conversable subjects. But conversable subjects have to be disposed to do more than just track such standards. They also have to be disposed to avow them as standards to which they can be held.

When someone engages us in conversation, they have to put themselves forward as worthy of being addressed and worthy of being heard; there is no point in talking or listening to the wall. And this means that they have to avow those standards that any worthy interlocutor—any conversable subject—must generally be expected to satisfy. The interlocutor who proved indifferent to criteria of inductive evidence, logical consistency, or argumentative coherence would soon lose any hold on us. And this being a matter that is knowable in common to all, any interlocutor who aspires to connect with us—to reach our minds—must avow such standards as criteria on the basis of which we are entitled to assess and respond to their performance (Pettit, 2001; Pettit and Smith, 1996).

We will treat subjects as conversable only so far as we think that they have the capacity to connect with us, or at least to connect with some others, in this way. And so we must treat such subjects as being disposed, not just to track the standards in question, but also to avow them as standards they embrace. It follows then that with any subjects that we take to be conversable—as we must take anyone whom we praise or blame to be conversable—we have to think that they are disposed both to track and to avow the standards of reason illustrated earlier. We must assume, not just that they will generally adjust as the standards of reason require, but that in any discursive engagement with others they will avow those standards as guidelines by which they can permissibly be judged.

When we assume in this way that conversable subjects are disposed to avow standards of reason, what we assume is quite substantive. It is not just that such subjects are disposed in discursive exchanges with others to avow those standards as guidelines by which they can be judged in the course of such exchanges. Rather it is that they are disposed to avow those standards as guidelines by which they can be judged in the course of any performance, whether in discursive exchange with others or in non-discursive contexts. An interlocutor would have little claim on being treated as someone worth talking to if we thought that while he or she could do quite well within the confines of an exchange, they were not generally a creature of reason. This being so, the interlocutor who lays claim to being taken seriously has got to avow standards of reason as standards by which they can generally be judged. And if we take a subject to be conversable, therefore, we must take them to be disposed to avow standards of reason in that general way.

The upshot of this line of argument is that the avowal proviso that I mentioned above is going to be reliably fulfilled with agents that are fit to be praised or blamed. Such agents will have to be conversable. Conversable agents will have to be disposed not just to track but also to avow the

standards of reason illustrated. And being disposed to avow those standards, they will be equally disposed to give us permission to react to their performance—their performance generally—with feelings of resentment and gratification, blame and praise. They cannot put aside the disposition to avow the standards and so they cannot withdraw the disposition to permit such reactions.

There is a difference, of course, between a person's avowing standards of reason and actually giving us a licence to react with resentment or gratification and their just being disposed to do those things; but the difference does not make for a problem with the argument. For suppose that we react on the basis of the agent's disposition alone, manifesting resentment or gratification, blame or praise. The agent will not be in a position to deny our title to react in that way, since they will have to admit that they were disposed to license such reactions and that this was indeed something that we were able—and that they were able to see that we were able—to discern. We may have been a little presumptuous in reacting as we did but we were not presumptuous in a degree that the agent could seriously condemn.

CONCLUSION

That an agent could have done otherwise in any action means, under the agent-centred account on offer, that the agent acted in the presence of a capacity to track standards of reason, in particular standards of reason that he or she is disposed to avow. And that an agent could have done otherwise in this sense, so it transpires, explains why it is permissible to praise or blame them for what they did. In particular, it explains why this is permissible without forcing us to regard them as creatures with capacities that transcend the resources of the natural world. The agent-centred account meets both the naturalistic and the normative constraints introduced at the beginning.

The position defended offers an ontologically distinctive account of what makes true the claim that the agent could have done otherwise in a given choice. The truth-maker is not a discrete and punctual feature of the process in which the choice was generated. It does not reduce to the sort of thing postulated in those incompatibilist and compatibilist theories discussed in the first section: not to a distinctive sort of agent-causation, for example, and not to a desire that might have been otherwise than it was. It consists rather in the nature of the agent at the time of action: in the fact that he or she was possessed at that moment of a capacity to track those demands of reason that are avowed in any discursive relationship.

This feature of the approach has an interesting implication. Consider two agents who each act under the immediate influence of certain beliefs and desires, performing more or less identical actions in more or less identical situations. According to the agent-centred account defended here, it may still be true that one agent could have done otherwise and the other not. The difference will consist in a difference in their general make-up, not in a difference in the specific aetiologies of their actions. It will consist in the fact that the one agent had the capacity to track the demands of reason and the other not, though that capacity may have played no causal role in generating the first agent's behaviour.

One question, in conclusion. If an agent had the capacity to do otherwise in a given choice, as that capacity is interpreted here, does this mean that the agent enjoyed the fullest degree of freedom in making that choice? It certainly means that the agent had the basic capacity to track the avowed demands of reason; the agent was *compos mentis* and faced a choice in which the demands of reason were relevant. But freedom in the full sense imposes two other sorts of requirements that may or may not have been satisfied in the case on hand (Pettit, 2001).

First, if an agent is to count as fully free in a certain choice then their capacity to track the avowed demands of reason must not have been reduced in any measure—that is, rendered difficult but not impossible of exercise—by psychological factors like obsessiveness, compulsion, fatigue and the like. And, second, if an agent is to count as fully free in the choice then the demands of reason that the agent had the capacity to track must not have been primed or rigged by others—at least not primed in the negative manner associated with coercion—so as to lead the agent in a particular direction; things must not have been fixed against the agent's will, for example, so that he or she faces a threat of punishment in the event of taking a certain action. Freedom in a basic sense may be assured by the presence of a capacity to track the demands of reason, as that has been elucidated here, but freedom will exist in full measure only where the capacity is unreduced and the demands unrigged.

REFERENCES

- Ayer, A.J. (1982) "Freedom and Necessity" in G. Watson (ed.), *Free Will* (Oxford: Oxford University Press).
 Bilgrami, A. (1998) "Self-knowledge and Resentment" in B.S.C. Wright and C. Macdonald (eds), *Knowing Our Own Minds* (Oxford: Oxford University Press), 207–41.

- Chisholm, R.M. (1982) "Human Freedom and the Self" in G. Watson (ed.), *Free Will*, 24–35.
 Davidson, D. (1980) *Essays on Actions and Events* (Oxford: Oxford University Press).
 Fischer, J.M. (1999) "Recent Work on Moral Responsibility", *Ethics* 110, 93–139.
 Frankfurt, H.G. (1988) *The Importance of What We Care About* (Cambridge: Cambridge University Press).
 Honoré, T. (1999) *Responsibility and Fault* (Oxford: Hart Publishing).
 Jackson, F. (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Oxford University Press).
 McDowell, J. (1996) *Mind and World* (Cambridge, Mass.: Harvard University Press).
 McGeer, V. and Pettit, P. (2001) "The Self-regulating Mind", *Language and Communication* 21.
 Moore, G.E. (1911) *Ethics* (Oxford: Oxford University Press).
 Otsuka, M. (1998) "Incompatibilism and the Avoidability of Blame", *Ethics* 108, 685–701.
 Pettit, P. (1993a) *The Common Mind: An Essay on Psychology, Society and Politics* (pb edn 1996, New York: Oxford University Press).
 — (1993b) "A Definition of Physicalism", *Analysis* 53, 213–23.
 — (2001) *A Theory of Freedom: From the Psychology to the Politics of Agency* (Cambridge: Polity Press).
 — and Smith, M. (1996) "Freedom in Belief and Desire", *Journal of Philosophy* 93, 429–49.
 Strawson, P. (1982) "Freedom and Resentment" in G. Watson (ed.), *Free Will*.
 Van Inwagen, P. (1983) *An Essay on Free Will* (Oxford: Oxford University Press).
 Watson, G. (1982) "Free Agency" in G. Watson (ed.), *Free Will*.