AGENT-NEUTRAL CONSEQUENTIALISM FROM THE INSIDE-OUT: CONCERN FOR INTEGRITY WITHOUT SELF-INDULGENCE

Is there a justification of concern for one's own integrity that agent-neutral consequentialism cannot explain? In addressing this question, it is important to be clear about what is meant by 'agent-neutral', 'consequentialism', and 'integrity'. Let 'consequentialism' be constituted by the following two theses:

(1) An action is morally right if and only if it produces consequences at least as good as any of the alternative actions available to the agent.

(2) Any action that does not produce consequences that are at least as good as any of the alternative actions available to the agent is wrong.

Agent-neutral consequentialism is the view one gets when one adds that all value is agent-neutral, which is to say that the value of a state of affairs is not relativized to a given agent bearing a specified relation to that state of affairs. In particular, a value is agent-neutral just in case the principle underwriting that value makes no ineliminable and non-trivial pronominal back-reference to a given agent. Whereas a value underwritten by a principle that does involve such back-reference is agent-centered. Henceforth, I shall use 'consequentialism' to refer to agent-neutral consequentialism unless otherwise noted. By 'integrity' I shall mean a person's commitment to following her own considered all-things-considered moral judgments. With these definitions in hand, let us return to the question of whether there is a justification of concern for one's own integrity that agent-neutral consequentialism cannot explain.

A consequentialist can explain why a person should be concerned for her integrity in one of two ways: instrumentally or non-instrumentally. A person's integrity typically will bear on what the consequentialist takes to have non-instrumental value (happiness, say), and therefore will be worthy of at least an instrumental concern. It is, however, open to a consequentialist to hold that a person's integrity has non-instrumental value as well. Notably, these justifications provide each agent just as much reason, in principle, to be concerned about the integrity of others as they provide her to be concerned about her own. In this respect the consequentialist explains why an agent should be concerned with

her own integrity, but the explanation does not make that concern's legitimacy a direct function of the fact that the integrity is *her own*. There may be derivative reasons for the consequentialist to be especially concerned about her own integrity, but those will depend on contingent circumstances. The question is whether there is any justification for a person 's caring about *her own* integrity in some further non-derivative way.

A powerful line of argument for thinking there is such a justification begins with the fact that many of our moral intuitions provide prima facie evidence for thinking there are deontological restrictions, or "side constraints." (I shall here use these terms interchangeably)[1] Roughly, a deontological restriction forbids the performance of an action of type X even when your X-ing is necessary to minimize the total amount of X-ing in the world. So, for example, such a constraint might hold that it would be wrong for me to lie even if my lying is necessary to prevent more lies of the same kind. The acceptance of fundamental deontological restrictions is famously inconsistent with consequentialism, as on the consequentialist account whatever non-instrumental reason an agent has for refraining from X-ing is a direct function of the non-instrumental disvalue of an instance of X-ing. In that case, though, whatever reason there is for me to refrain, in general, from X-ing is equally a reason for me to X where my X-ing is necessary to minimize the total amount of X-ing in the world. The consequentialist has two main strategies for dealing with the intuitions which seem to favor deontology. First, they can cite the good consequences of adopting a practice of not performing certain kinds of actions even when a person strongly suspects that doing so will prevent events with even more disvalue. Second, they can cite other of our considered moral judgments according to which we should, in many "hard cases," perform distasteful actions of the kind the deontologist argues are forbidden. The consequentialist can then urge that there is a tension between some of our moral intuitions and some others, and argue that we should abandon the pro-deontological-restriction judgments insofar as they cannot be explained on a consequentialist model. Often, part of the consequentialist's defense of

this second strategy involves arguing that the moral judgments favoring deontology are "dangling intuitions." The idea is that while the intuitions favoring consequentialism can be given a deeper rationale, judgments favoring deontology can find no such theoretical home.

At this point in the dialectic the deontologist might argue that the consequentialist's account of why a person should care about her own integrity is inadequate. An especially clear and powerful statement of this argument comes from Stephen Darwall, and I shall therefore focus his presentation of the argument.[2] The argument begins by describing two radically different approaches to moral theory. First, we might begin with the value of states of affairs impersonally characterized. In light of the value of those states of affairs, we might then work our way "inward" from the world to our actions and ultimately our integrity, and explain the value of the latter in terms of their contribution to the total amount of impersonal value in the world. This is the "outside-in" approach, and Darwall suggests that it is most well-suited to defending consequentialism. Alternatively, we might begin with the perspective of a conscientious deliberating moral agent's concern for her own integrity, and try to determine what she ought to do in virtue of taking that perspective seriously. We then work our way "outward," building up a conception of the value of various states of affairs out of what an agent must take to be valuable in virtue of taking the perspective of an appropriate concern for her own integrity seriously. This is the "inside-out" approach. On this account, there is a kind of agent-centered concern each agent should have for her own integrity, s*imply because it is her own*. The inside-out approach suggests a possible rationale for deontology: One must not perform an action of a given type even when doing so is necessary to preventing others from performing even more of such actions, *if* performing the action would threaten one's integrity.

The inside-out/outside-in distinction provides a useful frame for thinking about whether there is a justification of concern for one's own integrity that is inaccessible to

consequentialist. I begin by laying out Darwall's argument in more detail (section one). I then distinguish two ways of understanding what is involved in the inside-out approach, which I shall call the "synchronic account" and the "diachronic account." Whereas the synchronic account is one that a consequentialist can and should embrace, the diachronic account commits one to an implausible kind of moral self-indulgence (section two). Finally, I argue that there is a further important sense in which a moral theory might be understood as justified from the inside-out, and that consequentialism can, in principle at least, be defended in this way (section three). This defense would, however, come at a price. Contemporary consequentialists distinguish principles as "decision-procedures" from principles as "standards," and argue that consequentialism is the latter rather than the former. Indeed, this move can seem essential for the consequentialist to avoid objections to the effect that relying on consequentialism as a decision procedure is unlikely to be optimal. However, a defense of consequentialism from the "inside-out" requires conceiving the theory as providing, in the first instance, a decision-procedure.

I.

Darwall highlights three prima facie implausible features of how a consequentialist apparently must think about an agent's concern for her own integrity, and then motivates the inside-out approach by showing how it avoids these features. First, he argues that consequentialism implausibly "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." (Darwall 1986: 305) Insofar as it is important to be concerned about the maintenance of integrity, on a consequentialist view, it seems that an agent should be equally concerned about the integrity of others as she is about her own integrity. Call this the "Backward-Looking Point." Second, he argues that for any given agent, consequentialism, "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." (Darwall 1986: 306) This seems to be out of line with the thought that each person should be especially

concerned about the consequences of her actions for her own integrity as such. Here Darwall motivates this idea with a case, borrowed and then modified from Thomas E. Hill, Jr., of an artist who paints a masterpiece unappreciated by his contemporaries. The artist then cynically, and "for money and social status," and with some self-disgust, modifies the painting to please the "tasteless public and then turns out copies in machine-like fashion." (Hill 1982) Prima facie, such an artist seems to lack self-respect in that he "fails to live by a set of personal standards by which [he] is prepared to judge [himself]." (Hill 1982: 133) Darwall then continues the story and asks us to imagine that this artist's selling out so disgusts another budding artist who had been bent on pursuing the same commercialized path that he decides he cannot do it, and does not. So while a consequence of the first artist's action is that he loses his integrity, another consequence is that it preserves the integrity of another. This suggests that a consequentialist would be forced to admit that it makes no difference to what the first artist did that it violated his integrity, for "a loss of integrity is a loss of integrity." Darwall argues that this contradicts our intuitive sense that the artist should have some special concern not to violate his own integrity. Call this the "Forward-Looking Point." Third, Darwall argues that the consequentialist must deny 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." (Darwall 1986: 306) The idea seems to be that there are deeper reasons for caring about whether one can bear one's own survey. Call this last point the "Reflective Endorsement Point."

Having briefly discussed these three prima facie problems for the consequentialist, Darwall explicitly introduces the distinction between the outside-in approach and the inside-out approach, and suggests that the consequentialist's implicit commitment to the former is what lands her with these three problems. On the outside-in approach, one begins with an account of the non-instrumental value (and disvalue) of various states of affairs, understood apart from any moral evaluation of person's actions

or integrity. One then works from the outside (external states of affairs) in to an account of the rightness of action, and, ultimately, to the evaluation of integrity. By taking this approach, Darwall suggests, the consequentialist inevitably lands with the preceding three difficulties, as that account makes concern for integrity derivative in a way that makes all three of those problems inescapable. On the inside-out approach, by contrast, we begin with the perspective of a moral agent and her concern for her own integrity, and then work our way outwards to an account of the value of states of affairs. Paradigms of this approach are apparently to be found in the work of Bishop Butler and Immanuel Kant.[3]

In addition to avoiding the three problems facing the consequentialist, Darwall argues that the inside-out approach can provide a rationale for agent-centered restrictions. He holds this for two reasons. First, he argues that because the inside-out approach begins with the idea that each agent is responsible for her own integrity in a way that she is not responsible for the integrity of others, that, "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." (Darwall 1986: 311) In this respect, "the rationale for agent-centered restrictions is itself agent-centered." (Darwall 1986: 305) Second, the inside-out approach focuses on the principles that a person should be guided by in her deliberations. In this respect, the inside-out approach differs from the way in which most consequentialists understand their theory. For they typically distinguish between the ultimate standards of right and wrong, on the one hand, and the decision procedures that we should use in our actual deliberations, on the other. In referring to a rule as a "decision procedure," I do not mean to imply that the rule provides anything like an algorithm from descriptive facts to moral conclusions. Such rules most likely will fall short of being algorithms, but may nonetheless provide plausible guides for one's moral deliberation. It might be best to think of moral rules qua decision procedures as more analogous to recipes, which (unlike algorithms) leave much unstated and can require sensitive judgments in light of the situation at hand. With this rough distinction between

standards and decision procedures in place, the consequentialist typically urges that her theory is to be understood as providing a standard, not a decision procedure. As Darwall understands it, though, the inside-out approach, "refuses to make the sharp distinction between criteria of right and choice-guiding considerations." (Darwall 1986: 313) Since even consequentialists typically admit that our actual deliberation should quite often be in terms of non-consequentialist considerations, such as a derivative concern for agent-centered restrictions, this seems to suggest another, more indirect way, in which the inside-out approach can help provide a rationale for such restrictions. Namely, the inside-out approach may rob the consequentialist of a distinction that is essential to defending their view and its associated rejection of agent-centered restrictions at the level of the standard of right and wrong *as opposed to* the level of decision procedure(s).

The crucial issue, then, is whether the inside-out approach provides a plausible account of why a person should be concerned with her own integrity, and thereby provide a rationale for deontological restrictions understood as "agent-centered" restrictions. To resolve this issue, we must first get a clearer picture of just what is involved in the inside-out approach itself. It turns out that there are at least two ways of understanding the inside-out approach. On the first version of the inside-out account, which I shall call the "synchronic version," that account is quite plausible but poses no threat to consequentialism. On the second version of the inside-out account, which I shall call the "diachronic version," the account would, if sound, pose a serious threat to consequentialism. However, I shall argue that this second, diachronic version of the account is subject to a powerful version of the narcissism objection pressed against the agent-centered construal of deontological restrictions in the preceding section.

II.

On a synchronic reading, the inside-out approach simply holds that each of us must, at each point in time, be committed to doing what, in our best judgment, is morally required of us at that point in time. Construed in this way, the inside-out approach borders on

triviality, and is in no way incompatible with consequentialism. After all, the consequentialist can perfectly well say that each of us must, at each point in time, do our best to maximize the good. Nor would Darwall deny that this interpretation is within the reach of a consequentialist, as he remarks that, "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." (Darwall 1986: 306)

This suggests that Darwall has in mind the second interpretation of the inside-out approach. On this second interpretation, an agent's highest level commitment is not merely to doing, at each point in time, what one judges to be morally required of one. In addition to this synchronic requirement, the second interpretation requires that one also refrain from performing actions *now* that will most likely have as a consequence the result that *in the future* one will act wrongly or even become an evil person. It is in this respect that the second interpretation is diachronic. Unlike the synchronic account, the diachronic account really is incompatible with consequentialism, since even a consequentialist who assigns non-instrumental value to integrity will have to admit that if the only way to produce the best consequences is to perform an action that is likely to lead to your becoming evil in the future, then one is required to perform the action in spite of its unpalatable consequences for one's own integrity.[4]

On the diachronic interpretation, the inside-out approach is not so easily absorbed by an agent-neutralist. However, on this reading, the inside-out approach is subject to the that it is implausibly narcissistic. For on the diachronic interpretation, I have a special responsibility for my own integrity not simply because I am the only one who can directly control my actions, but also because it is *mine*. The mere fact that my integrity is *mine*, on this way of thinking, gives me special reason to be concerned about it. Of

course, the consequentialist can admit that the fact that my integrity is mine does indirectly give me special reasons to be concerned about it as opposed to the integrity of others. For one thing, the fact that typically I am most well-situated to know how to preserve my own integrity as opposed to knowing how to preserving someone else's may be relevant. Also, there may be good agent-neutral reasons for me not to pry into other people's business so much that I could become better situated to preserve their own integrity than they are. So the diachronic interpretation of the inside-out approach must involve something stronger than this. The thought must be that the *mere fact* that my integrity is my own, as such, gives me a special reason to be concerned about it as opposed to the integrity of other people.

Put this starkly, as it must be if it is going to be incompatible with agent-neutral consequentialism, the inside-out approach is implausibly narcissistic. Suppose that a moral agent faces a choice between an action that will almost certainly have the long-run consequence of ruining her own integrity, and another action that she knows will just as surely devastate the integrity of five other people.[5] To make things more concrete and vivid, consider the following admittedly fanciful case. Suppose that earlier in life a person has been addicted to gambling, and has now recovered from this addiction. It is also the case that this person knows for certain that there are lots of people in his community who are addicted to gambling, and who, as a result of this addiction are likely to lose their integrity. They will, for example, become less concerned about their families, more willing to do things they know they should not do to get more time in at the casino, etc. However, she also knows that there is a clinic that can help a good number of these people get over their addiction, and that this clinic is strapped for funds. She realizes she could help save the clinic if she had enough money, and she also happens to know that she is in fact a masterful professional gambler. She gave up gambling not because she was losing money, but because of the way her obsessive/compulsive behavior was destroying her integrity. So she could almost

certainly raise tons of money to help the clinic by returning to the casinos, though in the process she would be very likely to slide back into her old patterns of behavior and sacrifice her own integrity, at least for a substantial period of time. Nor does she know of any other way in which she realistically could raise such large sums. In such a case, it seems that she might very well be able to reduce substantially the risk of others losing their integrity only by doing something that puts her at substantial risk of sacrificing his own integrity. Prima facie, and keeping firmly in mind that we are stipulating away any non-negligible epistemic uncertainty about whether the person could win the money and donate it, or as to whether the clinic is effective, this person's returning to the casinos and giving all of the profits to the clinic seems like a morally virtuous, if ironic, sacrifice. However, insofar as it charges each of us with a non-derivative special concern for preserving our own integrity over time (and not merely a duty to do what we conscientiously think is right at each moment in time), the diachronic interpretation of the inside-out approach seems implausibly committed to supposing that in such a case the agent *must not* make this sacrifice. On that account, returning to the casinos is not only not morally virtuous, it is morally forbidden. For to do so would involve knowingly placing oneself at a substantial risk of damaging one's own integrity, against one's apparently highest-level commitment on the diachronic interpretation. Recall that on the diachronic interpretation one's most fundamental commitment is to preserving one's own integrity over time. Hence on that account, I must preserve my own integrity even if doing so ensures that numerous other people's integrity will be devastated. This in itself constitutes an important objection to the diachronic interpretation, for it smacks of a distasteful sort of moral narcissism. To avoid confusion that might allow the intuitive force of the objection to be lost, it is absolutely crucial to be clear that the objection here is that the diachronic version of the inside-out theory incorrectly implies that an agent *must not* make the sacrifice (that is, it implies that it is impermissible to make it), *not* that the theory incorrectly implies that it is permissible not to make the sacrifice.

On the other hand, the inside-out approach might very well allow or even require me to perform an action which will have the consequence that someone else is tempted into doing something that will be likely to ruin her own integrity if this is necessary to prevent a greater number of other people from losing their integrity. For example, it might well allow me to convince someone *else* to return to the casinos if I knew she faced the kind of circumstances just sketched. For in doing so I would not be endangering my *own* integrity, and could in the process preserve the integrity of several other people at the (likely) cost of one *other* person's integrity. In this respect, the inside-out approach in its diachronic guise would license a potentially unappealing asymmetry between risking one's own integrity and convincing someone else to risk their integrity. I do not mean to imply that the inside-out approach *must* hold this; there *may* be plausible ways in which it could deny it without invoking anything ad hoc. Rather the point is simply to indicate that there is a burden on the defender of that approach to show how this result can plausibly be blocked, since nothing essential to the account blocks it. Moreover, insofar as the account allows that, in general, it is at least permissible to do whatever maximizes the good whenever deontological restrictions are not in play it may be hard to block this result, since preserving the integrity of a greater number of people would often maximize the good. Finally, if we are not simply to have solved one "dangling intuition" problem only to replace it with another, any attempt to block this consequence of the inside-out approach must be fit into a larger theoretical rationale, and not simply rely on an ad hoc appeal to first-order intuitions. So far as I can tell, it is simply an open question whether the advocate of the inside-out approach in its diachronic guise can block this result in a principled way.

This possible asymmetry between the diachronic interpretation of the inside-out approach's attitude toward (i) sacrificing one's own integrity to prevent even more people from losing their integrity, on the one hand, and, in effect, (ii) sacrificing someone else's integrity to prevent even more people from losing their integrity, smacks of an

unappealing sort of self-indulgence. On the diachronic interpretation so understood, it seems that while I can do something that will cause one *other* person to be tempted into a life of crime when that is necessary to prevent the temptation of even more others, I apparently *cannot* do something that will tempt myself when this is necessary to prevent the temptation of others. In other words, while I may not lead myself into temptation to prevent the temptation of others, I nonetheless may lead others into temptation to prevent the temptation of even more others. Only an unappealingly self-indulgent concern to preserve one's own integrity over time could underwrite such an attitude. Since, as I indicated above, this asymmetry might be avoidable for the defender of the inside-out approach, I do not rest my objection against that approach on her commitment to that asymmetry. The mere fact that the account precludes an agent from risking her own integrity when this is absolutely necessary to preserve the integrity of others (and the agent knows this for certain) is itself sufficient, in my view, to show that the view falls prey to a serious worry about moral self-indulgence. The possibility of this sort of asymmetry only adds to what I take to be an independently powerful objection.

So on the diachronic interpretation, the inside-out approach falls prey to the objection of inappropriately valuing a distasteful sort of self-indulgence. Nor does it seem that the consequentialist must defend the outside-in approach against the inside-out approach, as Darwall's account might seem to suggest. For on the only apparent interpretation of the inside-out approach that avoids the narcissism objection (the synchronic interpretation), the consequentialist can happily embrace the inside-out approach herself. Perhaps, though, there is a third interpretation of the "inside-out" approach that is worth considering. Before turning to the question of whether the consequentialist can plausibly be seen as adopting an "inside-out" approach in a sense that is both stronger than the trivial synchronic interpretation but that does not suffer from the narcissism endemic to the diachronic interpretation, we should first return to the Darwall three objections to consequentialism.

III.

Darwall argues that the consequentialist's account of why an agent should care about her own integrity is inadequate on three grounds, which I have called the "Backward-Looking Point," the "Forward-Looking Point," and the "Reflective Endorsement Point." In light of the discussion of the diachronic interpretation, we are now in a position to see how a consequentialist might reply to each of these points.

The Backward-Looking Point holds that the consequentialist is stuck with the implausible view that an agent's "*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." (Darwall 1986: 305) In fact, this claim may be too quick, depending on what is meant by 'directly relevant'. For a consequentialist can hold that there is non-instrumental agent-neutral value in people's leading certain kinds of lives (e.g., lives in which they repay debts of gratitude). In that case, though, the fact that I accepted a favor from you in the past, say, could be directly relevant to what I should now do. For given those histories, it directly follows that if I do not act in certain ways that there will be fewer people in the world who are as grateful, say, as there could be. If we take the paradigm of an indirectly relevant backward-looking consideration to be one in which one looks at the past simply to compute what the likely consequences of one's actions will be in the future, then this sort of relevance is at least not indirect in *that* way.

Perhaps, though, a fact's being 'directly relevant', in Darwall's sense involves more than that the fact has implications for what an agent should do now, apart from serving as evidence as to what the future consequences of the person's actions would be. In particular, perhaps the suggestion is that what is directly relevant is not merely how those backward-looking considerations bear on how my actions will reflect on some person's integrity, but how they will reflect on *my* integrity. In other words, this sort of direct relevance presupposes that each agent should be especially concerned about preserving the purity of her own integrity *as such*. In that case, the consequentialist may well have

trouble accommodating the direct relevance of backward-looking considerations. However, if direct relevance is understood in this way, then worries about the self-indulgence of the non-consequentialist alternative surface again, so that the consequentialist's inability to accommodate the point is no longer obviously a vice.

Consider the Forward-Looking Point. Here the suggestion is that consequentialism is inadequate because for any given agent it, "denies that the consequences of acts for her character are any more relevant in themselves to what she should do than are the consequences for the character of others." (Darwall 1986: 306) Here, even more clearly than in the case of the Backward-Looking Point, the objection falls prey to charges of narcissism. For here the point is explicitly couched in terms of whether the consequentialist can make sense of the way in which each moral agent should be *more* concerned with the purity of her own integrity over time than she is about the preservation of the integrity of others.

Lastly, consider the Reflective Endorsement Point. Here Darwall alludes to a famous passage from David Hume, in which he emphasizes the importance of one's being able to "bear one's own survey," and argues that the consequentialist cannot fully accommodate Hume's insight. To assess Darwall's suggestion on this point requires us to be clear about what it was that Hume had in mind. A good deal has been written on the apparently Humean idea of reflective endorsement, and I lack the space here to do that literature justice. Very roughly, Hume's basic idea here seems to be that a normative perspective, such as a moral point of view or the point of view of theoretical reasoning, has genuine normative authority for a person only if the person would, upon reflection, endorse the perspective in question. There are many different ways of developing this basic idea, but one of them that seems to have been especially important for Hume emphasized the need to endorse a normative perspective from that very perspective. Understood in this way, reflective endorsement is best understood as a contribution to moral epistemology. This connects nicely with the inside-out approach, insofar as the

latter approach strongly suggests that one has a highest-order duty to determine one's first-order duty. In which case, one might then need to employ the reflective endorsement methodology to fulfill that highest-order duty.

The contrast with some versions of consequentialism can be non-trivial in this respect. For on some forms of consequentialism, particularly those justified from the "outside-in," one might, in principle, at least, have no duty at all *ever* to determine what duty demands. For example, if for a given person it is true that "going with the flow with no concern for duty" would maximize value, then that person will act rightly in going with the flow and without ever as much as thinking of her duty. On this reading, it is also not hard to see why Kant is thought to be an especially clear representative of the "inside-out" approach, given his emphasis on acting "from duty." Furthermore, it may well be that the inside-out approach, understood in this way, is more well-suited to providing a rationale for deontological restrictions than the outside-in approach. For the apparently paradoxical nature of such restrictions may well seem pressing if we begin from the subjective perspective of the deliberating agent rather than with an account of the objective value (or disvalue) of various states of affairs.

The possibility remains, however, that a consequentialist could mount a plausible argument for her view from the inside-out as well. Granting, in other words, that deontology stands a better chance of being defensible on the inside-out approach than on the outside-in approach, it remains an open question whether the consequentialist might nonetheless be able to defend her view from the inside-out as well. For if this interpretation of the inside-out approach qua reflective endorsement is correct, then there is no obvious reason that a consequentialist could not accept it as a plausible account of moral epistemology.[6] The consequentialist could then argue that when each of us examines our moral faculty, to stick with Hume's now quaint-sounding way of putting it, we find that the faculty garners its own support *and* that the substantive content revealed by our moral faculty is consequentialist in form. In this way, it seems that a

consequentialist might consistently find a respect in which each person's inspecting the most fundamental principles of her own character to see whether she can bear her own survey can acquire a significance that is far from a mere "contingently instrumental obligation." (Darwall 1986: 306) For this exercise would be crucial to each agent's conscientiously carrying out her commitment to morality insofar as it is justifiable. Moreover, insofar as the reflective endorsement approach has each agent begin with the most fundamental principles of her character and then work outward from there to see what she should do, there is a recognizable sense in which this account is "inside-out." So there is an intriguing conception of the inside-out approach that seems to be plausible, non-narcissistic,[7] and accessible to a consequentialist.

Here, however, Darwall's second argument from the inside-out approach to agent-centered restrictions is worth considering. Recall that his first argument was a straightforward appeal to the claim that the inside-out approach entails a duty to be especially concerned with one's own character, and we have seen that this argument is valid only if we interpret the inside-out approach diachronically and so narcissistically. Darwall's second argument for such restrictions is more indirect and more promising, particularly when we conceive of the inside-out approach as essentially identical with the reflective endorsement approach. For the second argument appealed to the way in which the inside-out approach leaves no room for a sharp distinction between standards and decision procedures, while the consequentialist apparently needs to draw some such distinction. This argument is especially forceful in light of the reflective endorsement approach which does seem not immediately to suggest any useful distinction should be drawn between decision procedures and standards. For if the reflective endorsement approach is the whole of our moral epistemology, it seems that the only moral principles we could ever know about are the ones that we suppose are adequate as decision procedures. If there were sound principles that were merely sound qua standards and not sound qua decision procedures, they would on this account be forever beyond our ken.

In fact, there may be room even on the reflective endorsement approach, for a rough and useful distinction between objective standards and the appropriate decision-procedures. For example, if the correct decision-procedure for the conscientious moral agent turns out to be the universal law formulation of the categorical imperative, then there is an obvious candidate for an objective standard: whether the agent's maxim, *in fact*, is universalizable. The fact that the agent has carefully come to the conclusion that her maxim meets the test does not mean that she is correct in so doing. So there is room, even on the reflective endorsement approach for an appearance/reality gap with respect to our judgments of which actions are right, even where our judgments are as careful and conscientious as one could reasonably demand.[8] It is not, then, that there is no useful distinction to be drawn between decision procedures and standards. Rather, it is that we start with decision-procedures and build objective standards out of them, so to speak, instead of the doing things the other way around. In principle, then, there is no reason that the consequentialist might not establish that the correct subjective decision-procedure is for an agent to aim to perform the action that maximizes the good, and then infer that the correct objective standard of right and wrong is the consequentialist one, as defined at the beginning of the present essay.

This may still mean, however, that the consequentialist's employment of the inside-out approach qua reflective endorsement comes at a steep price. For consequentialists often see themselves as going in exactly the opposite direction: first determining the correct objective standard and *then* deducing what the correct decision-procedure would be for a given agent in her actual empirical circumstances from the empirical facts and that standard. If the consequentialist is going to justify her account via the inside-out approach qua reflective endorsement, this approach will need to be reversed. This may make it much more difficult to justify consequentialism, given the worry that under many conditions consequentialism would recommend its own rejection.

Whether this will pose an insurmountable obstacle to a defense of consequentialism from the inside-out will depend upon how stringent we are in our interpretation of what it takes for a principle to pass the reflective endorsement test. If, for example, we require that a decision procedure must be consciously employed whenever the agent faces a number of possible options then the consequentialist account will look quite implausible indeed, and for very familiar reasons. Quite often employing consequentialist reasoning will predictably lead to worse consequences than many alternatives would; often "going with the flow" might produce better results, for example. However, it seems dubious that any plausible principle could satisfy this requirement. Even the categorical imperative will look quite implausible if we suppose that the agent must consciously employ it at every choice point, however trivial, in her life. A considerably more plausible requirement would be that the principle always serve as a regulative one, in that if the agent were considering doing something that were clearly incompatible with the principle then the principle would, so to speak, "kick in" and alert him to the moral status of that course of action. On this account, the principle in question might typically lay in the agent's unconscious, and not enter into his actual deliberations at all. An even more modest requirement would be that to count as a sound decision procedure, a principle must serve as a regulative one in this sense, not all of the time, but simply most of the time. Finally, of course, one might go so far as to require only that the principle serve this regulative function at least some of the time. Certainly these last two conceptions of a sound decision procedure are not obviously beyond the grasp of sophisticated consequentialists, and even the second most demanding constraint, requiring a constantly regulative principle, might be one that they could meet. Indeed, a number of consequentialists have tried to defend their view in just these terms, distinguishing two levels of thinking - one appropriate to everyday thinking and one appropriate to reflective contexts in which one's more everyday ways of making decisions seem not to be the best decision procedures for the case at hand.[9]

Consider again Darwall's (borrowed and modified) case of the struggling artist. At the outset, it is worth admitting that for many this case does produce strong anti-consequentialist intuitions, and perhaps lends substantial prima facie support for an deontology. I do not pretend to settle the debate between deontologists and consequentialists here. Rather, my point is the more modest one that if Darwall is right that such pro-deontology intuitions stand in need of a deeper rationale, then the inside-out approach on the diachronic interpretation could provide such a rationale only at the cost of making the deontologist's position unduly self-indulgent. With that caveat in hand, it is worth distinguishing a couple of variations of the case. In the first variation, which is the most natural reading of Darwall's description, the first artist does not know that his selling-out will prevent another artist from doing the same thing, or if he does know this he does not act on the basis of that consideration. In that case, then, a consequentialist could argue that subjectively speaking the artist acted wrongly, for he may not have correctly deployed the appropriate consequentialist decision procedure. If, however, the rest of the consequences are fixed appropriately, the consequentialist might have to admit that the artist did in fact do the right action from the point of view of the consequentialist's objective standard. Furthermore, if the agent did know all those consequences for certain and sacrificed his own integrity for the sake of another struggling artist, then the consequentialist would have to admit that his action was right both subjectively (he correctly applied the right decision procedure) and objectively (his action is in accordance with the correct objective standard, which is itself "built out of" the correct decision procedure). However, it is not obvious that this is such an implausible result. We must distinguish two ways in which the artist might be tarnishing his integrity. First, he might be doing something that, as a consequence, will lead to his integrity being damaged. In that case, to insist that he nonetheless must not sell-out would smack of narcissism again. On the other hand, though, we might suppose that the action tarnishes his integrity not because of its consequences, but in and of itself, simply

because it is an instance of his doing what he believes to be wrong. No doubt part of the reason the consequentialist account of this case can seem implausible is that we might have the intuition that the kind of action the artist performs is wrong apart from its consequences, simply because it displays a lack of respect for humanity, say. Moreover, in that case it might be that the agent would be violating even the trivially true synchronic version of the inside-out approach, by knowingly doing something wrong to prevent someone else from doing something wrong. Rather clearly, though, a consequentialist will resist this description of the case. In particular, a consequentialist will resist the suggestion that actions of this sort are wrong, all things considered, in themselves (though the consequentialist may happily admit that they are *bad* in and of themselves). For to assume that would be to rely on "dangling" deontological intuitions that are, as yet, in search of a foundation. Perhaps, as was noted above, some such deeper rationale can be found from the inside-out approach; as Darwall emphasizes, Kantian moral theory as developed by Rawls and others might well provide such a foundation. I agree that the inside-out approach is a more promising one for the deontologist than the outside-in approach. My point is not to impugn that project. Rather, my aim is simply to emphasize that there is also room for a defense of consequentialism from the inside-out on any interpretation that is not implausibly self-indulgent.

Conclusion.

The plausibility of the inside-out approach has been thought to suggest that consequentialists lack an adequate account of why an individual moral agent should care about her own integrity. I have argued that there is a crucial ambiguity in the conception of the inside-out approach. If it is understood synchronically, then the consequentialist can accommodate it. If, by contrast, it is understood diachronically, then the consequentialist cannot consistently embrace it, but this is no embarrassment. For in its diachronic form, the inside-out approach is vulnerable to a charge of narcissism and moral self-indulgence. Finally, if we instead understand the inside-out approach as

embodying the Humean idea of reflective endorsement in moral epistemology, then the consequentialist may still be able to embrace that approach. This is not to suggest that the reflective endorsement approach does not face its own share of problems; worries about relativism, for example, may be very pressing once we begin to think more critically about that approach. Indeed, this is not the place even to begin assessing the overall merits of reflective endorsement. For present purposes, the important point is that embracing the inside-out approach conceived as reflective endorsement means that the consequentialist must admit that moral principles qua decision-procedures are more fundamental than moral principles qua standards. This may well create serious problems for the consequentialist for the familiar reason that consequentialism can be self-effacing. Still, depending on just what is involved in internalizing a principle as a decision procedure, these problems may be ones that a sophisticated consequentialist could handle. Defending a "two-level" version of consequentialism, like R.M. Hare's theory, is probably the consequentialist's best strategy here. Moreover, if the consequentialist could defend their view in this way, they might be able to explain the lingering intuition that we should be especially concerned with our own integrity without falling prey to worries about self-indulgence. For reflective endorsement makes concern for one's own integrity important for reasons of moral epistemology that do not seem especially narcissistic. In any event, the possibility of defending consequentialism from the inside-out qua reflective endorsement is worth taking seriously.[10]

WORKS CITED

Blackburn, 1998. *Ruling Passions.* Oxford: Oxford University Press.

Butler, J. 1983. *Butler's Five Sermons*. ed. S.L. Darwall. Indianapolis: Hackett.

Darwall, S. 1980. "Is there a Kantian Interpretation of Rawlsian Justice?" in Blocker and Smith 1980. 311-345.

_____. 1986. "Agent-centered Restrictions From the Inside Out." *Philosophical Studies*. 50. 291-319.

French, P., et. al. (eds.) 1995. *Midwest Studies in Philosophy*. XX.

Green, O.H., ed. 1982. *Respect for Persons*. Tulane Studies in Philosophy. Volume XXXI. New Orleans: Tulane University.

Hare, R.M. 1981. *Moral Thinking: Its Levels, Method, and Point*.

Hill, T.E. 1982. "Self-Respect Reconsidered." in Green 1982.

Kant, I. 1990. *The Groundwork of the Metaphysics of Morals*. Translated into English by Lewis White Beck. New York: Macmillan Publishing Company.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Sidgwick, H. 1907. *The Methods of Ethics*. 7th edition. Chicago: University of Chicago.

---

[1]See Nozick 1974.

[2]See Darwall 1986.

[3]See Butler 1983 and Kant 1990.

[4]That this interpretation provides further resources for arguing against the consequentialist also provides evidence that this is the interpretation that Darwall has in mind, since he took himself to be defending deontology against consequentialism. There is also a good deal of direct textual evidence that Darwall had in mind the diachronic account. He remarks, for example, that, "consequentialism...denies that the consequences of acts for her character are any more relevant in themselves to what she should do than are the consequences for the character of others." (Darwall 1986: 306) In the context it is relatively clear that he means to be contrasting consequentialism with the inside-out approach, and this in turn suggests that he has in mind the diachronic interpretation.

[5]Oddly, Sidgwick seemed to think such cases could never arise; see Sidgwick 1907.

[6]At some points, Darwall seems to deny this, though it is unclear in those contexts whether he has in mind the diachronic interpretation or the reflective endorsement interpretation. See, for example, his suggestion that the outside-in approach is, "*the* line of thought leading to consequentialism," (Darwall 1986: 305) (rather than simply a line of thought leading to that conclusion), and his suggestion that, "the consequentialist approaches moral theory from the outside-in." It must be admitted, though, that such passages are not decisive, as Darwall may only mean to suggest that the outside-in approach is the one that is most well-suited to defending consequentialism and not that it is the only plausible way of doing so. Indeed, at one point he seems quite alive to the possibility, in principle, of a defense of consequentialism from the inside-out, when he qualifies his discussion of consequentialism in passing with the clause, "at least when the latter [indirect consequentialism] is grounded in an outside-in rationale." (Darwall 1986: 314)

[7]It might be argued that it is narcissistic in that the agent focuses so much on the principles of her own character as opposed to what others think, but this objection underestimates the resources of the reflective endorsement approach. For a crucial part of what is required for adequate reflection upon which principles one should embrace might well involve discussion with others who disagree with one. Indeed, it is fairly plausible to suppose that a moral agent who did not give any weight to the dissent of other reasonable moral judges would be behaving dogmatically and unreasonably herself. The point behind the reflective endorsement test is one that is perfectly compatible with a strong commitment to considering the point of view of others in one's reflection. Still, each agent must think for herself, and must subject these various perspectives to her own reflection. I must not, on this account, embrace a principle simply because a majority of *other* reflective agents have adopted if it does not meet with reflective endorsement upon my own conscientious and open-minded consideration of the principle in question. It is in this respect that the

reflective endorsement approach, while not being solipsistic or dogmatic, does still recognizably work from the inside-out.

[8]Darwall makes a very similar point, though not in terms of standards and decision-procedures; see Darwall 1986: 316-317.

[9]R.M. Hare famously defends such a view; see Hare 1981. For a recent defense of such a view, see Blackburn 1998.

[10]Thanks to Thomas E. Hill, Jr., Keith Horton, Sean McKeever, Philip Pettit, and Susan Mendus for useful comments on an earlier draft of this paper.