STEPHEN L. DARWALL

AGENT-CENTERED RESTRICTIONS FROM THE INSIDE OUT

(Received 6 January, 1986)

In what follows I shall be concerned with what might be called "the problem of agent-centered restrictions." Briefly put: How can any restriction on what a person may do to promote the best states of affairs that concerns an act's relation to *himself* (for example, that it would be a breaking by *him* of *his* promise, or that it would be a harming by *him* of *another*) possibly be justified? I shall argue that while any case for agent-centered restrictions remains elusive as long as ethics is approached in one, quite common way, there is another approach, one that begins with the idea of a responsible moral individual, that makes agent-centered restrictions intelligible.

I begin in Section I by sketching a line of thought that puts enormous pressure on agent-centered restrictions in order to set the problem. In Section II I quickly review the earlier history of this problem in Moore, Broad, and Ross. In Section III I discuss a current version of the problem as posed by Samuel Scheffler, and in Section IV I show how some preliminary attempts to advert to individual responsibility to solve it fail. In Section V I show how deeply a theory of right without agent-centered restrictions conflicts with a widely held view about an individual's responsibility for his own moral integrity.

Finally, in Section VI, I sketch an approach to ethical theory, one that can be found in Butler and Kant, that takes this view of responsibility and integrity as fundamental. Apparently no justification for agent-centered restrictions can be found so long as we begin by looking *outside* the moral agent – whether to states of affairs that acts bring about or to the nature of acts themselves considered independently of motivation. If we approach ethics from the outside we are led to consequentialism as a theory of right, unless, like Ross, we simply assert certain agent-centered restrictions as fundamental and underived. The alternative I sketch approaches the theory of right from the direction of

Philosophical Studies **50** (1986) 291–319.

an account of responsible moral character. It works, as it were, from the inside-out. My thesis is that this approach is much likelier to provide a rationale for agent-centered restrictions in its theory of right.

I

There is a way of thinking about ethics that makes consequentialism a very appealing position. We begin by thinking of actions as the initiating of changes in the world; actions have consequences. Some of these consequences are causal, but not all are. There is also a sense in which the state of affairs consisting of an act's performance is a consequence of the act: had the act not been performed, the state would not have obtained.

What a person ought to do, what it would be right for her to do, is the best thing she can do. If she does something else she does something worse, and surely there is more justification for doing what is better. We should do the best we can.

One act is better than another if, and only if, the states of affairs brought into existence by the former are better, on the whole, than the states produced by the latter. This may be because the former act has consequences, in the narrow sense, that are better than the latter; its causal consequences may be better. But that will be insufficient by itself. The latter act may partly constitute a state of affairs that is intrinsically better, and its greater intrinsic value may outbalance the greater value of the effects of the first. So if the consequences of one act are better, on the whole, than those of another, then the total value it produces, both the intrinsic value of states it partly constitutes and that of further states it causes, must be greater than that of the other. An act is the best thing one can do just in case the value of its consequences, construed broadly, is greatest. So an act is right just in case it has the best consequences.

Three things should be noted about this line of thought. First, its conclusion is that consequentialism *in the broad sense* is the correct theory of right. Consequentialism has perhaps more usually been associated with theories of the good, such as hedonism, according to which an act's being performed has no value in itself. On these views, there is nothing intrinsically bad about disloyalty, say, or good about

distributing resources equally, or in accordance with merit. There is nothing intrinsically good or bad about any act. Acts, and states of affairs that consist in their performance, can have only extrinsic value.

But there is nothing in the intuitive idea that a right act brings about the best states that requires this view. A keeping of a promise brings about a promise's being kept. If that is a good thing in itself, then its value must be reckoned into a consequentialist calculus. To the extent that consequantialism as a whole has been rejected because consequentialist theories in the field have had implausibly narrow theories of good, that rejection must be rethought. What may appear to be arguments against consequentialism, and for deontology, may actually be arguments for a more sophisticated consequentialism.

But what, then, is the difference between deontology and consequentialism? If every objection to consequentialism can be absorbed by suitable changes in its theory of good, then is there any remaining difference? Does the intuitive line of thought just sketched construe consequences so broadly that the issue is lost? The second thing to notice is that it does not.

Consequentialism holds that an agent ought to do what will bring about the best states of affairs. The requisite value of a state of affairs is fundamentally independent of any relation to the agent – it is 'agent-neutral'. Even if the valuable state of affairs essentially includes an action, its value is independent of being *the agent's* action – of being *his*. For example, if *S*'s keeping his promise is intrinsically valuable, it is so independently of its being *his* keeping of *his* promise.

Consequentialism is an agent-neutral theory of right. Deontological theories are not agent-neutral; they often include principles that are agent-centered.[1] For example, a deontological theory might include a *prima facie* duty to keep promises. This is different from treating promisekeeping as intrinsically valuable. A *prima facie* injunction to keep promises is a *prima facie* injunction to keep *one's* promises, not to bring about the intrinsically good state of affairs of people keeping their promises.

If the intuitive line of thought sketched at the beginning tells in favor of consequentialism, it tells against any agent-centered theory. If it is right to keep my promise, then keeping my promise must be the best thing I can do. If it is the best thing I can do, then the state of affairs

of my keeping it, together with its further consequences, is best. But whether that *state* is best in the requisite sense depends in no way on its being *my* keeping *my* promise. There can be, consequently, no agent-centered *prima-facie* duty to keep *one's* promises. At best there might be a *prima facie* duty to promote promisekeeping, if that state of affairs is intrinsically good.

Now it might be thought unreasonably demanding to hold that it is always wrong to do what will have less than optimal consequences. It is one thing to say that the best act brings about the best states of affairs, but quite another to say that anything less than the best is wrong.[2] Thinking along these lines may lead one to conclude with Samuel Scheffler that it is often not wrong for a person to do what is less than optimific when her own projects and commitments would be sufficiently sacrified by the optimific act. Scheffler proposes an 'agent-centered prerogative' according to which a person is permitted to pursue her own projects out of proportion to their agent-neutral value.[3]

Still, even if it is not always wrong not to do what will have the best consequences, it may still never be wrong to do what will. A weaker version of the initial line of thought goes in this direction. Even if it is not always wrong not to do the best act, it can never be wrong to do the best act. After all, how could the best thing one could do be wrong for one to do? If an act is best, then the state of affairs of its being performed is best. If it is never wrong to do what is best, then it is never wrong to bring about the best states.

This, then, is the third thing to notice about the line of thought. While some sort of agent-centered prerogative or permission is compatible with a weaker version, no agent-centered *restriction* or prohibition can be. If the weaker version is correct, it can never be wrong to do what will bring about the best states. If so, there can be no requirement that is agent-centered (say, to keep *one's* promises, or not to harm *others*) that it is wrong to violate even when doing so would produce better consequences.

The plausibility of this line of thought poses an important challenge to deontological theories of right. It is a very common way of thinking about ethics, and it lies behind, I believe, a recent resurgence of support for consequentialist theories.[4] On the one hand, it blunts past

criticisms of consequentialism by showing how many can be absorbed within the consequentialist framework. On the other hand, it apparently shows why what is truly distinctive about a deontological theory, agent-centered restrictions, cannot possibly be justified.

The most trenchant version of the challenge can be found in Scheffler's *The Rejection of Consequentialism*. This is ironic in a way since Scheffler's aim is partly to reject consequentialism – or, as he more cautiously puts it, to argue that a rationale exists for an agent-centered prerogative that is independent of consequentialist considerations. He argues that a case for an agent-centered prerogative can be mounted on the basis of two considerations: one, that our motivations naturally arise from our own personal points of view, and, two, that a theory of right that directly 'reflects' this with an agent-centered prerogative is a 'rational response' to that fact.

Scheffler's challenge to deontology is that the justification for an agent-centered prerogative provides no rationale for any agent-centered restriction (the 'independence thesis'), and, moreover, that there is reason to think that no justification of any kind can be given for agent-centered restrictions (the 'asymmetry thesis'). He canvasses various proposed justifications and concludes that they all fail.

II

The debate between consequentialism and deontology has been continuous in moral philosophy, in some form or other, since the eighteenth century British moralists. But it is, I believe, only in this century that it has been cast in terms of agent-neutrality versus agent-centeredness in the theory of right. The first place I know this dialectic to have arisen is in connection with the consequentialism, or as he called it "ideal utilitarianism," advanced by G. E. Moore in *Principia Ethica*. Moore's consequentialism is of the sophisticated variety; Ross called it "the culmination of all the attempts to base rightness on productivity of some sort of result."[5] Because of its pluralistic theory of good, Moore's theory of right resists some deontological criticisms of simpler consequentialisms. And, more important for contemporary consequentialists, it shows the resources that a sophisticated con-

sequentialism has available. Moreover, Moore articulated, perhaps more clearly than anyone, the underlying rationale for consequentialism.

Moore's general argument against deontology, as well as his famous "refutation" of egoism, are versions of the intuitive line of thought with which we began. Both egoism and deontology are agent-centered theories. Egoism maintains that each ought to advance his own happiness, and deontology includes duties that are agent-centered. Moore's argument against each was the same. If a person *ought* to do something, then it must be good, indeed *best*, that he so act. The act must promote something with intrinsic value. Perhaps it is the act itself, perhaps some further consequence of the act. In either case, if that state of affairs is good, there will be the same reason for others to produce it as there is for the agent. So, against egoism: there can only be a reason for someone to advance her happiness if her being happy, or acting for the sake of it, is good absolutely. But if that is so, then others will have the same reason to promote that state of affairs, and she will have the same reason to act similarly with respect to others. And against deontology: there can only be a reason for a person to fulfill some (apparently agent-centered) duty if her doing so is good, indeed best. But if it is good that she do so, then others will have the same reason to promote that state of affairs, as will she to promote their so acting. So there is no *agent-centered* duty. At best there are intrinsically good acts the performance of which every agent has a reason to promote.

> It is plain that when we assert that a certain action is our absolute duty, we are asserting that the performance of that action at that time is unique in respect of value. But ... its value cannot be unique in the sense that it has more intrinsic value than anything else in the world; since *every* act of duty would then be the *best* thing in the world, which is ... a contradiction. It can, therefore, be unique only in the sense that the whole world will be better, if it be performed, than if any possible alternative were taken.[6]

Although Moore did not put his argument in terms of agent-centeredness and agent-neutrality, C. D. Broad later noted that this was its thrust. He characterized what he called Moore's "ethical neutralism" in this way:

> Ethical neutralism assumes that there is a certain *one* state of affairs – "the sole good" – at which *everyone* ought to aim as an *ultimate* end. Differences in the proximate ends of different persons can be justified only in so far as the one ultimate end is best secured in

practice by different persons aiming, not directly at it, but at different proximate ends of a more limited kind.[7]

Actually, since sophisticated consequentialism has a pluralistic theory of good, it would be more correct to say that it provides one *ranking* of states of affairs, and that agents ought to do whatever will bring about the best states of affairs as determined by that ranking.

Broad went on to point out that Moore's argument against egoism, and he could have added, his argument against agent-centered deontology, depended on this commitment to ethical neutralism. But was ethical neutralism a premise or a conclusion for Moore? In *Principia* it is more natural to think of it as a conclusion derived from a deeper premise: namely, that if an act is a duty, if it is right to perform it, then that must be because performing it is "unique in value." If an act is right, then it is best. And "in asserting that the act is *the* best thing to do, we assert that it together with its consequences presents a greater sum of intrinsic value than any possible alternative."[8] For the Moore of *Principia*, there is only one fundamental ethical notion, the *good*, or intrinsic value, and the right can be defined in terms of it.

Ross also noticed the agent-neutrality of Moorean consequentialism, though not in so many words, and pointed out the sharp contrast with his own deontological theory of *prima facie* duties. And recognizing what led Moore in this direction, he steadfastly refused to follow.

Against Moore, Ross argued that the agent's specific context, his relations to others, the history of his past acts and of others' acts towards him, and his special relationship to himself, are all directly relevant to what it would be right for him to do. Moore's theory, he argued,

seems to simplify unduly our relations to our fellows. It says, in effect, that the only moral ly significant relation in which my neighbours stand to me is that of being possible beneficiaries by my action. They do stand in this relation to me, and this relation is morally significant. But they may also stand to me in the relation of promisee to promiser, of creditor to debtor, of wife to husband, of child to parent, of friend to friend, of fellow countryman to fellow countryman, and the like.[9]

Strictly speaking, Ross's criticism is a bit wide of the mark. For while Moore called himself an "ideal utilitarian," he did not hold that an act is right only if it maximizes total net *benefit*. A right act maximizes

intrinsic value. And Moore held that friendship, or at any rate, personal affection, is among the things that have intrinsic value. So he held that *that* relation does have moral significance. And he *could* have held, and still remained a consequentialist, that all of the relations Ross mentions have intrinsic moral significance. That is, the flourishing of each of these relations might be held to have intrinsic value and to be worth promoting for its own sake.

Ross's criticism becomes clearer when we read the rest of his last sentence.

> and each of these relations is the foundation of a *prima facie* duty, which is more or less incumbent on me according to the circumstances of the case.[10]

Ross's view was not that since these relationships are intrinsically valuable there is moral reason for every person to promote them. Rather, he held that the fact a person is *himself* related to others in various ways creates *prima facie* duties, to care for his children, to be loyal to his friends, to keep his promises, and so on. Moore's view was that *no* relation was relevant in an agent-centered way.

But how did Ross resist the line of thought that led to Moore's neutralist consequentialism – the line from right act, to best available act, to act productive of the most intrinsic value? Ross maintained that "'right" does not stand for a form of value at all."[11] Moore's mistake was to suppose that an act's being the right thing to do *just is* its being productive of the most intrinsic value. As against Moore, Ross argued that the concept of right is no less fundamental to ethics and irreducible than Moore had argued that of intrinsic value to be. Once he had opened a logical space between claims about intrinsic value and claims about what it is right or wrong *to do*, Ross was in a position simply to assert the common sense position that agent-centered characteristics of acts can be right- or wrong-making – that, for example, its being a betrayal of *one's* close friend is directly relevant to whether an act would be right or wrong.

When sophisticated consequentialists objected that a pluralistic theory of right, with no unifying rationale, was arbitrary and unmotivated, Ross replied that he was in no worse a position with respect to the objection than those who generally made it, since they also held pluralistic theories, allbeit of the good. This reply is especially strong

when made to a sophisticated consequentialist whose theory of the good itself seems formulated expressly to meet deontological criticisms of simpler versions.

III

That agent-centered restrictions have the support of common sense is generally not in dispute. If there is a burden to be carried at the level of considered judgments about specific cases it certainly belongs to the neutral consequentialist. The "problem" of agent-centered restrictions is that there is no apparent rationale for them. The intuitions that support them remain, as Scheffler has put it, "intuition[s] in search of a foundation." (112)

It is at the level of deeper justification that consequentialism appears to be in a stronger position. At least we can identify an intuitive line of thought that underlies it. Like Ross, the deontologist may choose to reject this line of thought. He may urge that an act's being right and the state of affairs of its being performed being best are different things. Being right, he may say, is not a form of value. And he may insist, as did Ross, that there are agent-centered *prima facie* duties. But even if he can defend his position, he may be unable to say what is deeply appealing about it. That is "the problem."

Scheffler puts the problem in this way. He considers a specimen agent-centered restriction, R, against harming innocent others. He then asks us to

> suppose that if Agent A_1 fails to violate ... R by harming some undeserving person P_1, then five other agents, A_2 ... A_6, will each violate restriction R by identically harming five other persons, P_2 ... P_6 who are just as undeserving as P_1, and whom it would be just as undesirable from an impersonal standpoint to have harmed. (84)

What, he asks, could be the rationale for holding it to be wrong for A_1 to violate R in such a case?

Now it might seem implausible that the debate between consequentialism and deontology should come down to this question. The situation seems contrived, and it may be difficult to see what hangs on it. But there is a point to the question.

The point, to a first approximation, is that unless R has the feature that it is wrong to violate it even if doing so would bring about fewer

violations, then *R* can be fully captured within an agent-neutral consequentialism that holds acting contrary to *R* to be intrinsically disvaluable. If it is wrong for one to violate *R* even though that would lead to fewer violations of *R*, then *R* is inconsistent even with a neutral consequentialism that holds violations of *R* to be intrinsically bad.

Two things about the question deserve further comment, however. First, as Scheffler certainly realizes, an agent-centered deontological theory need not be committed to absolutism. That is, it can include agent-centered restrictions that are *prima facie* and that are overridden by other considerations, both agent-centered and agent-neutral. So Ross held, for example, that there is a *prima facie* duty to keep one's promise, but other *prima facie* duties as well that can conflict with it, both agent-centered duties, such as those on one's family, community, and so on, and agent-neutral duties such as the general duty of beneficence.

The point is that there is nothing magic about the number five in Scheffler's question. *R* might be an agent-centered restriction even if it would not be wrong to violate it to prevent five violations. What matters is that the wrongness of violating *R* not be reducible to the *disvalue* of its being violated. This could be true even if it would not be wrong to violate it to prevent four violations; or three. In fact, it could be true even if it would not be wrong to violate it to prevent *two* violations. This is so because its being justifiable to violate it in that case need not consist in the violation's producing more value.

This is the second thing to notice. It would be sufficient for *R* to be an agent-centered restriction, the wrongness of violating which is irreducible to the agent-neutral intrinsic disvalue of its being violated, if it would be wrong to violate *R* when doing so would promote greater, or equal, value. But this could be true at the same time that it would not be wrong to violate *R* to prevent two violations. To see this suppose that a violation of *R* would prevent *one* other violation of *R*. In this case an agent-neutral consequentialism will hold that, other things equal, there is nothing to choose between abiding by *R oneself*, thereby bringing it about that another person violates *R*, and violating *R* oneself, thereby preventing the other from violating *R*. Each violation would be an equally bad occurrence. From the point of view of an agent-neutral consequentialism it simply does not matter whether the

intrinsically bad consequence is one's violation of R or another person's. If a theory holds that that does make a difference, that it matters to what one should do whether it will be one or someone else violating R, then R will be an agent-centered restriction. So it is not necessary for R to be an agent-centered restriction that it be wrong to violate it even to prevent two violations. It is sufficient that it be wrong to violate it to prevent one exactly similar violation by someone else.

That said, I intend to make nothing hang on it. It does seem to make the job of justifying agent-centered restrictions less onerous, but the fundamental problem still remains: if the state of affairs of someone's violating an agent-centered restriction would be better, why would it be wrong for her to violate it?

Scheffler's challenge is that while a justification for an agent-centered prerogative can be identified in the "independence of the personal point of view," none can be identified for any agent-centered restrictions. An agent-centered prerogative does not conflict with the intuitive idea that it cannot be wrong to perform an act when so acting would be part of the best state of affairs that could occur.

IV

As I indicated earlier, my proposal will be that a rationale for some form of agent-centered restrictions is likelier to emerge if we approach ethics from the point of view of individual moral responsibility. In some form or other this suggestion is not new, and Scheffler explicitly considers a version of it. It is instructive to see why various versions are nonstarters.

Bernard Williams pointed out over a decade ago that consequentialism includes a doctrine of *negative responsibility*:

> if I am ever responsible for anything, then I must be just as much responsible for things that I allow or fail to prevent, as I am for things that I myself, in the more everyday restricted sense, bring about.[12]

But, he also noted, common sense recognizes an important difference between consequences that would not have occurred if the agent had acted differently, but whose occurence is the direct result of some *other* person's action, and direct consequences of the agent's own acts. On a neutral consequentialism, however, all that matters for what a given

agent, S, should do in some circumstance are the values of V_i, associated with each possible act A_i, in the conditional: If S had done A_i, then states of affairs with value V_i would have obtained. It is simply irrelevant whether the causal chain goes through other agents' acts.

There is a sense, then, in which on a neutral consequentialism, one is as responsible for bad consequences of others' acts one could have prevented, but did not, as one is for bad consequences resulting directly from acts of one's own. It simply follows from agent-neutrality that whether consequences result directly from *the agent's* act is irrelevant to its being right or wrong. Of course, whether a person should be *held* equally responsible for indirect as for direct consequences will be a different question for the consequentialist. That will depend on the consequences of further acts involved in holding people responsible.

Scheffler considers an attempt to motivate agent-centered restrictions by pointing to the common sense idea that one is responsible for the direct effects of one's acts in a way that one simply is not for the effects of the acts of others that one could have prevented. His response is quite reasonable: there is no question that the doctrine of negative responsibility is implausible at the level of common sense, but that is not the issue. The issue is whether there exists some deeper rationale for rejecting it. So far the assertion that people are more responsible for the direct consequences of their acts is no deeper than the assertion that they have no similar duty to prevent the bad consequences that would directly result from the acts of others. To assert the former is virtually to assert the latter; it does not justify it.

A second strategy might be to argue that neutral consequentialism is inconsistent with respect for persons as independent responsible moral agents. Since it holds the consequences of others' acts to be relevant to the rightness of a given act to the extent that the latter can affect the former, neutral consequentialism appears simply to 'look through' or disregard the moral agency of any person other than the agent whose act is being evaluated as right or wrong. To the extent that moral agents internalize a neutral consequentialism they will then have a way of regarding others that might be thought morally pernicious. Consequentialism apparently requires that one simply *assume* responsibility for others in an obnoxious way. Even God is thought to do no wrong in leaving us free to act in ways that have ill effects. True, one

would not want an ethic to recommend simple quiescence in the face of evil potentially resulting from the acts of others. But the other extreme, that a person regard preventable bad consequences of the acts of others as warranting intervention in every case in which it would be warranted were the consequences the direct result of her own acts, seems unpalatable also. That seems inconsistent with respect for others as independent moral agents.

There is, however, a sort of respect for autonomy that a neutral consequentialism can recognize. A sophisticated consequentialist can hold that autonomy is intrinsically valuable – to respect it is to promote it. If so, neutral consequentialism can hold interference with others' agency to be intrinsically disvaluable, other things equal, thereby avoiding the unpalatable extreme without an agent-centered restriction.

There is justification for an agent-centered restriction only if there is justification not simply for weighing in the intrinsic disvalue of interference, but for a restriction the violation of which is not warranted by an increase in value even when the intrinsic value of autonomy is taken into account. And the intrinsic obnoxiousness of regarding others simply as the conduit of one's own agency does not evidently provide any justification for that. If neutral consequentialism is to be rejected because it conflicts with respect for others as independent responsible agents, therefore, deeper considerations will have to be marshalled.

Finally, one might try to argue against what Scheffler calls the 'independence thesis', against, that is, his claim that the rationale he provides for an agent-centered prerogative does not also justify any agent-centered restriction. The prerogative, recall, permits agents to devote energy and attention to their own projects and commitments out of proportion to the (objective) value of their doing so. And its rationale according to Scheffler is that most of our projects and commitments naturally develop from within our own personal points of view. We become committed to particular pursuits, people, communities, and so on, as a result of our own individual personal histories. Scheffler argues that a theory of right should directly reflect that fact with a prerogative. The best that a sophisticated consequentialist can do is a 'consequentialist dispensation' that gives intrinsic value to person's pursuit of their own commitments and projects, but

that does not permit them to pursue them when they could promote more self-realization by others by sacrificing their own projects.

When one considers what an ethical theory with an agent-centered prerogative, so justified, but without any agent-centered restrictions, would look like, the result may seem unstable. How can there be a justification for a prerogative, but none for a restriction on interference with its exercise? Won't the theory both permit agent *A* to pursue a nonoptimific personal commitment but require agent *B* to prevent *A* from doing so if that would be optimific, assuming that forbearing interference is not covered by *B*'s prerogative? The idea of a prerogative suggests the idea of a morally protected sphere of personal action, but without an accompanying restriction on the acts of others, the sphere will not be protected against morally sanctioned interference.

The situation is analogous to that considered just above. A neutral consequentialist can treat autonomy, or self-realization, as intrinsically valuable. This value, then, can be weighed in determining whether interference with another's exercise of his prerogative would have the best consequences. Because the value is agent-neutral, however, there will be no case for failing to interfere with *A* if doing so is necessary to prevent yet greater interference with others. But even if the independence of the personal standpoint justifies an agent-centered prerogative, and not merely a consequentialist dispensation, it is hard to see why it justifies an agent-centered restriction on interference rather than a neutral consequentialism that gives intrinsic value to self-realization and intrinsic disvalue to interference. There is at least some plausibility to the view that the importance of the personal standpoint provides a rationale for persons having some freedom to pursue their own personal projects even when their doing so is at some cost to general self-realization. But when we shift from the agent exercizing the prerogative to others, the importance of the personal standpoint provides no apparent rationale for restricting *others* from interfering when doing so would promote greater self-realization. An *agent-centered* restriction on interference seems unmotivated.

V

It seems, then, that no rationale for agent-centered restrictions emerges

in any simple and direct way from considerations of responsibility, respect for others as responsible agents, or the independence of the personal standpoint. There is a way of conceiving of each of these within a fundamental rationale that leads to neutral consequentialism and away from any agent-centered restriction. If there is something about responsible moral agency that provides a justification for agent-centered restrictions it will apparently have to be framed within a wholly different line of thought.

In the next section I will sketch a fundamental approach to ethics on which agent-centered restrictions are, as such, unproblematic, and contrast it with the line of thought leading to consequentialism. The latter begins with a view about the intrinsic value of states of affairs conceived independently of any moral evaluation of conduct or character, while the point of departure of the alternative I shall suggest is a fundamental view of character, moral integrity, and of responsibilities relating to these. To put it in a rough and preliminary way, moral integrity involves a person's guiding his life by his own moral judgment, properly understood, and the fundamental responsibility of the moral life is the maintenance of integrity, so conceived. Instead of beginning outside the moral agent with a view about states of affairs that are intrinsically worthy of promotion, the alternative begins inside the moral agent with a view about moral character and integrity. The rationale for agent-centered restrictions is itself agent-centered.

To prepare the way for a discussion of this approach I want first to consider how consequentialists are bound to view the proper relation of integrity and character to what a person ought to do.

To begin with, because on a neutralist view the history of a person's own conduct is not directly relevant to what she should do, there is a sense in which a person bears no direct responsibility for what she has done. Her *own* past conduct leaves no directly relevant trace in determining what she should subsequently do, since were it to do so it would have to be *via* an agent-centered restriction. Neutral consequentialism thus rejects any special duty to try to comprehend, understand, or come to grips with, one's own past conduct, and by doing so to repair moral integrity. Of course, a neutralist can explain why we should do this on many occasions, so that we will be better able to maximize intrinsic value. But we have no special responsibility for our past in the

sense that what we should do is intrinsically unaffected by what *we* have done.

Neutral consequentialism does hold that a person has a special responsibility for her acts at the time of their performance, that she does not have for the acts of others, in at least one sense. A theory of right action *just is* a theory of what a person is responsible for *doing* given what, at the time of action, she has it in her power to do. To act contrary to the theory is to do wrong and, in this sense, to fail to discharge one's moral responsibility.

But consequentialism denies that a person has a special responsibility for her character or integrity in the sense that it denies that considerations regarding *her* character and integrity are in any way directly relevant to what she should do. It denies that the consequences of acts for her character are any more relevant in themselves to what she should do than are consequences for the character of others. It denies that an act's constituting a diminishing of her moral integrity, or a violation of her own principles and values, is any more intrinsically relevant to what she should do because it is her own moral integrity that is at stake. And it denies that a person has any but a contingently instrumental obligation to take thought of what she has done and is doing in her life, to "bear [her] own survey," in Hume's phrase, and conduct her life in a way of which she can on honest reflection approve.

A vivid example will be helpful. In a recent essay Tomas E. Hill, Jr. describes "an artist of genius and originality" who "paints a masterwork unappreciated by his contempararies," but who "cynically, for money and social status," and with some self-disgust, "alters the painting to please the tasteless public and then turns out copies in machine-like fashion."[13] Hill argues that there is a well understood sense in which the artist fails to respect himself: he fails to "live by a set of personal standards by which [he] is prepared to judge [himself]."[14]

Suppose, however, that the story continues. There is another similarly talented artist who is bent on pursuing the same path, but the spectacle of the first artist so sickens him that he decides he cannot do it, and does not. So the consequence of the first artist's conduct is the loss of his integrity, but the prevention of the loss of the other's. A neutral consequentialism will hold that it makes no difference to what

the first artist should have done that it violated his integrity. A loss of integrity is a loss of integrity. Other things equal, there was no moral reason for him not to sell out that did not also exist for him to prevent the other's selling out.

Two clarificatory remarks are in order at this point. Though neutral consequentialism is indeed committed to these counterintuitive propositions about what it is right to do, the neutralist may respond that we find these propositions counterintuitive partly because we run together matters of right and wrong with matters of praise and blame, evaluations of acts and evaluations of agents. That the first artist does no worse wrong in violating his own integrity than he would in failing to prevent another from violating his, does not mean that *he* should be judged the same in both cases. Evaluations of acts as right or wrong is a wholly different matter from evaluating persons. For various reasons, it could be argued, lack of self-respect is a worse trait of character than is unwillingness to prevent another's loss of self-respect if it requires losing one's own respect.

Also, the neutralist insists, we must distinguish between subjective and objective rightness – between which act is right given what would actually have happened, and what act would have been right judged relatively to what the agent believed or could reasonably have believed, that is, on the assumption that those beliefs were true. The first artist's act may have been, objectively, no worse than his keeping his moral integrity intact if that would in fact have led to the other's compromising of himself. Nonetheless, if he was ignorant of this consequence his act was subjectively wrong.

For reasons that will become apparent in the next section, I am skeptical that evaluations of acts and agents should be kept separate in the way the consequentialist insists. The point is not that we cannot distinguish, at least in many cases, between what act a person should perform, regardless of motive, and how an agent is to be appraised for performing it from some particular motive, some particular set of beliefs, and so on. My point will be that one can approach the theory of right, in a general way, from a view of moral character.

Second, concern about personal integrity may lead one to think, along with Scheffler, that what consequentialism requires is simply an agent-centered prerogative that protects action for such ends. But the

sort of integrity with which I am concerned is not Williams' identification of a person with his 'ground projects', alienation from which is threatened by neutral consequentialism.[15] My concern is with *moral* integrity, a person's responsibility to live by principles he can reflectively accept, and, consequently, not to do what is wrong by his own lights. Here a prerogative will be insufficient.

VI

The line of thought leading to consequentialism begins, as I said, outside the moral agent with a view about the intrinsic value of states of affairs. It then works its way inside, first with a theory of right action, and then with a theory of moral character. Acts are right if they maximize the value of states of affairs. A character trait is good if inculcating it maximizes valuable states or, perhaps, if praising it does so. In this progression of external to internal, acts are the natural midpoint. They are the effect of internal causes or, less committally, the output of creatures with a certain internal constitution. But they are individuated independently of their specific internal cause or motive. And, for the consequentialist, their signal feature is that they are part of an objective external order; they partly constitute and bring about states of affairs. So acts have both an external and an internal aspect.

We may say, then, that the consequentialist approaches moral theory from the outside-in. From some basic premises about intrinsically valuable states of affairs, he builds both his theories of conduct and of character.

Now because, on this approach, both conduct and character are evaluated by their respective relation to valuable states, there is an important sense in which consequentialist appraisals of them are instrumental. Conduct is right if it brings about the best states of affairs. A trait is part of good character if it reliably produces the best states.

This may seem to be blunted by the sophisticated consequentialist's holding that the performance of an action, or the having or expressing of a character trait, may be good in itself; but that is only partly true. While a sophisticated consequentialism can hold acts and traits to be

part of, and not simply means to, intrinsically good states, it will hold that a person should perform such an act, or that such a trait is a virtuous one, only if they bring about states with the most value overall. Thus a given act held to be intrinsically good will only be something one should do, or an intrinsically good trait be part of good moral character, if there is no other act or trait available that would produce even more value.

The point is really the same as the "problem of agent-centered restrictions." Even if we think of a character trait as good in itself, there will be a rationale for a person's having it only if that will bring about the most valuable states. If her having some quite contrary trait, even one held by the consequentialist to be an essential part of a state that is bad in itself, would promote greater value, say, if it would promote more people having the intrinsically good trait, then the "evil" trait will be the one the person should have, and she will be a better person for having it.

Approaching ethics from the outside-in forces one to treat moral character as derivative and instrumental. And that suggests a different approach. What I shall call the Butler/Kant view turns the line of thought leading to consequentialism on its head. It begins not with a view about the value of states of affairs but with a very general theory of moral character. It then proceeds to work toward a theory of conduct from its theory of character.

It is, I think, significant that both Butler and Kant held deontological normative positions. Kant, of course, is the paradigmatic deontologist. But Butler may seem harder to peg since, like many eighteenth century British moralists, he rarely addressed the question of what to do considered independently of motive. He did not have a theory of right properly so called. Nonetheless, in arguing that the virtues cannot simply be resolved into benevolence he anticipated what were to become stock objections to consequentialism: that "fraud," "violence," and "treachery" can be wrong, even when their overall consequences are good, and "fidelity" and "strict justice" right though their overall consequences be bad.[16]

Had they been faced with the categories of agent-centered and agent-neutral there is little doubt that both Butler and Kant would have accepted agent-centered restrictions and rejected any wholly

agent-neutral theory of right. So much is familiar and uncontroversial. What is less appreciated is that these philosophers shared a fundamental approach to moral philosophy, one based on a conception of moral integrity and character, that offers hope of a rationale for agent-centered restrictions.

Very roughly put, the notion that is common to Butler and Kant is that to be subject to morality is to have a complex moral capacity, the having of which creates a fundamental responsibility to lead one's life in a way that exercises it. Exercising this capacity, moreover, is both essential to good character and constitutive of moral integrity.

The common notion, therefore, is of a sort of competence that is constitutive of character and integrity and which there is a fundamental responsibility to exercise. Thus on this view there is a link between character and right conduct that is not derivative from their respective relations to some third thing, in particular to states of affairs held to be intrinsically good. Persons ought to conduct themselves in ways necessary to maintain their moral integrity.

The requisite competence is a complex of capacities: (a) to be aware, not only of situations confronting one, but also of the sorts of motives or reasons, ('maxims', in Kant's term, 'principles', in Butler's) that might move one to act in them, (b) to reflect in a certain way, and from a certain point of view, on the idea of *a person's* acting on a given reason or principle in a kind of situation (for Kant, by considering whether one could will that everyone act on the reason), (c) to take an attitude toward acting-on-that-reason-in-that-sort-of-situation on the basis of the appropriate reflection, a reflective attitude or choice that constitutes a *judgment* of so acting, and (d) to regulate one's own conduct by that judgment.[17]

For both Butler and Kant, the person of good character is one who guides his life by exercising the complex competence necessary to be subject to moral demands, a competence, more or less, for independent moral judgment. Only beings with this capacity, Butler argued, can be moral agents in the strictest sense, and by virtue of it all moral agents are "a law unto themselves."[18]

Kant, of course, held that conduct expresses good character only if it issues from the agent's own sense of what she should do. Otherwise, no matter how intrinsically "amiable" the motive of a person's act is, it

will, seen from the agent's own point of view, lead her to do what she should do only "if fortunate enough to hit on something beneficial or right."[19] It will not express moral self-government.[20]

The motivation for the Butler/Kant view of moral character is not, as on the outside-in line of thought, that having it leads to intrinsically valuable states of affairs. The inside-out approach is compatible with, indeed congenial to, a profound skepticism that states can have the sort of intrinsic value they must be able to have on the outside-in approach. A theory of conduct can be justified from the outside-in only if states can have an intrinsic worth-bringing-aboutness that not only creates a *prima facie* justification for any moral agent to promote them regardless of his specific motivational and affective susceptibilities. It must also provide justification for thinking it *prima facie wrong* for him to fail to promote it.[21]

The inside-out view of character is motivated, rather, by the thought that this is what character must be if a person can be responsible for her own moral integrity simply by virtue of having the power to constitute it. Thus Butler: "[W]e are agents. Our constitution is put in our own power. We are charged with it; and therefore are accountable for any disorder or violation of it".[22]

The Butler/Kant approach is agent-centered at the outset. It begins with the idea that each person is responsible for her own moral integrity. But how is agent-centeredness at this level likely to be translated into a theory of right? There are, I think, reasons to expect agent-centeredness of at least two different kinds in a theory of right justified from the inside-out along Butler/Kant lines.

First, because it begins with the proposition that agents bear a responsibility for their own moral integrity that they do not for that of others, it will follow that persons have a duty not to compromise their own moral integrity that they do not have to do what would prevent others from compromising theirs. From the outside-in a loss of moral integrity is a loss of moral integrity. But from the inside-out it is the agent's own moral integrity that is his fundamental responsibility.

Now this duty, though fundamental on the inside-out view, is second-order. An agent violates his moral integrity by doing things he would authentically judge wrong. Ordinarily, however, we think of a theory of right as addressed to the level of the agent's first order

thoughts. True as this doubtless is, an inside-out view must hold there to be a genuine second-order duty not to do what one honestly thinks wrong using one's own best judgment, even, indeed, when one's first-order judgment is mistaken. Acting contrary to one's best judgment threatens moral integrity even if the first-order judgment is mistaken.

But is there any reason to think that approaching a theory of right in the Butler/Kant way, from the inside-out, will lead to agent-centered restrictions at the first-order level? There is, in fact, a very interesting reason for expecting that it would.

If we approach the theory of conduct from the outside-in then we think we have a rationale for evaluating acts by their relation to valuable states of affairs. What matters to us is which states of affairs would actually be brought about by the act.

If we approach the theory of conduct from the inside-out, however, our focus will rather be on the principles, considerations, or reasons that persons should be *guided by* in their deliberations about and choice of acts. Whether an act is right will depend on whether it is recommended by principles or considerations that would weigh with a person of good character.

Consequentialists are at pains to distinguish between criteria of right and wrong and considerations that should be taken account of in deliberation and choice. Their theory concerns the former and not the latter. In fact, they are often quick to point out that while the theory of right is agent-neutral, a consequentialist theory of decisionmaking may well dictate that persons take account of agent-centered considerations in deciding what to do. The best consequences may be produced only *indirectly*, that is, if persons guide their choices not by a neutral consequentialist theory of right, but by other considerations, perhaps by agent-centered ones.

A particularly good recent example of this position is advanced by Derek Parfit. After arguing against what he calls Common-Sense Morality as a theory of right because of its agent-centeredness, he then says that nonetheless "for most of us, the best *dispositions* would in the following sense roughly *correspond* to Common-Sense Morality. We should often be strongly disposed to act in the ways that this morality requires."[23]

But what is important in evaluating a consideration's status as right-

or wrong-making on the inside-out view *just is* whether persons should be guided by it in making their choices, or more precisely, whether a person of good character would be guided by it. The inside-out view refuses to make the sharp distinction between criteria of right and choice-guiding considerations. Indeed, it is worth asking what the force of the consequentialist's assertion that an act was wrong *is*, over and above its simply meaning that it produced less than optimal states of affairs, when he simultaneously asserts that considerations by which the person should have been guided recommended against the act and that the person was a better person for being so guided. That is, what is the force of asserting that not only did the act have less than optimal consequences, but also that, because of that, it was the wrong thing for the agent to have done?[24]

Even consequentialists agree that the considerations a person should be guided by likely include agent-centered restrictions. The "problem" of agent-centered restrictions does not arise for them at this level. The problem concerns whether, even though it is better that agents be guided by agent-centered restrictions, they do what is right when they are so guided. It arises here because if the rationale for a view of what a person should do is to be found in the intrinsic value of states of affairs, then it seems natural to conclude that what a person should do is whatever would bring about the best states.

There are, of course, broadly consequentialist positions that deny that an act's being right depends in any simple way on the value of its consequences. Rule-consequentialists, for example, would roughly agree that since persons should be guided by agent-centered restrictions, then they act rightly when so guided. They are thus likely to endorse agent-centered restrictions. But if the rationale offered for "indirectly" consequentialist normative positions is an outside-in one, if it is argued that inculcating the relevant agent-centered rules and motives will maximize valuable states, then the "problem" reemerges. What is directly at issue in a theory of conduct is what a person should *do*, and not, directly anyway, how people should be motivated or guided in choice. So if the rationale adopted is outside-in, then a neutralist act-consequentialism seems better justified than any indirect consequentialist view.

It is open, of course, for someone who pursues the outside-in

approach to define a concept of right in the way Mill did, as connected to rule-governed practices of approbation and disapprobation. If the concept is so defined, then there may well be an outside-in rationale for a rule-consequentialist account of right that will include agent-centered restrictions. But with any such definition it will still be possible to raise the further question whether a person should do what it would be right to do so defined. And if the fundamental rationale for holding a position on the latter question is outside-in, then any agent-centered response will seem problematic.

The inside-out approach does not face the problem of agent-centered restrictions in the same way. While it is similar to indirect consequentialisms in holding that what a person should do depends on what considerations and principles a person of good character would be guided by, it differs from the latter, at least when the latter is grounded in an outside-in rationale, by not basing its theory of character on a more fundamental view of objectively valuable states. Consequently the relation it asserts between principles a person of good character would be guided by and the rightness of acts is not liable to be undermined in the way indirectly consequentialisms are when their alleged support is outside-in.

The Butler/Kant approach advances a fundamental theory of character that is independent of any view of the intrinsic value of states of affairs. But if its formalist, or as I prefer to say, constitutionalist theory of character enables it to avoid self-undermining of the sort that threatens indirect consequentialisms derived from the outside-in, this very aspect seems to raise other serious problems, problems that I can no more than mention here.

Quite apart from the plausibility of its account of character and moral integrity on the one hand, and of its claim of a fundamental responsibility to maintain integrity on the other, there is a serious question whether any rationale can be mounted from these for any specific theory of right, in particular for a theory of right with specific agent-centered restrictions. The problem is that if what is fundamental is a more or less formal or procedural ideal of moral judgment, together with the proposition that no person is bound by a principle unless she could in principle approve herself of or legislate it from a certain standpoint in a way that satisfies the procedural constraints, what reason is

there to think that any principles, much less any agent-centered ones, will on this basis be binding on all?

Butler and Kant, of course, thought universal principles could be so grounded, but when we consider why they did we may be less confident. Butler seems to have rested his case on a common human nature, created by a God who, by making us so that "there are certain ... actions, which are themselves approved or disapproved by mankind, abstracted from the consideration of their tendency to the happiness or misery of the world", "may have laid us under particular obligations."[25] So, he thought, the existence of a general obligation of "fidelity," rests on fidelity's being universally approved, other things equal, and infidelity disapproved; or at least, on these judgments being universal when informed and reflectively considered in the appropriate way.

Kant's case for universal principles of duty grounded in an ideal of moral judgment, on the other hand, depends at least in part on the unpromising idea that an act is wrong if its maxim cannot consistently be conceived to hold as a universal law.[26] But if the very existence of a practice of promising (Kant's example), is vulnerable to violations in such a way that it is simply impossible for everyone to make false promises whenever it would be to their advantage to make them, so also might the existence of some thoroughly repugnant practice (such as Rawls's example of "telishment") be vulnerable to universal departures under some similar condition.[27] It might be, for example, that individual officials find telishing innocents a burden they would often like to escape, and that if they all did so when it was to their advantage then it would be impossible for anyone to telish because the practice would collapse. But this hardly seems to provide any justification for thinking it wrong not to telish.

Ignoring the "contradiction in thought" test in the Categorical Imperative, however, requires one to emphasize Kant's test of universal legislation in the *will*. The relevant question then becomes whether one could rationally will, perhaps from a standpoint that is impartial between persons (the "kingdom of ends") that everyone be guided by a given principle. But what reason is there to expect universal agreement on principles here?

Unlike Butler's, Kant's case for universal principles of right rests on

no controversial theses about a common human nature that could be expected to lead to universal agreement in reflectively informed and impartial attitude. But because it lacks this common basis the question arises why there is any determinate answer to the question, What principles would it be rational to choose persons be guided by when that choice is made from a standpoint that is impartial between them?

The best hope for the Kantian project, it seems to me, is to pursue it in something like the way Rawls attempts in *A Theory of Justice.* Impartiality is modeled by a veil of ignorance and the basis for choice from this standpoint is then the agent's own interests as a rational and moral person.[28]

For this approach to provide a rationale for any principle of right, two things will have to be true. First, there must be interests that rational and moral persons have as such, relative to which a choice of principles behind a veil of ignorance can be more or less rational.[29] And second, it must be the case that relative to those interests there are principles it is rational to choose from behind a veil of ignorance. Both of these assumptions are far from trivial, but it does not seem unlikely to me that there are agent-centered principles it would be rational to choose from this standpoint.

A second problem concerns the relation of such principles, if there be any, to moral integrity. Even if a specific principle is one it would in fact be rational to choose all to act on, it may nonetheless be one that a given individual's conscientious judgment conflicts with. The Kantian can presumably rule out cases where a person simply believes something is a principle of right but has not herself genuinely embraced (legislated) it in the appropriate way. But what if she takes up the appropriate standpoint, or comes as close as can reasonably be expected, and embraces a principle that conflicts with the one it would be rational for her to choose from that point of view?

When this happens the person will apparently be under conflicting obligations on the inside-out approach. She will have a fundamental obligation to maintain moral integrity and hence not to act contrary to her own authentic moral views. On the other hand, if she does so she will contravene a principle of right grounded in a more adequate exercise of the capacity on which moral integrity depends.

This second problem should not, it seems to me, be viewed as an

unwelcome consequence of the inside-out approach. For surely it is a problem that is central to the moral life and not one we should expect a philosophical account of morality to explain away. Its oddness, if not its sting, may be eased by thinking of principles of right, on the inside-out view, as primarily addressed not to the question of what a person should do when his own moral judgment on some issue is settled, but rather to the question of what his judgment on that issue should be. But if a person has a settled, and authentically gained, view on some matter, it does seem a mistake to think that the question of what he should then do is essentially unchanged.

Agent-centered restrictions seem mysterious or essentially problematic only when moral philosophy is approached from the outside-in. Whatever contribution of disvalue an act makes to the world, however bad it is, no rationale follows from that for refusing to perform it when doing so prevents more performances.

The inside-out approach is not value-based, however. It is integrity- and character-based. If moral philosophy is approached in this way, the 'problem' of agent-centered restrictions dissolves in its outside-in form.

NOTES

[1] Two points should be kept in mind. First, I am assuming a consequentialist/deontological distinction made with respect to the content of a theory of right. A theory is consequentialist if, and only if, it determines whether an act is right by whether the act maximizes good consequences. Otherwise the theory is deontological. So rule-utilitarian would, for present purposes, count as a deontological theory. Consequentialism, in the present context, includes only act-consequentialist theories.

Second, deontological theories may well include principles that are not agent-centered – for example, a *prima facie* principle of general harm prevention. The point is that any such principles could also be part of a consequentialist theory if it had a suitable theory of good.

There is another way of making the consequentialist/deontological distinction, viz., with respect to a theory of right's underlying rationale. On that distinction any theory of right based on propositions about objective or impersonal value – that certain states of affairs are good or bad in a way that creates a reason for any person to promote them – is consequentialist. Deontological theories are those advanced without such a rationale. The "problem of agent-centered restrictions" is whether there exists any other rationale for a theory of right.

[2] This point is made by Judith Lichtenberg in her review of Scheffler's *The Rejection of Consequentialism*, 'The good, the right, and the all right', *Yale Law Journal* 92 (1983),

pp. 531f. See also Michael Slote, *Common-Sense Morality and Consequentialism* (London: Routledge & Kegan Paul, 1985).

[3] Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982), pp. 1–79. Further references to this work will be placed parenthetically in the text.

[4] In addition to Scheffler, see Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984); Peter Railton, 'Alienation, consequentialism, and the demands of morality', *Philosophy and Public Affairs* 13 (1984), 134–171; and Donald Regan, *Cooperative Utilitarianism* (Oxford: Oxford University Press, 1980).

[5] W. D. Ross, *The Right and the Good* (Oxford University Press, 1967), p. 16.

[6] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1966), p. 147.

[7] C. D. Broad, 'Certain features in Moore's ethical doctrines', in *The Philosophy of G. E. Moore*, ed. P. A. Schilpp (La Salle, Ill.: Open Court, 1968), p. 46. Compare Parfit's definition that "agent-relative" principles or theories give "*different* agent *different* aims" (Reasons and Persons, p. 55).

[8] *Principia Ethica*, p. 25.

[9] *The Right and the Good*, p. 19.

[10] *Ibid.*

[11] *Ibid.*, p. 122.

[12] Bernard Williams, 'A critique of utilitarianism', in *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), p. 95.

[13] Thomas E. Hill, Jr., 'Self-respect reconsidered', in *Respect for Persons*, Tulane Studies in Philosophy v. xxxi, ed. O. H. Green (New Orleans, Tulane University, 1982), p. 130.

[14] *Ibid.*, p. 133.

[15] Bernard Williams, 'Persons, character, and morality', in *Moral Luck* (Cambridge: Cambridge University Press, 1981), esp. pp. 12f. See also the section titled 'Integrity' in 'A critique of utilitarianism'.

[16] Joseph Butler, *Five Sermons*, ed. S. L. Darwall (Indianapolis: Hackett, 1983), p. 66n.

[17] Because this view holds there to be no external standard of moral legislation, to put the point in Kantian terms, but only procedural and formal constraints of 'duly constituted' moral judgment, it is illuminating to describe it as *constitutionalist*. ('Constitution' is a central Butlerian notion.) I discuss this aspect of the view in more detail in 'Self-deception, autonomy, and moral constitution', in *The Forms of Self-Deception*, ed. B. McLaughlin and A. Rorty, University of California Press, forthcoming.

[18] The phrase comes from Paul's *Letter to the Romans* (II:14). This passage provides the text for Butler's Sermon II and III. See *Five Sermons*, pp. 34–45 esp. p. 37.

[19] Kant, *Groundwork of the Metaphysics of Morals*, trans. H. J. Paton (New York: Harper & Row, 1964), p. 66; *Preussische Akademie*, p. 398.

[20] In this respect the Butler/Kant approach differs from what is often called an "ethics of virtue", such as Aristotle's, that might also be considered to be inside-out. It seems to me that "constitutionalist" projects such as Butler's and Kant's are purer cases of an inside-out approach since they do not include as essential any particular concern for states outside of the moral agent within their ideal of moral character. They thus differ in this way from a view like Hutcheson's also.

[21] Whether states can be good in any purely objective or impersonal sense that provides any agent a justification to produce them, regardless of the agent's specific nature, is a completely different question than whether there can be facts of the matter about a state's being good for a person or group, or from a person or group's point of view. There might be objective values in this second sense even if there are none in the first.

[22] Butler, *Five Sermons*, p. 15.

[23] *Reasons and Persons*, p. 112.

[24] It is, of course, open to the consequentialist to hold as did Moore in *Principia* that the only fundamental ethical notion is that of intrinsic value.

I should also point out that what I say here does not take into account the possibility that the consequentialist's account of right might be held to coincide with an account of what considerations should guide the deliberations of a perfectly impartial cognizer, like Hare's archangel, though not the deliberations of us less than perfect decisionmakers. This suggestion has promise for some cases but not for others. See Hare, *Moral Thinking* (Oxford: Oxford University Press, 1981).

[25] Butler, *Five Sermons*, p. 66n.

[26] Specifically, he held that the distinction between perfect and imperfect duties is to be explained by the difference between maxims that could not be conceived to hold as a universal law and those that could not be willed so to hold. See *Groundwork, Ak.* p. 424.

[27] In 'Two concepts of rules', *Philosophical Review* 64 (1955), 11–12.

[28] It is often complained that Rawls's original position cannot model Kant's 'realm of ends' because the choice behind the veil is one of instrumental rationality relative to self-interest. But the argument would be essentially unchanged if the parties were assumed to be completely self-sacrificing trustees for the interests (as rational person) of another person. The veil makes it impossible to tailor principles to any particular individual, so by assuming a concern for the rational interests of one person the standpoint effectively expresses a concern for an arbitrary, rather than any particular, person.

I discuss this way of modelling Kant in 'Is there a Kantian interpretation of Rawlsian justice?', in *John Rawls' Theory of Social Justice*, ed. H. G. Blocker and E. Smith (Athens, Ohio: Ohio University Press, 1980), pp. 311–345.

[29] The assumption that there are interests persons have as such plays a more prominent role in Rawls's writings since *A Theory of Justice*. In particular, see 'Kantian constructivism in moral theory', *The Journal of Philosophy* 77 (1980), 525–527; and 'Social unity and primary goods', in *Utilitarianism and Beyond*, ed. A. Sen and B. Williams (Cambridge: Cambridge University Press, 1982), p. 164–165. Thus, from the latter: "In formulating a conception of justice for the basic structure of society, we start by viewing each person as a moral person moved by two highest-order interests, namely, the interests to realise and to exercise the two powers of moral personality. These two powers are the capacity for a sense of right and justice (the capacity to honour fair terms of cooperation), and the capacity to decide upon, to revise and rationally to pursue a conception of the good".

Department of Philosophy,
University of Michigan,
Ann Arbor, MI 48109,
U.S.A.