1. IMPARTIALITY AND PRACTICAL REASON

LEIGH B. KELLEY
UNIVERSITY OF TULSA

Abstract. The paper constitutes a detailed critical commentary on Stephen Darwall's *Impartial Reason* (Ithaca: Cornell University Press, 1983). Its central thesis is that Darwall's attempt to integrate a naturalist theory of substantive reasons for acting with a neo-rationalist derivation of moral requirements from the very concept of practical rationality is faced with insurmountable theoretic problems. The author argues that anyone who would accept a plausible internalist account of reasons, that justificatory reasons for an agent to act are facts which must be capable of motivating that agent under certain conditions, cannot establish on an a priori or rationalist basis claims for the intersubjective validity of reasons or substantive normative requirements of any kind, but rather must acknowledge that such claims are both irreducibly empirical and epistemically risky.

"This is an instructive case. There is neither money nor credit in it, and yet one would wish to tidy it up."
A.C.D.

1. Introduction

Not long ago, Stephen Darwall published an ambitious work on metaethics, *Impartial Reason*. I aim to show that it is somewhat too ambitious. This will require detailed examination of portions of the book. Its arguments are intricate and the issues they raise complex. But here, critical effort is justified, for the problems with Darwall's project are profound; attending to them profoundly important.

A bit of background is in order. Darwall begins by reminding us of a controversy:

Contemporary discussion of rationality in action has produced two widely divergent and contending camps. One takes the view that practical rationality is wholly self-centered. Any reason to act, it holds, can arise only from an agent's own desires, preferences, interests, or good. If to
ignore the welfare of others or considerations of morality is not contrary to their own desires or interests, people may do so without acting contrary to reason in any way.

Members of the opposing camp deny this. They believe that the person who acts without consideration either for others or for morality ignores an important class of reasons. Indeed, many would add, to ignore morality is to ignore reasons that are often, if not always, weightier than self-centered ones. (13)

For Darwall,

... what underlies many of the arguments for self-centered theories of rationality is a deeply attractive view about the relation between reasons and motivation: the internalist view that reasons for a person to act are considerations capable of motivating the person (when considered in the right way). Those who have held self-centered theories have been right, I think, to embrace internalism. (20)

Which side Darwall will ultimately take is clear:

As its title suggests, this book seeks to vindicate the feeling of the moralists that considerations other than self-centered ones are reasons to act, indeed, that moral requirements, suitably understood, provide reasons for any agent that generally override those based on the agent's own individual preferences. It maintains that practical reason is, at its base, impartial rather than self-centered. (17)

In my view, Darwall's central project involves the attempt to integrate two quite distinct theoretic paradigms. In his exposition, the first travels under the banner of "internalism" and is further articulated in an "initial account of rational consideration". The second, which emerges slowly as one proceeds through the text, we might term "neo-rationalism", which in all its versions seeks to derive substantive value judgments as or from "theorems" of practical rationality. (A principal advantage to this form of argumentation is that those who do not accept the substantive conclusions so derived can be charged, not just with judgmental impropriety in matters moral, but with irrationality.)

As we shall see, the type of neo-rationalism Darwall will ultimately embrace is in part Rawlsian in character. In such accounts, the derivation of substantive normative requirements and constraints proceeds by way of arguing the rational necessity of making certain overriding choices and commitments under conditions of radical ignorance. In particular, Darwall will seek to move from the initial (internalist) account of rational consideration by showing "... how certain kinds of facts can be considered only if one makes oneself aware of them from a perspective that abstracts from one's own personal relation to what one considers: what Nagel has called the impersonal standpoint". (21) And the "overriding rationality of right conduct" will in part be revealed through attention to "more formal aspects of practical rationality". (19)

The basic problem with this project is that internalism and neo-rationalism don't mix very well. A proper understanding of the first will
clearly indicate either (1) that there can be no relevant conflict between morality and "self-interest" (properly understood) or (2) that, in order to avoid the first conclusion and at the same time maintain even the semblance of a coherent internalism, one must in the end embrace a deeply unmotivated and implausible doctrine. This we shall call the Doctrine of Poisonous Truth. We shall also see that this unpalatable dilemma flows in part from the attempt to advance contingent substantive evaluative claims as requirements of practical rationality.

Now in the first seven Chapters of Impartial Reason, Darwall is primarily concerned to show that the actual desires and preferences an agent has do not, even in conjunction with a rich set of principles of relative rationality, determine categorically what an agent ought rationally to do nor what valid reasons there are for an agent to act. But he also rightly accepts the basic thesis of internalism that reasons for an agent to perform a certain act are facts or considerations which must be capable of motivating that agent to perform the act when considered by him under certain conditions. Otherwise, we are led either to embrace some form of ethical nonnaturalism, with all its attendant problems, or to give up the view that reasons are normative or guiding, that is, that reasons, by their very nature, have a necessary practical relevance to agents for whom they are reasons.

2. The Initial Account of Rational Consideration

"Reasons to act are facts that motivate us to prefer an act when we give consideration to them in a rational way." (85) In Chapters 8 and 9 Darwall provides "an initial account" of what that way is. My immediate concern is to provide an accurate sketch of that account. Let me say at the outset, however, that the theme or image of a "perspective", "standpoint", or "point of view" also surfaces in this context. It is nicely interwoven in Darwall's discussion of rational consideration, but we shall want to keep a sharp eye on how it is used, for this theoretic device, having later undergone a sea-change, will play a profound role in Darwall's attempt to go beyond the internalist contours of the initial account. "The initial perspective of rational consideration, then, will be the person's own point of view, albeit as informed and dispassionate and as having a life that extends through time." (85)

"Rational consideration of a fact requires, obviously enough, being aware of it on reflection--thinking about it in connection with the act in question." (85) This observation leads to a definition of a "presumptive reason": if p is a fact about A [an action] reflective awareness of which would move S to prefer his doing A (to his not doing A), then p is a (presumptive) reason for S to do A. (86) Why only presumptive? Darwall writes:

Further reflection, say in light of other facts, or perhaps from some other point of view from which we might rationally deliberate, may altogether cancel any initial motivation we had to prefer A. So any reason that p provides is only presumptive, and defeasible. Not only may a reason be overridden by other, weightier reasons (and still hold its weight as a reason), but further reflection may cancel any motivating power it had and convince us that what we took to be a reason was in fact no reason at all. (86)
On Darwall's view, preferences can be of two kinds, intrinsic and extrinsic. Since the rational grounding of the latter depends essentially on that of the former, the problem now is to give an account of the rational consideration of facts that motivate intrinsic preferences. (87) In the first instance, these facts will include features or properties that are "internal" to the object of preference. This, however, raises a question: "How are we rationally to adjudicate between conflicting intrinsic preferences, each of which is supported by reasons?" (89)

The image of a perspective or standpoint is now forcefully deployed:

In Part I, I suggested that in having a desire or preference we, to some extent, experience the world from the perspective of that desire or preference. We are fastened, to some extent and in a variety of ways, on its object. This way of speaking is metaphorical, but it shows that a preference for something is a cluster of dispositional states concerning not merely behavior but also attention and affective response. (90)

Again,

An agent with two conflicting preferences has a problem of integration. . . . in order to remove the conflict and integrate herself, she must decide between them. And to do this she must be capable of assuming a perspective distinct from that internal to each individual preference and of getting the object of both in view. (90)

Rationally to adjudicate the conflict he must bring the reasons for his two individual conflicting preferences together and reflectively consider them at the same time. This enables him to judge whether, all things considered, he prefers to . . . [act on one preference or the other]. (90-1)

Rational consideration of facts supporting our intrinsic preferences may require that we determine which sets of reasons for conflicting preferences are weightier. (94) In determining this, we will determine which preferences are more rationally justified, i.e., which have stronger reasons supporting them. Moreover, we must give the reasons for the relevant preferences "dispassionate consideration". (94) "We may think of a dispassionate judgment between the two preferences as analogous to the idea of an impartial or unprejudiced judgment between two contending claimants." (94) "To allow the one preference to suppress appreciation of the object of the other is to undermine genuine joint consideration of them." (95)

Dispassionate judgment between our intrinsic preferences requires that we represent to ourselves in an imaginatively vivid way what we know to be relevant to their objects. [footnote omitted] Genuine consideration is, the Oxford English Dictionary reminds us, an attentive view or survey. [footnote omitted] This corrects both for errors of perspective (whether temporal, spatial, or some other sort with psychological relevance) and for errors that arise because of
the sheer sweep of a present desire, whether the blindness of infatuation or of rage. (95-6)

So far then, "the point of view from which one rationally adjudicates between two conflicting intrinsic preferences is not internal to either individual preference. It is a reflective standpoint from which one can dispassionately consider facts that support the two preferences". (97) Yet, such consideration still may not be rationally sufficient so long as we restrict it to facts regarding specifically the objects of the conflicting preferences at hand. As Darwall notes, "What one prefers, all things considered, may not be to act on either of the initial conflicting individual preferences but rather to act on a third that arises in the process of rational consideration of facts related to one's situation". (98) In sum,

Rational consideration involves in the limit, then, the ideal of reflection on all relevant considerations. According to our initial account, relevant considerations include those that provide an agent some motivation to prefer an act when she reflects on them in an imaginatively vivid way.

* * *

In each case, the preference is based on reasons not necessarily in the sense that it was itself generated by a consideration of reasons but in the sense that reasons exist for it. In the internalist account this means that the preference would result from the right sort of consideration of facts that are reasons for it. (98-9)

This is the initial account of rational consideration. Let us note some important things about it, including several that Darwall himself brings to our attention. Some questions are in order as well.

First, one of Darwall's cardinal theses is that reasons (and practical rationality) have a motivational and a normative aspect. The initial account reflects both:

The motivational aspect of reasons to act is clear on this account: a fact can only be a reason for someone to act if consideration of it, under certain conditions, would motivate him. But simply being moved by a consideration is not the same thing as judging it to be a reason. Even more clearly, the degree of motivation that a consideration provides is not identical with its weight as a reason. Reasons also have a normative aspect; they are grounds for a person's judgment of what it would be best to do, all things considered, or of what he ought (rationally) to do. The way in which this normative aspect is brought into an internalist account of reasons is through the idea that what establishes some fact as a reason (and its weight) for someone to act is his rational consideration of it resulting in some motivation to prefer the act. The normativity of reasons, according to an internalist account, arises from an ideal of rational consideration and not through the externalist idea that there are simply some facts that have the intrinsic property of being action-guiding and some that do not. (81)
Second, we must also bear in mind another of Darwall's central claims, that the "initial account" of rational consideration is incomplete, apparently not just in some detail or owing to technical insufficiencies in its various formulations, but owing to the fact that it must be supplemented by other essential theoretical elements. "At this point [toward the end of Chapter 8] we have no reason to suppose that rational consideration is limited to this sort of reflective awareness (from the agent's own standpoint), and I shall argue . . . that it is not". (98-9) Darwall will later claim, for example, ". . . that reasons grounded in our initial account are not the only reasons to act and that they cannot systematically override considerations of moral requirement". (233) A basic question, then, is whether there is any significant sense in which it is true that the initial account is incomplete.

3. The Normativity of the Initial Account

It should be obvious that the two central factors invoked in Darwall's account are motivation and the consideration of fact. It is by adding qualifications to the latter and specifying its proper relation to the former that he hopes to explicate an important aspect of normative justification. At least on the initial account, only reasons can justify motivation and action. Of course, that is not the whole story of normative justification, for Darwall also maintains the incompleteness thesis. However, whether or not the latter can be made to seem plausible may significantly depend on whether we have the details of internalism right. For only then can we be sure that we have appreciated its scope and power and so have not been prematurely induced to search for theoretic "supplements" that are unnecessary or ill-advised. In this and the next section, then, I want to raise some further questions about whether Darwall has gotten the details right.

Several times in the exposition of the initial account Darwall invokes the notion of a "point of view" or its equivalent. The touted subject of that exposition is the rational justification of desires and preferences. But as the "point of view" notion is successively deployed, the author's emphasis is almost entirely on how it serves to explain psychological capacities we all admittedly have, for example, the ability to resolve conflicts between two desires, or to come to or develop alternative preferences and abandon both of the original ones we were deliberating about, etc. The focus seems to be on how we integrate and order our various preferences. Darwall tells us very little, however, about the operations which must occur for us to exercise these capacities. We simply assume, according to him, different standpoints. But what has all this to do with the normative justification of preferences, however integrated or ordered? For Darwall does not think that de facto preferences carry much or any justificatory weight on their own. (21, 107)

At one point (90), Darwall suggests that to talk in terms of a "perspective", at least as regards one's "experiencing the world from the perspective of . . . [a] desire or preference", is to speak metaphori­cally. Then what, literally, does this notion involve in the exposition of the initial account in Chapter 8?

At every point, it seems that a standpoint or perspective is definable in terms of a set or range of facts that can be considered in a given actual or hypothetical instance of practical deliberation. The
"standpoint" of a particular preference is constituted by facts about or features internal to its object. The "standpoint" of two or more preferences is likewise constituted by facts about and relations among their objects. The limiting "standpoint" of rational consideration under the initial account is simply (indeed, would have to be) the aggregate of the relevant facts, facts that might affect an agent's motivation at a given time were they to be "dispassionately" considered by him. One assumes these varicus "standpoints" by considering the set or range of facts which define them. So far then, the concept of a "standpoint" is basically eliminable. In the context of Darwall's discussion in Chapter 8, where the initial account is presented, I find nothing more in the concept.

However, perhaps the most remarkable thing about Chapter 8 is that Darwall fails to focus on or even explicitly acknowledge the following rather striking fact: The order of increasing rational justification, or the strength of that justification, of preferences and desires (and ultimately of actions) generally tracks increments in the range of facts considered by the agent and the extent of consideration given them. One's motivation and consequent action are prima facie justified to the extent that reality is directly or indirectly reflected in one's deliberation. That suggests that there is or may be a deep connection between reality and the cognition of it, on the one hand, and, by way of the effect (or non-effect) of the latter on one's motivation, the category of normative justification, on the other.

So far then, Darwall's discussion in Chapter 8 is perfectly consistent with the view that a "standpoint" or "point of view" can be defined in terms of a range of facts considered by an agent under conditions of rational consideration, the broader the range of facts, the more inclusive or encompassing is the perspective, and the stronger is the presumptive justification of the action that that agent would be motivated to perform under those conditions. But in a brief passage in the next Chapter, this way of looking at the matter appears to go by the board:

To this juncture no challenge has yet arisen to a fundamentally self-centered view of practical reason. While the considerations that our initial account certifies as reasons need not be restricted to self-regarding ones, that account appears to make the standpoint of rational consideration the agent's own. According to it, we cannot judge whether a fact is a reason for someone to act unless we are somehow able to assume his point of view and see whether he would be motivated by reflective awareness of it. (112)

Here, Darwall's use of the term 'standpoint' changes significantly. The focus is no longer on the set of facts considered by the agent, but (and what else could it be?) on the actual or hypothetical motivation he would have upon rationally considering them. My "standpoint" is that aspect of myself that I bring to rational consideration which is independent of the cognitive and inferential character of the latter. But in this sense, talk of "points of view" is again metaphorical. What does it mean to "assume another agent's point of view"? Surely not that I somehow bring myself to have the same sort of resultant motivation that he would have upon rationally considering the relevant facts. Why not? Simply because doing that is quite unnecessary to judging that he would have the indicated motivation upon his rationally considering those facts. To
make that judgment, I needn't be similarly motivated, any more than to judge that someone else is in pain, I need also to have a headache, i.e., assume or adopt his "point of view" with respect to pain.4

So what is going on here? This conceptual shift is unheralded in the text. No warning is given that a new conception now marches forth under the banner of "standpoint". Let us dub these two conceptions of a standpoint, respectively, the cognitive standpoint and the motivational standpoint. The former allows us to see how we might resolve conflicts between intrinsic desires by assuming a standpoint external to each, which means jointly considering facts relevant to both. With respect to facts and reality, such cognitive standpoints are additive. A standpoint external to two more limited standpoints is generated simply by the addition of the relevant sets of facts under the operations of rational consideration, and so on, until the limit is reached wherein one's intrinsic desires and their relative ordering are grounded in an adequate conception of all (or at least enough) motivationally relevant reality. What unifies the lesser standpoints is nothing less than the unity of that single reality which provides the subsets of facts defining them.

But is there a corresponding unity and limit regarding motivational standpoints? What could unify our potentially contrary and divergent motivations under rational consideration? After all, reasons are not, in their motivational aspect, necessarily universal. (88) What might a "universal motivation" be? Perhaps, but here I say only perhaps, such a motivation can be delineated if we abstract from our potentially divergent motivations under rational consideration just those factors which could produce that divergence.

4. The Productive Aspect of Reasons

Relative to the initial account, what one ought rationally to do, or the best thing to do, all things considered, is the action best supported by reasons. Also on that account, reasons are facts (considerations) which would motivate (85) or lead (81) one to prefer a certain act when they are considered in a rational way. "In the internalist account this means that the preference would result from the right sort of consideration of facts that are reasons for it." (99, my italics) Darwall is fond of reminding us (for example, in his Chapter 6) that the mere fact that one has an actual desire or preference, even an overriding second-order one somehow derived by summing over the aggregate of first-order desires and preferences, is no sufficient justification for acting on it. Indeed, Darwall says, "In our initial account, the reasons that there are for agents to act are not given by the fact of desire or preference itself but by facts that motivate preference when dispassionately considered". (107, footnote omitted; my italics) Again, desires and preferences without the support of reasons have no inherent normative or justificatory force.

But how are we to understand this notion of "support"? So far, reasons are facts that motivate or result in a preference or desire when they are rationally considered. At one place (and I might add at only one place in the entire book), Darwall seems to recognize an alternative way to understand this concept. "Facts that when we are reflectively aware of them move us to have (or maintain) an intrinsic preference for something are presumptive reasons to prefer it." (87, my italics in
Parentheses) Presumably, when the other elements of rational consideration are added to reflective awareness, for example, "maximal vividness", and so forth, presumptive reasons become genuine full-fledged reasons.

(99) The issue now concerns this concept of "maintenance" and how it contrasts with the parallel concept of "support".

The earlier formulations in which the motivic aspect of reasons is expressed suggest that Darwall views the "support of reasons" as something, at least projectively or hypothetically, productive of justified preferences and desires. But what if an agent already has the desire in question? Must we suppose that he doesn't have it, and then further assume that the rational consideration of facts would lead him to have it? This is perhaps (I say "perhaps" because we are given no relevant explanation in the text) where the idea of reasons serving to maintain the desire comes in. But again, how is this to be understood? Since the agent, we assume, is motivated, partially or overall, to act on his desire (that's just the sort of thing desires are), must the available reasons, or the agent's awareness of them, increase that quantum of motivation? Or again, suppose that the desire begins to wane, sua sponte? Were the agent rationally to consider the relevant facts, would he (and must he) then be motivated to some extent to prevent the desire from waning?

Consider now an agent who has an overriding intrinsic desire for something. Let us suppose that this desire in fact was not generated by his rational consideration of relevant facts, but rather arose in him ontogenetically, owing to biological or other pre-rational factors. (Here you may pick your favorite background theory, sociobiology, Freudian psychology, whatever, to fill out the case.) Clearly, were he to give rational consideration to the facts, such would not produce the desire. He has it. If we assume both that he doesn't have it and that he gives rational consideration to the facts, must that produce the desire? No, for the sort of desire in question might only arise in the subject owing in part to the operation of such non-rational (but not necessarily irrational) factors. (We might even have trouble supposing that he doesn't have the desire in the first place, for then we might have a very different sort of being than the one with which we started.) But if this desire started to wane or disappear, wouldn't it at least be the case that the agent would be moved to maintain it? I don't see why this has to be so, especially if having or satisfying the intrinsic desire in question was not positively related, extrinsically, to the satisfaction of other intrinsic desires. So, with respect to this desire, neither the notion of "support" nor that of "maintenance" seems apt.

But then, is acting on such a desire rationally justified in part or overall, or is it not, on the initial account? This question appears even more important when we realize that in the case of many desires, satisfying them takes time and effort. If the hypothesized overriding desire meets either condition, and if the agent has other intrinsic desires, however weak, that have been or would be generated by the rational consideration of facts, and if finally their satisfaction is even minimally curtailed by satisfaction of the former, then, given the initial account, it seems that there are genuine reasons against acting on the original desire, no matter how strong it is. But suppose that the original desire would remain after ("endure" if you will) rational consideration of all relevant facts. There is still no reason for it, unless perhaps we are willing to say that the fact that there is no conclusive (motivationally overriding) reason against it is itself a reason for it. But it seems, even
this we cannot say on the initial account. And why? First, assuming that
an agent already has a dominant intrinsic desire, there is no guarantee
that rational consideration of the fact that there are no reasons against
acting on it would result in any additional motivation, beyond that al­
ready involved in the agent's having the desire, to act on or "maintain"
that desire.10 In that case, there just is no reason, on the initial ac­
count, supporting the desire.

But a further problem is this. As indicated earlier there are (or
quite probably are) genuine reasons against acting on the dominant in­
trinsic desire we have supposed the agent to have. Acting on this de­
sire will probably interfere with satisfaction of other, motivationally
weaker, intrinsic desires. But in that case, what reason (or rationally
grounded motivation) is there for acting on the original desire? Cer­
tainly not that ingredient in the original desire itself. If not that, then
what? It seems that here the only consideration available on the initial
account is the following: the fact (if it is a fact) that the motivation
against acting on the original desire that would be generated by ratio­
nal consideration is overall weaker than that ingredient in the original
desire itself. So let us suppose, again an entirely contingent matter,
that rational consideration of the very fact just indicated would itself
add some quantum of new motivation to act on the original desire. It is
certainly possible and indeed likely that this additional motivation would
not be stronger than that against acting on the original desire. Hence,
the only rationally supported motivation for acting on it is weaker than
the like motivation against acting on it. Thus reason would not overall
support acting on the original desire.

It is therefore difficult to see how an agent's pre-existing intrin­
sic desires that, even hypothetically, would not result from the rational
consideration of fact, could ever have the "support of reasons" in Dar­
wall's sense, or, therefore, how the motivation ingredient in such desires
could be rationally grounded per the initial account. If we press these
matters far enough, back through various chains of extrinsic support, it
may turn out that many of our intrinsic (yet normatively unproblematic)
desires will lack, in the broader sense, the overall support of reason.11

This really does threaten to bring down the entire edifice of the
initial account. And the fact that there are no obvious solutions to these
problems indicates that some important matters are not well sorted out
in that account. (As it turns out, the supplements to the initial account
Darwall will later advance do not resolve or even address these prob­
lems, but rather amplify them, as we shall see.)

I take it that it is intuitively plausible that acting on an intrinsic
desire or preference which, though not generated or maintained by the
rational consideration of facts, would withstand any such consideration,
is to that extent normatively justified. To deny this seems theoretically
gratuitous in the context of a broader and more sensible internalism.
That justificatory reasons for an agent must somehow connect with that
agent's actual or potential motivation is true. In this sense, the neces­
sary practical relevance of reasons is secured. But for precisely the
same reason, the truth or validity of entire normative judgments binding
on an agent must also be appropriately related to his motivation. Hence,
internalism need not be viewed solely as laying down a necessary con­
dition for a theory of reasons. But on the initial account, as we have
seen, the relevant motivation is limited to that which would be caused
by consideration of relevant fact. It is this restriction, I think, which is arbitrary and unmotivated. The basic idea is, rather, to specify those conditions under which one's motivation, again actual or potential, has been maximally qualified or tested by reason and reality. Doing what one ought is a function of two parameters, motivation and the cognitive or rational qualification of it. The world brings something to the agent, relations, facts, properties, laws, etc., by way of his cognitive activity, that is, by way of his representation of them and warranted inference from what he thereby represents. But the agent himself brings something to the function as well, his antecedent motivational character, potential or actual. The former may be the dominant factor which produces or structures manifestations of the latter in the context of rational consideration, but again, it may not be. If one's actual motivation would remain despite the maximal effect of reality by way of one's representation of it, that too partakes of justification. Hence, if normative justification consists in the rational sufficiency of motivation in the face of nonnormative truth, then motivation may be rationally justified even though, actually or hypothetically, it was not "produced" by reasons. But to recognize this, one must, I shall argue, allow for the possibility of an irreducibly personal element in normative judgment.

5. A Revised Internalist Account: Cognitive Naturalism

As we have seen, Darwall's initial account of rational consideration clearly reflects the internalist thesis that "reasons for a person to act are considerations [which must be] capable of motivating the person (when considered in the right way)". (20) That reasons, and in general the normative requirements they support, must at a minimum satisfy this condition is clear: If they do not, their practical relevance is obliterated. If reasons cannot affect an agent's motivation, then neither can they serve to determine his actions. And in that case, such putative reasons for an agent cannot be normative for that agent. But, on the internalist view, that is absurd.

What I should now like to do, before proceeding to a direct examination of Darwall's attempts to extend the initial account of reasons, is set up a revised version of just that sort of account.12 It is important to do this for, as I have already hinted and shall later argue, Darwall's supplementation of the initial account may rest on a failure to appreciate the power inherent in that account, and more particularly, the constraints it establishes both in regard to maintaining a robust and coherent internalism and to marking a sensible distinction between metatheoretic and substantive claims in ethics.

Darwall focuses primarily on the notion of "a reason" (or simply on there being "reason") for someone to do something. It is in terms of the explication of that notion that we are to understand more global evaluative notions. For example, what a person rationally ought to do is that which reason recommends. (31) The best thing to do, all things considered, is that which is best supported by reasons, all things considered. (70) I shall reverse this procedure and begin with an explication of the notion of "what one overridingly ought to do", where this is intended to express a theoretically adequate conception of the strongest type of normative requirement applicable to an agent. (From this, a definition of a (sound/good) reason can easily be derived.)
Where S is an agent and A is an action or action type within S's power either to perform or not to perform,

(Tw) S ought overridingly to perform A iff adequate consideration of a (minimal) evaluatively complete set of conditions holding with respect to the actional alternative, S's performing A;

(i) would generate a decision (choice, dominant desire) in S to perform A, or would not extinguish one already operative; and

(ii) would generate, or not extinguish, a preference that S perform A in those subjects whose overriding normative (or evaluative) judgment regarding S's performing A could be expressed by a sentence token of the form, "S ought overridingly to perform A.", in the sense that the truth of such a token would determine the ultimate reasoned preference of those subjects in regard to S's performing (or not performing) A.

Some ancillary definitions are, I think, in order:

(a) By a "condition holding with respect to a thing or action X", I mean anything which can be represented as a condition holding with respect to X.

(b) By "adequate consideration" of a set of conditions holding with respect to a thing X, I mean enough consideration of them such that no further would alter the subject's motivational orientation toward or preference regarding X.

(c) A "minimal evaluatively complete set" of conditions holding with respect to a thing or action X is a (possibly empty) set L of conditions, cl . . . cn, such that

(1) X satisfies the conditions in L; and

(2) X satisfies no further condition d such that adequate condition of d together with the elements of L would alter the decision or preference that would be caused by (or would remain after) adequate consideration of the elements of L alone; and

(3) no proper subset of L satisfies condition (2).13

Certainly, there is not space here to explore in detail the motivations behind or ramifications of the foregoing complex definition. But for present purposes, certain matters are essential.

First, in pre-theoretic terms, Tw pretty much comes to the following: One ought overridingly to perform an action insofar as one would attempt or choose to perform it were reality, by way of one's ad-
IMPARTIALITY AND PRACTICAL REASON

... equitable representation of it, and one's reflection on what is thereby represented, to determine one's choice as much as possible. And in part it is precisely in virtue of this claimed equivalence that one can sensibly talk in terms of "objective" value judgments and normative requirements. Indeed, the latter are, or can be, objective in two important respects on the present account. First, the complex fact asserted to obtain under Tw by a judgment that someone oughtw (i.e. ought overriding) to do something is as much an objective fact about the world as any other, in principle accessible to everyone alike as regards the categories of cognitive judgment, confirmation, disconfirmation, and so on. Second, evaluative and normative judgments are "objective" in the sense that the truth of such judgments is premised on the maximal determination of an agent's motivation and preferences by the subject-independent character of the relevant evaluative object.

Now, under Tw, predications of 'oughtw' are relativised, via subclause (ii), to an open "class" of subjects, as therein specified. When S himself asserts that he oughtw to perform a certain action A, part of what he is claiming is that, were he adequately to consider doing A, he would prefer to do so. An additional part of his claim, that indicated in subclause (ii), is that anyone else who would respond as S does upon adequately considering the matter of S's doing (or not doing) A, would in fact prefer that S do A. Thus, when S judges that he oughtw to do A, subclause (ii) is essentially irrelevant. As to him, it is a dummy claim. But when another person, say P, judges that S oughtw to do A, and we are to conceive of P's judgment as his own normative (or evaluative) judgment regarding S and A, the truth of the judgment requires that P would prefer that S do A were P to adequately consider the matter. And why? Because, on the present account, a person's value judgment is precisely a judgment the truth of which determines the ultimate or projected motivation of that person under some more or less extensive quantum of cognitive and rational qualification. The greater the latter, the stronger and more unqualified is the value judgment with respect to its object.

Tw assigns a complex structure of truth conditions to sentences of the form, 'S oughtw to do A.' But as just suggested, that may not be the whole of the semantics of this or other sentence-types in which strongly evaluative expressions have a primary occurrence. Such expressions also seem to invoke (to be subject to) the pragmatic condition that, at least without suitable qualification, their use by a speaker is properly "reserved" for the assertion of judgments which are evaluative or normative for him, and do not constitute merely another claim of fact about that thing, and indeed possibly, a fact of no motivational relevance to him at all.

We have dealt with first and third (and by implication second) person cases under Tw. But let us suppose that S and P both correctly judge that S oughtw to do A. What about everyone else? Merely in virtue of S's and P's assertion that S oughtw to do A, what is thereby supposed by them, on pain of inconsistency, about the motivation of other persons, other agents, upon adequate consideration of the matter? On the definition given, nothing at all. But two profoundly important points must be kept in mind. The present account does not rule out or make impossible the formulation, by S or P, of any claim about such matters, up to and including the claim by either that, say, every rational agent would have the same preference regarding S's doing A were
each to adequately consider S's doing or not doing A. But neither does our theory require or entail any such claim, or its negation for that matter. Rather, we regard such claims as distinct from and ancillary to the essential content of S's and P's value judgment as such. Claims of the former sort, which essentially have to do with the character of the "class" indicated in subclause (ii) of Tw, are, on the present state of any theory I know of, empirical, clearly and irreducibly so.

The second point is this. It is certainly possible, but again not required under Tw, that the fact that other agents would prefer an alternative upon rationally considering it constitutes a reason for S or P to prefer it, perhaps an overriding reason. For recall that on subdefinition (a), a condition holding with respect to a thing or action will include literally anything that can be expressed in a 'that ... ' clause about that thing or action. This broad notion of a condition is here appropriate owing to the fundamental idea embedded in definition Tw, namely, that the cognitive and rational qualification of motivation, of which normative justification consists, is in large measure a function of the accurate representation of reality, in whatever aspect and from whatever source derived. Hence, that other agents would prefer an alternative upon adequately considering it is as much a fact about that alternative as are other facts, and may be a normatively relevant fact, a reason, for S or P.

On this showing, a sound or good reason for an agent to perform an action is simply any element of a minimal evaluatively complete set holding with respect to that action. It should be clear that definition Tw not only reflects the internalist thesis that reasons are capable of motivating, but also allows sufficient room for the rational qualification of actual desires and preferences. But notice that it does this in a way that is different from that sometimes expressed in Darwall's initial account. What one ought overridingly to do is not that which one would choose to do, "all things considered". (70-1, 80) For on sober reflection, there is no such thing as "all things considered", especially in light of the fact that any action or action type will satisfy a non-denumerably infinite number of conditions. Even if such a requirement were intelligible, it is unnecessary in order to remain within the internalist framework.

To see how this is so, consider how the notion of a limit is employed in our basic definition, Tw. One's motivation or choice is sufficiently rationally qualified in case it has either been generated by or has withstood enough knowledge of and thought about reality such that any further knowledge or thought would make no difference to one's motivation. And that is enough, for once that limit has been reached, any further representation or reflection is motivationally irrelevant and therefore certainly would not alter one's actual or projected decision to perform a certain action. But notice that the potentially infinitary aspect of real objects need not be directly reflected in one's cognition. Indeed, it is quite possible that the limit in question might be reached by an agent upon his adequately considering a relatively small number of salient facts about a thing or action. Notice also that the "or not extinguish" clause in Tw solves the problem (discussed above in Section 4) which beset Darwall's view. Innately determined intrinsic preferences which would survive the process of adequate consideration are overall normatively justified. Moreover, in some cases, the only minimal evaluatively complete set of conditions holding with respect to their respective
objects will be the null set. Hence, such preferences, while normatively justified, are not grounded in or based on reasons. On reflection, this is consonant with linguistic practice. To say, "I just want to," in response to the query, "What are your reasons for doing that?", is not really to specify a substantive reason.

6. Transition—A Potential Clash of Paradigms

Tw is intended to provide a reductive analysis of the notion of an overriding 'ought'. For all that, it entails no substantive value judgment whatever, not even that we or one should use the notion so analyzed. Tw does, however, have the following interesting entailment: If an agent ought overridingingly to perform an action A, and yet does not even attempt to do so, then he is either ignorant of facts that would be relevant to him were he either to be aware of or to reflect further on them, or, his inferences from what he is aware of have been faulty in some logical or epistemological respect. Again, under Tw, the order of increasing justification overall tracks the extent of one's adequate consideration of fact, up to the internalist limit that no further consideration of fact would alter one's motivation. Underlying our entire account then, is the deep intuition that the categories of the evaluative and the normative are constituted by the concepts of truth, cognitive representation, coherent inference, and motivation, all standing in certain complex relations. We might appropriately call such a view, "Cognitive Naturalism."

What, then, of the "moral" 'ought'? The question is interesting and naturally there are several related issues that it raises. First, is the moral 'ought' equivalent to the overriding 'ought', as defined in Tw? And if it is not, how are we to understand it? By now, a number of ethical theoreticians have perhaps become accustomed to thinking of the moral 'ought' as indicating overriding normative requirements. But even if that view is correct, and if Tw does indeed provide an adequate analysis of the overriding 'ought', does it thereby follow that the expressions, 'morally ought' and 'overridingly ought', are equivalent in meaning?

This last question we can answer immediately in the negative. The reason for this is that it may be the case that the qualifier, 'morally', introduces a substantive restriction on the range or kinds of facts about actions which one would call "morally relevant" considerations. Indeed, that we are all inclined to make the distinction just indicated tends to support this view. On the other hand, the overriding 'ought', as explicated in Tw, introduces no such substantive restriction. As an empirical matter, it may turn out that the motivation of an agent (or every agent) that is generated by the adequate consideration of morally relevant facts may be overriding or determinative of that agent's (or every agents') actions in the context of adequate consideration. But this again is to assert the substantive claim that any partial (positive or negative) motivation that would be generated by the adequate consideration of a certain kind of fact about an action would, in the context of the potentially more global sort of adequate consideration required under Tw, be decisionally overriding, for most or all agents. If some such claim could be made out and adequately supported, then with respect to the class of agents one has in mind, one would have established, not that the moral 'ought' and 'oughtw' are identical in semantic content,
but rather that, as a complex matter of contingent fact, the normative requirements of each will coincide, again, with respect to the relevant class of agents.

But to bring matters to a head, let us reiterate one of Darwall's central claims: "... that reasons grounded in our initial account are not the only reasons to act and ... cannot systematically override considerations of moral requirement". (233) If we understand the claim that moral considerations are overriding in the way indicated in the last paragraph, this new claim is incoherent. Had Darwall said that "non-moral considerations grounded in our initial account are not the only reasons to act", etc., the incoherence disappears. But this claim hardly constitutes any significant revision or modification of the initial account. If alternatively, the moral 'ought' means the same as the overriding 'ought' as explicated in Tw, then, taking Tw and the theory in which it is embedded as our revised version of the "initial account", Darwall's claim again seems incoherent, for under that account, reasons include any conditions or considerations that can be expressed in a 'that . . .' clause.23 Try to give an example of a reason that cannot be so expressed! Indeed, let the entire argument of the remainder of Darwall's book be expressed in a 'that . . .' clause or set of them with respect to any action whatever. Under conditions of adequate consideration, those facts will be motivationally relevant or not. If they are not, they cannot constitute reasons under the initial account, or else we give up internalism. If they are, they are accommodated by the initial account. There had, therefore, better be something rather extraordinary in Darwall's argument supporting his "supplements" to the initial account.

But whither now for Darwall? To supplement the initial account. "In the next two parts, we shall consider challenges to the view that rational consideration is wholly self-centered." (112) By "self-centered" he means properly made from the agent's own point of view. He rejects the "self-centered" view and develops two supplements to the initial account. First, he will construct

... a notion of intersubjective value that, I shall argue, is our own. We implicitly take intersubjective values to give us (and others) reasons, I shall maintain, when we adopt certain extremely familiar attitudes towards ourselves and others. The standpoint from which we are moved by considerations of intersubjective value is not our own personal one but an intersubjective standpoint with which we identify. (112-3)

The second, and perhaps even more important supplement, is anticipated as follows:

But even if we do in fact identify with intersubjective standpoints, and implicitly hold others to doing so when we do, could not a person reject all such standpoints and still not act contrary to reason? That is, could not a person rest with our initial account of rational consideration as defined by his own standpoint, as such? I shall argue in Part IV that this is impossible. The normativity of reasons requires the perspective of a rational agent as such as the standpoint from which all reasons, including those grounded in
what motivates an agent from his own point of view, are ultimately assessed. (113)

**7. Objective Reasons and Impersonal Standpoints**

In Chapters 11 and 12, Darwall develops his first major supplement to the initial account, the notion of "intersubjective value". We shall examine with some care both how the latter is constructed and what important implications Darwall thinks it has. This we shall do in the next two sections. First, we must consider a preliminary distinction Darwall develops.

An essential element in Darwall's construction of the notion of intersubjective value is the distinction between objective and subjective reasons. Objective reasons are reasons, per the initial account, which make no essential reference to the agent himself, "as such", for whom it is a reason. (130, 131) Subjective reasons, on the other hand, do involve in their full and explicit formulations essential and ineliminable reference to the agent for whom the consideration is supposed to be a reason. The distinction in question is one that Darwall takes over from Thomas Nagel and reformulates in the terminology of the initial account.24 "Thus, if there is some linguistic item in the expression of the reason that functions as a variable referring to the person for whom the fact is a reason, then it is a subjective reason. Otherwise, it is an objective reason." (121) Remember that reasons, whether objective or subjective, so far must meet the internalist requirements of the initial account. They must be specifications of facts that would to some degree motivate one to prefer a certain action in the context of adequately considering whether or not to perform that action. To take a clear case, that a certain action A would result in there being more square objects in the world would presumably be an objective reason for me to perform A insofar as I would be given some motivation to prefer performing A upon adequately considering that fact. On the other hand, that a certain action B would bring me [the speaker or contemplator of A] discomfort would be a subjective reason for me to do A, provided that, were I to adequately consider that fact, I would be (perhaps partially) motivated to do A.25

Darwall's concern with the distinction between objective and subjective reasons first arose in connection with his discussion, and I think quite successful refutation, of Nagel's thesis that ". . . something can be a reason for one person to act only if it is also a reason for others to enable her so to act". (119) Relevant here are two subtheses Darwall distinguishes, the Thesis of Universality, TU, and the Thesis of Objectivity, TO:

(TU) No fact can be a reason for someone unless that same fact would be a reason for anyone to act similarly in relevantly similar circumstances. (117)

(TO) There are no subjective reasons; any reason for a person to act must be, or be based on, an objective consideration. (118)
According to Darwall,

If both the thesis of universality and the thesis of objectivity are true, then something can be a reason for one person to do something only if it is at least prima facie a reason for anyone to bring about that person's doing it. Notice that both these theses are required to obtain this result. Even if all reasons for acting are objective, a reason for one person to do something will not always give others reason to aid him unless universality holds. Universality guarantees that if a reason exists for a person, it is to do a kind of thing everyone has reason to do. Objectivity guarantees that the kind of thing is not instantiated differently for different persons and thus that the present action is an instance of something they also have reason to promote. (119)

Besides finding fault with Nagel's specific argument for the original thesis, (125-29), more general factors likely underlie Darwall's rejection of it. The first is that Nagel's thesis flatly contradicts the initial account of reasons, for on that account there can be subjective reasons, and Nagel denies this. "If, however, we accept, as we have throughout this study, that a reason for someone to do something is a fact about an act rational consideration of which would motivate the person to do it, other things equal, then there is no a priori reason to accept the universality of all reasons. Different agents may be motivated differently by their rational consideration of the same facts." (130-31)

Despite these reservations, Darwall quite properly does not wish to embrace an extreme subjectivism about reasons: "Now it seems plain that often our preferences are or can be motivated by considerations that make no essential reference to ourselves as such, and therefore, that there are objective reasons for us to act." (131) I see no reason to disagree. There can be objective reasons and subjective reasons. Far from being inconsistent with the initial account, or our revised version of it, such an observation is clearly implicit in both. Thus to mark the distinction between objective and subjective reasons is no more to effect a radical modification or extension of the initial account, or to show that that account is somehow fundamentally incomplete, than would be the ill-conceived attempt to establish similar claims about the axioms of set theory by pointing to the possibility of defining the notion of a Cartesian Product of two sets. Our immediate question, then, concerns what theoretic hay Darwall wishes to make of the possibility that there are OR's. Where does he wish to go with this notion? To make the journey may be interesting, and indeed, we shall again encounter an old conceptual friend, the notion of a perspective or standpoint.

"Whenever we are motivated by our reflective awareness of facts to prefer some act, there will be an associated intrinsic preference." (131) Darwall then defines two kinds of intrinsic preferences:

An intrinsic preference is impersonal if its object can be expressed without free agent variables. Impersonal intrinsic preferences are motivated by objective considerations.

An intrinsic preference is personal if its object can be expressed only with a free agent variable. Personal intrinsic
preferences are motivated by subjective considerations.
(133)

We may provisionally accept these definitions as definitions, but notice several things about them. First, two conditions are expressed, respectively, in each definition. Second, they are logically distinct. For example, from the fact that the object of an intrinsic preference of mine can only be expressed with a free agent variable, it does not follow that it will only be motivated by subjective considerations. It does not even follow that it will be motivated by any subjective considerations. Why, for instance, could it not be the case that reflective awareness of the following facts about a certain thing, that it is shiny, red, worth a great deal of money in the marketplace, all clearly objective considerations, might not motivate me intrinsically to prefer that I get one for myself? Indeed, if we view intrinsic preferences as being either impersonal or personal solely in relation to how their objects are properly specified, then the two definitions given above suggest four possibilities: impersonal preferences motivated solely by objective considerations; impersonal preferences motivated solely by subjective considerations; personal preferences motivated solely by subjective considerations; and personal preferences motivated solely by objective considerations.26

Third, even the four categories specified above, let alone the two distinguished by Darwall, are not exhaustive. Two potentially important ones are left out, namely, intrinsic preferences that would only be motivated by some combination of objective and subjective considerations, and intrinsic preferences that, while they would not be motivated by any considerations at all, would endure adequate consideration of relevant fact.

So, while the first strokes of Darwall’s brush are incomplete and involve some rather complicated cross-classifications, there is no immediate cause for alarm.27 Less innocuous are several claims Darwall makes just after introducing the distinction between impersonal and personal intrinsic preferences:

Personal preferences can only be for states of affairs viewed from a personal standpoint. They get no grip on the world seen from an impersonal perspective. What is preferred is something considered in relation to oneself. Impersonal preferences, on the other hand, are for states of affairs viewed independently of their relation to one—-from an impersonal standpoint.

* * *

The only thing lacking in the impersonal standpoint is certain information: who one is and how what one considers relates to one. Desires, cares, values, and preferences that can be formulated impersonally remain intact. What distinguishes impersonal from personal preferences are their respective objects and the sorts of considerations that can motivate them. (133)

Again,

Objective considerations move us, if they move us at all, from an impersonal standpoint. It is an important conse-
quence of this that the sort of reflective awareness necessary to establish whether an objective consideration is a reason according to our initial account is awareness of it from an impersonal perspective. But as long as an agent considers matters from his own personal standpoint he cannot give any consideration to objective considerations in themselves. to consider objective considerations rationally, then, one must adopt an impersonal standpoint. (134–35)

Finally,

Whether a person would be motivated by his vivid awareness of an objective consideration turns on how he would be motivated from an impersonal standpoint. As long as he considers matters with an eye to their relation to him he cannot be vividly aware of objective considerations themselves. (135)

Now it seems to me that something decidedly odd is going on in these passages. On one interpretation, these claims can be seen to follow trivially from definitions previously given. On another reading, however, certain of them and the picture they suggest are false, or at best extraordinarily misleading. Consider again Darwall's claims that the only thing lacking in the personal standpoint is certain information, viz. personal information, or, that as long as an agent considers something from his own personal standpoint he cannot give any consideration to objective considerations in themselves. If the "personal standpoint" is to be understood as constituted by the consideration only of subjective considerations, then trivially, one will not in the context of such consideration consider objective reasons. Again, when Darwall says that in order to consider objective considerations rationally, or indeed at all, one must adopt an impersonal standpoint, if what is meant is simply that one must consider objective considerations (and so if one considers only subjective considerations, one cannot do this), the claim is again trivial.

On the other hand, recall Darwall's claim that, "As long as he [an agent] considers matters with an eye to their relation to him he cannot be vividly aware of objective considerations themselves." One wonders why this should be so. In particular, why cannot one do both, that is, consider matters in relation to oneself and consider any motivationally relevant objective considerations? Indeed, to take the sum or union of all standpoints, which here simply means that one adequately considers all (or "enough") motivationally relevant truth, would seem to be, per our revised version of the initial account, the normatively and justifi­catorily preferred standpoint to adopt. The passages we have been considering, insofar as they are not given a trivial construal, seem to suggest that if one considers subjective considerations, then one cannot consider objective considerations as well or at all, and vice versa.

Consider the first claim quoted above: "Personal preferences can only be for states of affairs viewed from a personal standpoint. They get no grip on the world seen from an impersonal perspective." (133) Presumably, Darwall would wish to make the same claim about subjective considerations, viz., that subjective considerations can only be viewed from a personal standpoint and hence get no grip on the world seen from an impersonal perspective. In a sense, again, this is true by
definition: subjective considerations are subjective considerations. But in another sense, such claims are entirely misleading. For suppose that I consider the following fact: that a certain thing, X, that I might get for myself is bright red and worth a lot of money. By definition, this is a "subjective consideration", for it contains the "free agent variable", 'myself'. But suppose I imaginatively, vividly, laboriously consider this fact and its logical entailments. Will I not thereby also get a grip on the ingredient objective facts, namely, that X is bright red and worth a lot of money? Of course. Indeed, if we press this far enough, it seems clear that the personal standpoint, viz a viz information about the world, will subsume the impersonal standpoint, for consider the following consideration type, the fact that C is a fact for me to consider, where 'C' may be filled in by any objective consideration whatsoever. This resulting consideration is itself subjective.

Lest the reader think that I have been somewhat uncharitable to Darwall, especially in regard to the nontrivial but problematic interpretations of the various claims we have considered, let me add a final observation on the matter. Darwall says the following: "There is, actually, a puzzle about how a person can bring both subjective and objective considerations together in forming an all-things-considered preference, for that appears to require a person simultaneously to see things both from personal and the impersonal standpoints." (136-37) In a footnote to the passage just quoted, Darwall says: "The sort of disassociation threatened here is that involved in being unable to make an all-things-considered judgment that incorporates both subjective and objective reasons. It is a considerable strength of the position offered in this book that it affords a standpoint to integrate these two perspectives (see Chapter 15, sec. 8)." (137)

This is not merely odd, it is bizarre. The "puzzle" or problem, for the solution to which we are referred to the intricacies of Darwall's Chapter 15, is, on the initial account, utterly spurious. That Darwall sees it as a problem could only come, I think, from his viewing his own earlier claims and metaphors not as definitional entailments of the notions of an objective reason and impersonal preference, but as new substantive claims (and highly problematic ones at that). On the initial account, there is no puzzle and this can be seen at once. One simply considers all relevant considerations or facts, or enough such that consideration of any further would make no difference to one's motivation, and the overriding motivation will carry the day. True enough, specific subsets of considerations, of any kind you like, may set up partial motivations which may tend to direct one toward contrary or inconsistent choices. We somehow sum over these contrary tendencies. How we in fact do that or could do it given the constraints set by our physiology are matters for empirical psychology. Curiously enough, Darwall does not further inquire in his book into the functional mechanics of our psychological capacities.

The long and the short of it is that we seem to have wandered into something of a conceptual bog.

Having now distinguished objective from subjective reasons, impersonal intrinsic preferences from personal intrinsic preferences, Darwall warns of the dangers and hints of the glory to be found on the road ahead. First, the dangers. "The question of which, if any, impersonal preferences we have, or would have if adequately informed, and
therefore which, if any, objective reasons there are for us to act, appears at this point to be unavoidably empirical." (134) Again, "Neither subjective nor objective considerations have any a priori primacy as reasons". (136) Indeed,

We have not, however, attempted to establish, as Nagel did, that the only reasons to act are objective: nor could we. Moreover, as far as our initial account goes, a person may rationally choose personal rather than impersonal preferences if that is what he would be inclined to do on considering all, including personal, information. That is, his informed, all-things-considered preference orderings may systematically rank personal preferences over impersonal ones even if considering matters from an impersonal perspective does motivate him, other things equal, to have preferences. So far as our initial account goes, subjective reasons may generally be weightier reasons than objective ones. They are weightier, other things equal, if the agent would be moved by them more when he considers both. (136)

Now the glory:

In fact, however, some objective considerations have a much greater weight as reasons than our initial account indicates. In what remains of Part III, I shall argue that only a conception of value based on objective considerations is sufficiently rich to fund a number of deeply rooted evaluative attitudes that we take toward ourselves and others. In Part IV, I shall argue that objective considerations of morally right conduct generally override subjective reasons when they conflict. (137)

Sounds interesting. But before we recommence our journey, let us bear in mind what we were promised. It is something more than the truth of the two (somewhat vague) claims just made. While we shall certainly question whether adequate support for their truth will be given, we shall also be concerned to see whether subsequent discussion somehow shows the revised initial account to be incomplete as a general theory of practical rationality, normative justification and overriding normative requirement. Are there fundamentally new aspects of normativity and practical rationality that are missed in the initial account? The question is not idle. For anyone who has appreciated the power of the revised initial account, it should be clear that that account provides the way and contains more than enough theoretic resources for Darwall to make both the claims indicated in the last passage.

Indeed, the initial account broadly indicates the kind of empirical support that would be required to make such claims seem plausible. For example, one must specify what "moral considerations" are, in substantive terms, and then establish that, under the relevant conditions and for each member of the relevant group, perhaps all rational agents whatsoever, preferences generated by adequate consideration of "moral facts" would always override contrary preferences generated by the like consideration of other relevant facts. It should be equally clear that such support may prove hard to come by. Indeed, it may depend on having nothing less than an adequate general psychology and biology of all beings capable of conceptualized cognitive representation. Unless, one
might muse, some way can be found to avoid such stringent requirements, some way to have one's theoretic cake and eat it too, some way to gain the benefits of internalism without having to pay the epistemological price it exacts for maintaining one's preferred substantive normative claims, perhaps via the "derivation" of such claims from those aspects of practical reason itself that are not reflected in the initial account.

8. Intersubjective Value

The notion of intersubjective value is formulated by Darwall as follows:

Let us say, therefore, that something has intersubjective value (relative to a community of valuing beings S) if a preference for it is impersonally basable for any member of S. This means that if something has intersubjective value, any lack of preference (within the relevant group) toward that thing must be explainable either by failure to consider impersonally everything relevant or by failure or incapacity to appreciate what is apprehended. (140-1)

Certainly, but what is an "impersonally basable preference"? Darwall defines that notion this way:

If S would prefer that p be the case, were he to consider in an imaginatively vivid way all and only objective considerations relevant to p's being the case, then a preference that p is impersonally basable for S. (138)

Such a preference, according to Darwall, will both be impersonal and will not ". . . depend in any way on seeing things from a personal standpoint". (138)

Now one thing to notice at once is that the precise contours of these notions seem somewhat artificial. For example, embedded in them is the (rather arbitrary) requirement that the resulting preference among members of a "relevant" community be based only on objective considerations regarding the object of the preference. One could also and just as easily construct a notion of "intersubjective value" which is like Darwall's concept except that it omits this requirement. This second notion too would be, in fact is, coherent.

I suspect that part of Darwall's motivation for requiring that intersubjective value be based only on objective considerations is that he wants to ensure that a somewhat stronger form of "community" prevails when it is true that a thing has intersubjective value relative to that community. Perhaps this is because he thinks that by excluding subjective considerations, that will guarantee that the members of the community would agree not only what they prefer under conditions of (in this case partial) adequate consideration, but also on why they prefer it. And of course, virtually by definition, subjective considerations differ from person to person. Were it a relevant fact for me to prefer something that it would benefit me, Kelley, the consideration is subjective. Were it a relevant fact for Darwall to prefer that same thing that it would benefit him, the consideration is also subjective. In one sense,
both of us have the same reason for preferring the thing, but in the strictest sense not, for two different facts, and hence, presumably, two different reasons, are in question.\(^3\) On the other hand, "If something has intersubjective value relative to a community therefore, there are objective reasons for any member of the community to promote it." (144) And indeed, ". . . by making judgments of intersubjective value we commit ourselves to judgments about shared objective reasons for acting". (144)

At first blush, the move from there being objective reasons for any member of a community to there being shared objective reasons for them would seem to be invalid. For example, that a thing X has intersubjective value relative to a community S seems perfectly consistent with the following possibility. Suppose that S has five members: A, B, C, D and E. Suppose further that the objective considerations about X that would have any motivational effect on any of these agents under conditions of adequate consideration are 01-09. (Recall that certain internalist constraints of the initial account are still in force, in particular, that a reason must be a fact that would motivate a preference under certain conditions.) Finally suppose that the objective considerations which in fact would motivate a preference for X relative to each member of S are as follows: A(01, 02); B(03, 04); C(05); D(06, 07); and E(08, 09). In such a case, a preference for X is impersonally basable for each member of S, but there does not exist any objective consideration here which constitutes an objective reason for that preference for more than one member of S.\(^3\) Hence, it would seem, the existence of shared objective reasons is not guaranteed under either the notion of intersubjective value or that of an impersonally basable preference.\(^2\)

Other glosses Darwall attempts to place on the notion of intersubjective value as he defines it are also apt to give one pause. Consider the following four passages which I label sequentially:

(P1) 

\[
\begin{align*}
\text{\ldots whether or not a concept of intersubjective value can be articulated in a complete and coherent way, it is a well-entrenched part of the way we see things. For better or worse some concept of intersubjective value is our own. (144)}
\end{align*}
\]

(P2) 

\[
\begin{align*}
\text{We are willing to regard some of our impersonal preferences as idiosyncratic, but with others, we are loath to suppose that a conflicting attitude could be impersonally basable for someone. We take the value of these states to be independent of our own preferences as such, and see such preferences as appropriate to that value rather than its source. (139)}
\end{align*}
\]

(P3) 

\[
\begin{align*}
\text{We rarely suppose the value of these things [e.g., the welfare of loved ones, work we find meaningful, social goals we embrace, the enjoyment of nature] to be dependent on our own individual tastes as such. In valuing them as we do, we do not think them good simply because we, as individuals, happen to prefer them. (139)}
\end{align*}
\]

(P4) 

\[
\begin{align*}
\text{Judgments of intersubjective value are not made from a personal standpoint. To ascribe this sort of value to}
\end{align*}
\]
something is to suppose that it rests on features that can be apprehended and appreciated from a perspecti­ve common to a community of valuing beings: an in­tersubjective standpoint. (140)

These don't sound too bad, but nevertheless there are problems.

First, the theoretic relevance of "how we see things" or of what "we are loath to suppose" is not clear. Such observations may be, indeed likely are, relevant to the question of whether we in fact use or apply a notion of intersubjective value roughly analogous to the one Darwall has defined. However, whether or not this or that thing has intersubjective value relative to this or that "community" is quite another matter. Any such specific (substantive) claim is empirical and is not de­cided by what I or most people think or would like to be the case. That is part of the theoretic price one has to pay for specifying objective truth conditions for any type of evaluative claim. Moreover, we also on occasion "see things" non-intersubjectively, and indeed, see them intersubjectively relative to very small and narrowly circumscribed communi­ties. What is it after all to "see things" intersubjectively? Presumably, sometimes to apply the notion of intersubjective value. But that notion is defined in part by means of a floating term, viz., a "relevant commu­nity or group". Hence, I can, I suppose, see the world through the lens of the concept of intersubjective value if I but comprehendingly claim that X is intersubjective value relative to the community consisting of myself and a single friend Harry here.33

Regarding P2 and P3, a relevant point is this. They are offered as part of a brief for the claim that we use some concept of intersubjective value, and in particular, they are offered to explain our deep seeded conceptual (and what shall we say here?) wishes, hopes and desires about the realm of evaluative truth. What puzzles me is a certain aspect of the explanation. Use of the notion of intersubjective value, or perhaps more accurately, the fact that it actually applies to certain things we value, allows us to "validate" (and I can think of no more appropri­ate word here) our intuitions that the value of such things is ..., independent of our own preferences as such" or of "our own individual tastes as such"; that our preferences are rather "appropriate to [their] value rather than its source"; and finally, that some things have value not simply "because we, as individuals, happen to prefer them".

It appears to me that Darwall has, as it were, stacked the de­scriptive deck at a rather inappropriate place in the exposition. First, the issue at this juncture is not whether a thing is good or valuable in some appropriately strong sense merely because it is or happens to be preferred or desired. The internalist account already reflects the fact that desires and preferences are subject to rational and cognitive qual­ification. Resort to the notion of intersubjective value does not for the first time allow this, and on this score does not constitute a revision of the initial account. Second, with the remark that in some cases we think that our preferences are "appropriate" to the value of things rather than the "source" of such value, Darwall comes perilously close to falling back into the intuitionist way of thinking that he has already (and quite properly) rejected. The picture his claim may suggest is that we scan the world with preferential lenses and sometimes discover the value of certain things which exists and can be defined independently of any reference to our motivational and preferential capacity. But such
a picture is flatly inconsistent, not only with the revised initial account, but also with Darwall's definition of intersubjective value. That a thing is intersubjectively valuable, if a fact, is a fact that is constituted by satisfaction of the condition that all members of a relevant community would prefer it on adequately considering all (relevant) objective considerations. Such rationally qualified preferences, given that they would be forthcoming, are an essential ingredient in a thing's intersubjective value. It makes no more literal sense in this context to talk of such preferences being appropriate to the objective value of a thing rather than the source of that value than it would be to insist that the real "source" of combustion is the oxygen, and that to the resulting flame, the requisite heat and combustible fuel are only or merely "appropriate".

Now even if we read, as in all fairness we must, P2 and P3 as referring to rationally qualified preferences, etc., there remains therein the rather startling claim that when we correctly judge that a thing has intersubjective value, it somehow follows in a slightly different sense that the value of the thing, while not independent of any actual or hypothetical preferences whatsoever, is independent of one's own rationally and cognitively qualified preferences. If some such claim could be made out, then we might be able to avoid or at least move some distance away from the sort of relativity to the motivational capacities of particular agents that is inherent in the internalist account. Unfortunately, no such claim can be made out, and certainly not on the basis of Darwall's notion of intersubjective value. For suppose that I correctly judge that a thing has intersubjective value relative to a community S of which I am a member. How is it that its having that intersubjective value is supposed to be independent of or not to depend upon my projected preference under conditions of adequate consideration? How indeed? For if my projected preference were not as it would be, or were to change, it follows as night the day that the thing in question is not intersubjectively valuable relative to the community in question. True enough, my ultimate preference for the thing is no longer sufficient, but is clearly necessary, for its having intersubjective value relative to any community to which I belong. Hence, the agent-relative constraints of internalism are not here avoided.

Finally, with regard to P4, I must confess that I do not understand at all what claim is being made. What feature of, or fact about the world, about anything, cannot in principle be apprehended or appreciated cognitively from that "standpoint" that would permit me or you to consider any fact whatsoever, including subjective considerations? None. "That X would harm me [Kelley]" is a consideration that is just as much open for your awareness and appreciation as it is for mine. Only you would put it differently, viz., "that X would harm him [Kelley]". So whatever kind of value (or indeed intersubjective value relative to any community) is at issue, reality in the aggregate can constitute a common perspective. Of course, what may be bothering Darwall here is that there is a "perspective" of sorts that might not be common to all members of the community unless it is true that X has intersubjective value relative to that community. And what is that? The "perspective" of similarity in ultimate preference regarding X. But in that case, what sense does it make to say that the objective (non-evaluative) features of X can only be apprehended from that perspective? None. So again, the conceptual confusion engendered by this janus-faced notion of a "per-
spective", now cognitive, now motivational, noted earlier in Darwall\'s exposition of the initial account,\textsuperscript{25} has persisted.

9. The Significance of Intersubjective Value

Given the difficulties considered in the last section, one might be inclined simply to junk the concept of intersubjective value. But let us not do so. Indeed, my real quarrel is not with the coherence of Darwall\'s notion, or of some such notion that might be defined. Moreover, I have no doubt that a goodly number of human beings apply something analogous to it or at least think that it applies. In fact, if one carefully attends to the matter, one sees that Darwall\'s notion of intersubjective value is just one of the myriad of ways of filling out or restricting two parameters specified in definition Tw of the revised initial account.\textsuperscript{26} One concerns (substantive claims about) the character and make-up of the class indicated in subclause (ii) of that definition; the other concerns substantive restrictions or limits on the kind of property or consideration with respect to which one is, by one's evaluative claim, projecting positive or negative motivational responses.\textsuperscript{27} Viewed in this light, there are far more interesting questions: Why precisely is there need to place so great an emphasis on this specific conception of value? Why does the notion deserve any special theoretic prominence? Answers to these questions may be found if we attend to other elements in Chapters 11 and 12 of the book.

Let me begin by quoting extensively from the introduction to the latter Chapter.

Intersubjective values are not relative to individual preferences per se, but they do have a more widely based relativity. One conception of intersubjective value relativizes it to the impersonally basable responses of human beings generally. It entails the existence of objective reasons for all human beings, but not necessarily for other rational beings. What is intersubjectively valuable for us may not be so for them, and consequently there may be no [objective?] reason for them to promote it.

* * *

It might be thought, therefore, that our attachment to a concept of intersubjective value is simply a curious fact about us. Could there not be rational agents who were neither subject to motivation from an impersonal standpoint nor disposed to embrace intersubjective values?

In this chapter I shall seek to show that the concept of intersubjective value is not simply a projection of the human condition. Having such a notion is intimately tied to capacities both to view one\'s life as meaningful and to bear attitudes toward oneself that seem central to any being whom we would recognize as self-reflectively rational: self-respect, self-esteem, and self-concern. [Footnote omitted] A conception of shared objective reasons for acting based on intersubjective values is not simply a human fetish: something we can easily imagine rational agents without. Because of its connection to other aspects central to our picture of ratio-
Now to do a bit of reconstructing, there seem to be two central theses embedded in the foregoing passage:

I. An agent's capacity to take or bear certain centrally important attitudes toward himself (or others) entails, in each case, his having a concept of intersubjective value.

II. An agent's being rational entails taking toward himself (perhaps one or more or a sufficient number) of the centrally important attitudes involved in thesis I.

The implied conclusion is obvious:

III. An agent's being rational entails having (and perhaps using sufficiently) a concept of intersubjective value.

Now before proceeding, we simply must clear up a problem regarding the interpretation of theses I-III. These notions of "having" or "being attached to" or "being committed to" a concept just will not do, or will they? I have difficulty understanding what could be meant by one's being "attached" to a concept. Regarding "having a concept", it seems to me that that is simply too weak a notion to support Darwall's claims. Suppose I have a concept of intersubjective value, but never use or apply it. Surely Darwall would not wish to claim that that is all that is involved as a necessary condition in one's being rational. Nor, however, would it seem to be enough here to say that it is enough that one sufficiently "use" or "apply" the concept, and that that is what is being claimed to be necessary in order for one to take the centrally important attitudes toward oneself or others to be rational. For I also have and use the concept of a witch. I typically use it, however, negatively, by way of denying that this or that is a witch (in the literal sense) or in the course of contradicting positive applications of the concept by others. Likewise, I might be an agent who often applies the concept of intersubjective value (relative to some community or other) but only negatively, and indeed I might hold (perhaps somewhat implausibly) that nothing in the world has intersubjective value relative to any group of two or more persons. What I think Darwall really has in mind is that one must sufficiently judge, suppose, assume or believe that at least some things have intersubjective value. Even on this construal, the central theses remain somewhat vague, for we are yet told nothing about how much assuming we must do and what must be the relevant "communities", say in regard to their size or relation to us.

Recall that we are concerned with the overall significance and importance Darwall would attribute to the notion of intersubjective value. Note, however, that so far, Darwall's position is consistent with the view that intersubjective value, the ultimate motivation it would involve and our judgments of it may be limited in scope and of secondary importance compared with the like values, motivation and judgments premised on an agent relative and subjective basis. For example, intersubjective value may be relatively rare or limited to small groups with which particular agents interact. Or, the motivation entailed by true judgments of
intersubjective value may often or systematically be overridden when we, hypothetically, allow agents to take "personal" information into account.

I think, however, Darwall wants to insinuate, if not outright claim, that the foregoing possibilities do not in fact obtain. For example, speaking of the "impersonal perspective", he says:

Such a perspective may not be purely impersonal, since it does not abstract from information that one is a member of the relevant valuing community; but because many of our most important intersubjective values are held relative to very wide communities, such as the community of man, this difference tends in practice not to be too important. (140)

At another place, as you may recall, he claims that with regard to some of our impersonal preferences, we are "loath to suppose" that a conflicting attitude could be impersonally basable for "someone", i.e., anyone else. (139) Moreover, if one attends to the progression in Darwall's exposition, especially in Chapter 11, we move from the agent relative notion of an impersonally basable preference to the definition of intersubjective value with its floating variable term, "S", to be filled in in each case by terms referring to a "relevant community". Little is said about which sorts or sizes of communities, aside from the claims already noted, are likely to be those with respect to which, relative to human agents, intersubjective values exist. By the end of the chapter, and at the beginning of the next, we are suddenly comparing all human beings with Vulcans and pondering whether or not intersubjective values are common to all rational beings. Finally, it is noteworthy that Darwall mentions at the beginning of Chapter 12, and for no apparent reason, a very particular value for the variable 'S': "One conception of intersubjective value relativizes it to the impersonally basable responses of human beings generally." (147) (Later on, he speaks in terms of "one's general community" (151), "reasons valid for all persons" (152), "the set of individuals with whom one fundamentally identifies, most broadly, perhaps, within the human species". (153))

All this, together with the specific arguments Darwall gives to connect the centrally important attitudes with judgments of intersubjective value, suggest that on his view, intersubjective values, premised as they are on impersonally basable preferences and objective considerations, do, and under conditions of adequate consideration would, play a significant and perhaps overriding role in our lives, are held relative to very wide communities, and finally would not be overridden by whatever additional motivation would result from the consideration of personal or subjective information. Let us call this more or less implicit aggregation of claims, thesis IV, or the "scope" thesis.

We have, therefore, four chief claims on the table, namely, theses I, II, III and IV. I shall examine them in roughly the following order: I, IV, II and III.

Thesis I asserts that taking or having certain central attitudes toward oneself or making certain sorts of evaluative judgments about oneself involve in part or commit one to judgments of intersubjective value, some of very wide scope. These attitudes include self-respect, self-esteem, making positive judgments of oneself, having responses of
intrinsic favor or disfavor to oneself, self-concern, caring about oneself, and viewing or finding life meaningful or of worth.

Darwall presents a battery of arguments to support his claims that having or taking these attitudes, even toward oneself, require that one apply a concept of intersubjective value. Many of these arguments share certain overall similarities:

In each case, I have linked the capacity to have a particular sort of attitude toward ourselves to a capacity similarly to respond to others. There is a general explanation for these links. In each instance the response to ourselves is occasioned by our awareness of something in us or about us that is the basis for the attitude. If we have appraisal respect for ourselves [e.g., respect ourselves because we are courageous or have a morally good character] it is because we take ourselves to have qualities we would respect in anyone.

... If we like or esteem ourselves it is, again, in virtue of liking or esteeming certain things in us to which we would similarly respond in others. (163)

For example, suppose one simply has "... some kind of favorable attitude to oneself: liking oneself or being the object of one's own pleasure." (156) Although being favorably disposed toward oneself does not entail being so disposed toward others, it does entail or require "... the capacity to respond to similar aspects in others ...". (156) Indeed,

... one will be capable of liking, and taking pleasure in the awareness of, certain qualities in oneself only if one is also disposed to respond to their presence in others. The sheer idea of oneself gives one nothing to respond to. If one likes oneself it will be because there are things about or in oneself to which one responds favorably, the awareness of which provides one with pleasure. One's liking of various things about oneself cannot be constructed out of a basic liking for oneself ... for there is nothing in the simple idea of oneself to be liked or not liked. Rather, such a liking for oneself can be constructed only out of responses to qualities one takes to be one's own. Plainly, these are qualities that others may have as well, and since it is our responses to these qualities that are basic, the qualities will tend to be an object of favor whether perceived in oneself or in others.

* * *

... the only possible way in which an attitude to oneself might bear the imprint of the personal standpoint would be if it were an attitude to oneself, or to qualities in one, as related to oneself--for example, as being useful to oneself. If we are capable of direct or intrinsic response to certain qualities in us, to finding such qualities immediately agreeable, then our response to ourselves is an impersonal one. It is essentially informed neither by our beliefs that such
qualities are in ourselves nor by the belief that they are useful to ourselves. (156-7)

This is a fairly typical example of the sort of argument Darwall advances. Again, there are problems. First, it is not clear why one cannot have a favorable attitude toward oneself in part or overall, but not because of something about oneself. Suppose, for example, that we are simply designed to tend to have such an attitude, not by way of any response, under conditions of adequate consideration or otherwise, to the cognition of general features of ourselves. Here again Darwall just assumes that our positive attitudes must always be produced or generated by our awareness of facts about ourselves and as a response to what is thereby represented. Second, even if we restrict our attention to positive attitudes that would be so produced and so would constitute "responses", why can we not respond positively to ourselves as a whole, in virtue of all or most of the features constituting our entire identity?

Darwall will reply that the "bare" or "sheer" idea of oneself gives one nothing to respond to. But this idea, whatever it involves, is certainly not the only idea of oneself. In the fullest (perhaps non-Cartesian or Lockean) sense, my idea of myself is extraordinarily complex, involving elements pertaining to my personal history, psychology, beliefs, and so on. It is all or most of this that I may respond positively toward or positively appraise. Darwall can reply that this aggregate of features and qualities could be possessed by others. But of course, if it is at all an extensive set that comprises the basis for my projected response, the real chances of this are virtually nil. Moreover, even in cases where one's positive response to oneself saliently features a particular quality which could be possessed by many others, it will not always be the case that my attitude or response toward myself will merely be, as it were, in the abstract, to that quality or to its instantiation simpliciter. Thus, I may be disposed to respond positively, at least to some extent, to courage, whenever and wherever found. Hence, if I find that I am courageous, I will therefore respond positively to myself or to that aspect of myself. I will admire my courage because I admire courage generally. On the other hand, what I may positively respond to, or admire, is "my courage" or "my having courage", where what is admirable to me is that a being such as myself also has the quality of being courageous. Hence, the response may depend essentially on my awareness of a complex of features of or in myself which few if any other persons are likely to have.41

Moreover, it is difficult to know what to make of Darwall's repeated plaint about the "bare" or "sheer" idea of oneself. Supposedly, this idea gives one "nothing to respond to". But is that so? Let us suppose that you and I are identical twins with virtually identical histories. We have been raised apart, as an experiment, in controlled surroundings designed to maximize the similarities in our beliefs, interests, character traits, and so forth. The evil genius who has been conducting this experiment informs us of these facts and indicates to me that one of us, you or me, is to be severely tortured in the next few minutes. He then leaves it up to me to decide which of us it will be. If I decline to choose one or the other, we will both be severely tortured. Now it seems to me that in such a case, although there is nothing or little qualitatively different about you and me that I might lay hold of or respond to, there remains the fact that if I choose myself to undergo the forthcoming torture, it is I who will experience the pain, whereas if I choose
you, it will not be me, but rather you, who suffers. Call it a "bare idea" if you will, one would think that it is at least possible, if not likely, that the relevance of this bare 'I' could be positively motivating, under conditions of adequate consideration or not. And of course, a similar possibility will hold in virtually all cases pertaining to one's experiencing pleasures, pains, enjoyments, or regarding the distribution of resources, opportunities, etc.

Another important point is this. Darwall's claim that the bare idea of oneself is motivationally inert seems fundamentally inconsistent with a large part of his program and exposition to this point. For if the claim were true, it can make but little sense to go to the great lengths he does to exclude subjective considerations in his definitions of intersubjective value, objective reasons, impersonal preferences, standpoints or points of view. It seems that the only idea that could make a consideration or bit of information relevantly subjective is that of the bare "I" or "me" (even where in the process of deliberation, one arrives at it indirectly, e.g., by noting that X will harm him and so being led to distinguish him from myself). But, if this idea is itself everywhere and always motivationally inert, then it is quite gratuitous to exclude it, for its occurrence in some consideration cannot possibly make any difference to the character and direction of any resulting motivation or preference. Indeed, it follows here that there are no subjective reasons in anything but a paper sense, precisely owing to the point just made. Hence, either Darwall is wrong in claiming that this idea is motivationally inert, or there has so far been no need to make and emphasize any distinctions ultimately depending on the presence or absence of it.

In sum, an agent's positive attitudes toward himself may not imply that his positive attitudes toward others will be as extensive as or as deep as Darwall's initial picture suggests. But even so, let us grant this much of his position, making relevant qualifications explicit: If one has a (partial) positive response to oneself solely because of one's rational consideration of the fact that one instantiates a certain quality or feature, and if one's awareness of the fact that it is instantiated in oneself is not a necessary condition for the production of the positive response, and finally, if the quality or feature in question is one that is or can be instantiated in other agents, then, were there such agents and were one to rationally consider the fact that the given quality is instantiated in them, one would at least to some extent positively respond to those other agents. So much for basic causal analysis. But a further important question is this: In such cases, how does it follow that a particular agent who takes a positive attitude toward himself thereby commits himself to making a positive judgment about the existence of intersubjective value? So far, the argument has shown only that if I respond positively and solely to certain generally instantiable features, then I must to some extent respond positively to their instantiation in any case. But how does it follow that anything has intersubjective value relative to any group or that I am committed to making such a judgment? That's a different issue.

A typical instance of the sort of argument Darwall gives to bridge this final gap is the following. Consider the sort of respect one can give to a thing or kind of thing which is premised on a view both about the features of that thing that warrant respect and about the appropriate conduct toward it. Darwall calls this "recognition respect". (149) This sort of respect Darwall thinks "... involves the notion of intersub-
jective value". (149) And why? Suppose for example that an agent re-
spects his own or another person's right to privacy.

He will be inclined to disprefer some acts and prefer others
because of the way in which he believes the right to pri-
vacy bears on them. But his preferences here are not pri-
marily about his own conduct. Insofar as he prefers his
doing something because it is what proper respect for pri-
vacy requires, he equally prefers anyone's so acting.
Recognition respect, then, involves impersonal preference: A
preference that people conduct themselves in certain ways
toward the respected object. (149)

But so far, this is merely another version of the argument we have al-
ready examined, and it does not commit this agent to any positive judg-
ment regarding the intersubjective value of privacy. Then Darwall says,

Moreover, recognition respect involves the judgment that
something is worthy of respect, that is, worthy of anyone's
respect. [Footnote: Of anyone, that is, within the relevant
community of valuing beings.] This means that the person
who has respect for something believes that his impersonal
preference for certain treatment of it is not idiosyncratic
but is, rather, an attitude that is impersonally basable for
anyone (in the community). (149)

And, according to Darwall, this amounts to a judgment of intersubjective
value.

I confess that I do not follow this argument. I see nothing inco-
herent in the supposition that someone might impersonally respect his
own or another's privacy and yet might also believe that other beings
or persons, with whom he has contact or not, would not, even under
conditions of adequate consideration, share that attitude. Darwall's ar-

gument appears simply to beg the question. On the other hand, if Dar-
wall would wish to reply that the sort of recognition respect about
which he is talking is defined to be that which one would not have were
one to believe that the projected preference were not impersonally basa-
ble in the "relevant" community, his conclusion is trivial. Although it is
hard to tell, perhaps what Darwall is assuming here is that if one agent
responds positively toward something solely in virtue of its having cer-
tain generally instantiable features, then all agents must respond like-
wise. That just doesn't follow. So in any case, this sort of argument
Darwall offers to the effect that bearing certain attitudes towards one-
self or others requires that one make or is committed to judgments of
intersubjective value is either question begging or vacuous.

There is a way, however, to understand (or recast) Darwall's ar-
gument supporting his claim about commitment to and the existence of
intersubjective values. It is this. To the extent that agents do or would,
under conditions of adequate consideration, bear toward themselves the
sort of positive attitude (subject to the relevant qualifications indicated
above), it would follow that intersubjective values exist. Put another
way, if one has several rational egoists each of whom would bear this
sort of attitude toward himself, then each would bear a similar attitude
toward the others upon adequately considering the matter. And from
this it follows that some intersubjective values exist. But two points
cannot be avoided. First, granting the foregoing entailments does not establish that there is a conceptual connection between taking the relevant sort of attitude toward oneself and there being intersubjective values (and hence that one is somehow committed to judging or believing that there are). Either of the latter claims can only be reached with the addition of an empirical premise. Second, none of this weakens the points about to be made below.

So far, we have focused on thesis I. III is thereby also called in question, since I is an essential premise in Darwall’s central argument for it. The issues we have just considered also have relevance to thesis IV, the "scope" thesis regarding intersubjective values. Taken as a whole, Darwall’s exposition is apt to leave one with the impression that intersubjective values (or our judgments of them) hold relative to very wide communities and are perhaps dominant in the lives of human beings. If that were so, there would be a significant convergence of preference among human beings generally and hence, on Darwall’s view, often overriding shared objective reasons to promote the preferences and interests of others (though it seems possible that this is precisely because they are my preferences as well).

Now, despite the vagueness of thesis IV, there are important matters to bear in mind. First, Darwall has not given us any reason to think that personal preferences and interests will not play a significant and perhaps dominant role, in fact or under conditions of adequate consideration, in the lives of most persons. Preferences relating to my having certain resources, pleasurable experiences, the welfare of my loved ones, the flow of my fortunes, etc., cannot therefore be ignored. Recall, that the definition of intersubjective value contained the variable ‘S’ referring to a relevant group or community. Why suppose that that variable will typically be instantiated by very wide communities, for example the human species? Is it not possible that I might prefer the interests of certain groups to which I belong, even quite small ones such as my own family, to those of others in cases of potential conflict? Given the potential multiplicity of overlapping groups to which any agent belongs, there might arise, even under conditions of adequate consideration of objectively given facts about things, as much ultimate diversity in rationally qualified preferences as was possible under the initial account which focused on single agents. But whether or not these things are actually so, and what the extent of such divergence is, are empirical questions.42

Of course, if as a matter of substantive evaluation or preference, one does not much care for these possibilities, and yet one does not have the empirical evidence to rule them out, what is left but to suggest that, in effect, as far as human beings are concerned, our most important intersubjective values are held relative to very wide communities, e.g., the community of man, and hence, that any divergence and interest or preference will in the end be limited in scope, "in practice".

Sorry, no, not proven or yet even made plausible. If the history of the world as it has been, and the behavior of the people in general as it now is are to be any guide to what I or you would specifically prefer under conditions of adequate consideration, a strong case can be made that ultimately, and with any significant scarcity of resources, agents are likely to prefer the interests of those who are, for example, emotionally, genetically or culturally proximate to themselves. But
whether this is in fact so, and the extent of such differential preferencing, are not the issue, however. It is, rather, that these are empirical matters, even relative to the human species, and are not, one would think, properly a part of the meta-theory of value.43

This brings us round, at last, to Darwall's thesis II, that being rational entails or requires taking a sufficient number of the centrally important attitudes. To begin our discussion of this matter, let us consider what Darwall seems to regard as the crowning jewel in his set of arguments for thesis I. Here the specific claim is that in order to care about things, or for them to have meaning for us, and in general for us to regard our lives as having meaning or value or worth, we are committed to judgments of intersubjective value, relative to very wide communities indeed. What is the argument?

Let us suppose that there is something an agent cares about or finds meaningful.

Something can give meaning to our lives only if we believe it to have intersubjective value. The reason for this is simple. Values that give our lives meaning both inspire and root our lives. They give our spirits the very air they need to breathe. They give us a rootedness: a place to stand, to defend, and to hold precious. But the value of what both enlivens and supports us cannot of itself be based on our own individual preferences as such. Since we could derive neither breath nor support from them unless we perceive them to be of value, we cannot see their own value as emanating simply from us. The moment we are aware of that something has value only for us we cannot draw the crafts-person's distinction between the way she regards pickup sticks (which she may intrinsically like) and the way she regards her craft. The difference is not that one is liked more than the other but that one is taken to be more important or serious, a more meaningful enterprise. That which endows our life with meaning must be something whose value we regard as self-transcendent. (164-5)

For example, for an agent to "... see his own childrearing as a meaningful activity at all he must regard childbearing as intersubjectively valuable: one that is at least prima facie worthwhile for anyone to engage in". (165)

I have already suggested that if the relevant community Darwall has in mind is that of all human beings or all rational agents, his claim is grotesquely implausible. It is also clear that his claim constitutes an empirical generalization of extraordinary scope. But, we may ask, how large must the relevant community be believed by an agent to be before he will no longer care about or find valuable or meaningful a particular thing? Darwall has not shown that there is anything contradictory about the supposition that an agent might intrinsically prefer something under conditions of adequate consideration, and continue to so prefer it even though he firmly believed that no other agent shared his preference. Indeed, even if I knew that everyone in the world, upon adequately considering the matter, would not prefer to seek theoretic knowledge, or would not prefer the welfare of my loved ones, I seriously doubt that my preference for these things would disappear or weaken or occupy a
less central place in my life. Darwall's general claim appears to entail the contrary.

If one thinks about it, the real appeal that Darwall may here be making is to what we might term the "herding factor". That is, under conditions of adequate consideration, would my awareness of the fact that all or most or many other agents would prefer something be a necessary condition for my preferring it? Again, under conditions of adequate consideration, how extensive would this phenomenon be? These matters are empirical. But an important thing to notice, one that we have already pointed out, is that even if in certain cases such facts would be motivationally relevant to an agent, this marks no modification or extension of our revised initial account. If I have adequately considered, per definition $T_v$, the relevant evaluative object, then, even though I may not know whether anyone else would share my preference for it, my preference and its motivic strength are set, determined. Hence, the "importance" or "significance" to me of that object and of my preference for it will not be affected, unless they have already been affected, by any further facts about whom or how many other persons or agents would or do prefer the same object.44

Finally, how is it that taking a sufficient number of these "centrally important" attitudes is conceptually tied to rationality? Suppose that an agent, upon adequately considering all relevant matters, as a result has no intrinsic preferences whatever. In what does his irrationality consist? Has he thereby contradicted himself or made incoherent inferences? Not obviously. Are there reasons which he is ignoring or has not sufficiently taken into account? By definition, not. We may not envy his state, but so what? Again, and here we may take up thesis III more or less directly, suppose there is an agent, Charlie, who has three intrinsic preferences. The first two are to maximize his own pleasure and his acquisition of knowledge, and the third is to construct a gigantic spherical object that will orbit the sun between Venus and Earth. The first two preferences are personal, are things about which Charlie deeply cares, but not in Darwall's sense, that he would only care about them or find them "meaningful" or important unless he believed that others would as well. The third (impersonal) preference is shared by no one else, nor would it be, under conditions of adequate consideration. Finally, let us assume that Charlie has adequately considered all relevant facts and has no beliefs about the intersubjective value of the objects of his preferences. Is he therefore irrational? For the life of me, I cannot see why. We may not like Charlie or we may want to change him, but does that show that he is irrational? If Darwall would wish to claim that it does, I would suggest that he is merely and surreptitiously building contingent substantive evaluative judgments into the notion of rationality.

Indeed, the plausibility of thesis III is shattered by the following simple fact. It is conceivable that under conditions of adequate consideration each agent in the universe would be left with a set of intrinsic preferences shared by no other agent. But if that is so, then all propositions asserting the existence of intersubjective values will be false. Therefore, in such a case, if Darwall were right, an agent's rationality would depend on his having a sufficient number of false beliefs about the world. But that is absurd. Consider, for example, Harold, an ordinary fellow under the following circumstances. Every agent in the universe adequately considers all of relevant reality. All agents except Harold
come to have no intrinsic preferences whatever. Harold, on the other hand, remains pretty much as he was. Thesis III entails that Harold cannot be rational. Again, that is absurd.

In sum, Darwall has done nothing to undermine the following points:

1. Questions of what has intersubjective value relative to what communities, what has objective value relative to an agent, or what the extent of convergence or divergence in ultimate preference across agents would be, are empirical and substantive and not properly a part of the meta-theory of value.

2. Darwall's fairly optimistic assessment of the motivic importance in our lives of intersubjective values relative to large communities, e.g., the human species, is not established, and indeed is somewhat doubtful.

3. Applying the concept of intersubjective value, relative to small or large communities, has not been shown to be necessary to rationality, unless we simply wish to use the term 'rational' as a substantive evaluative term.

4. Darwall has not shown that there is any a priori normative priority either to intersubjective values or to preferences based on objective considerations. The so-called "essential rationality in motivation from the impersonal standpoint" (147) parallels and is matched by the corresponding rationality of motivation from the personal standpoint. And neither has been shown to be evaluatively or normatively overriding to motivation from the unrestricted cognitive standpoint that would range over subjective and objective considerations.

10. The Rational and the Moral

We turn at last to consideration of the second, and perhaps more important, attempt to supplement the initial account. It is here that we shall explicitly confront Darwall's incompleteness thesis as well as his metaethical neorationalism.

At the beginning of Chapter 13, Darwall raises anew the question of whether it is possible that the "central attitudes" treated in Chapter 12 and the "impersonal motivation" they involve might have anything to do "rationality per se". (171) "Could not perfectly rational beings exist without them and without, therefore, the concept of intersubjective value they involve?", "... beings for whom there are no objective reasons to act at all even though ... they are often in a position to benefit others or not to harm them, states we believe there is objective reason to promote?" (171) These are troubling possibilities because we take some considerations "... to be objective reasons for any rational agent". (172)
In particular, we have a conception of morality or right conduct whose most fundamental requirements and recommendations are thought to provide reasons for any agent who is capable of rational action and who exists in the company of others. (172)

In this connection, it is a relatively specific conception of morality which Darwall will seek to validate. For example, it is a "... thesis of common sense that any rational agent has overriding reason to do as he is morally required". (198) "Moreover, morality is thought to generate reasons superior [to] and more weighty than reasons of other sorts. It claims by its very nature final jurisdiction for itself, and its requirements must override all others when they conflict." (175, footnote omitted.) In addition, "we" think that the validity of certain moral requirements cannot solely be based on the intersubjective value of states of affairs we bring about by moral action. We are committed to a conception of the moral value of conduct that is sui generis and not simply derivative from the nonmoral intersubjective value of states of affairs. (172, footnote omitted.)

And why?

The most basic moral requirements cannot be applicable to any beings capable of moral agency if they derive from nonmoral value, since both subjective and intersubjective nonmoral values derive from the informed responses of beings, and no such responses are, of necessity, common to all rational agents. [footnote omitted.] The sui generis character of moral requirements, then, goes hand in hand with their validity for all rational beings. (174)

But all this raises the question: "... how is the claim to be sustained that morality provides all agents with weightiest reasons for acting?" (175)

The assertion that morality provides all agents with reasons for acting is not itself simply a moral one. Were we to understand it in that way it would be no different from the assertion that considerations internal to any point of view are considerations from that point of view: for example, that the considerations of etiquette bind from the point of view of etiquette. Rather, it is a view about the rationality of moral conduct. (175-6)

We are therefore concerned with the rational justification of morality (or at least the Kantian conception of it Darwall embraces).47

Darwall considers two attempts, by Baier and Gauthier, to provide such justification, but finds them wanting. For our purposes, it is not necessary here to examine the details of Darwall's analysis and criticism of these approaches.48 It will be worthwhile, however, to look at certain general features of these approaches and the assumptions underlying them which will structure Darwall's conception of his central task in the latter part of the book. For example, each assumes that moral require-
ments can conflict with self-interest. (184, 194) It is clear that Darwall also believes that this is a coherent possibility. Thus, Baier, Gauthier and Darwall attempt to confront the rational egoist. Yet, "self-interest" here does not, indeed cannot relevantly, equate solely with selfish or non-altruistic concerns and preferences. Nor does self-interest concern mere de facto or cognitively unqualified desires and preferences that one might happen to have. This becomes clear in Darwall's discussion of Gauthier's approach. For the latter, rational egoism is viewed as "unconstrained individual utility maximization". (194) Indeed,

... Gauthier is understanding utility as a measure of preferences the agent would have "were he sufficiently informed and reflective" [reference omitted]. We may take it, then, that Gauthier's agents have the preferences they would have were they rationally to consider all relevant information as per our initial account in Part II. (195)

(We shall assume in what follows that we are operating on the revised initial account explicated in Section 5 above.)

Darwall it seems will attempt to show not only that fundamental substantive moral requirements are universally valid for all rational agents, and indeed override all other preferences when they conflict, but that these claims are tied to and somehow may be derived from the very concept of practical rationality, "per se". Yet, these requirements will not be based on the informed responses of agents, since, in that case, they would be, as Darwall holds they must not be, derivable from either subjective or intersubjective nonmoral values. (174) And how, generally, is this to be accomplished? By attending to the "normative aspect of practical rationality". (199)

Reasons for acting are considerations by which an agent ought (rationally) to be guided in acting. Indeed this normative aspect is itself concealed in our definition of the motivational feature. A fact is reason for someone if he would be motivated by it on considering it as he rationally ought.

* * *

I shall argue that since the most basic notion of practical reason is normative, an ideally rational agent would be disposed to act on whatever principles constitute his conception of rationality on account of there being principles on which any agent ought rationally to act. (199)

This will constitute a theory of practical reason that is "fundamentally deontological". (199) Neither it nor its connection with moral requirements will make "... the rational person's desire to act for reasons conditional on his desire that he, or that people generally, flourish". (191) Indeed, a fully rational person's desire to act for reasons would not, according to Darwall, seem to be conditional in that way. "That there is [may we say, overriding] reason for him to act is necessarily regarded by the rational person as a deliberation stopper." (191)

As we shall see, and consonant with the remarks quoted above, Darwall's general strategy will in part be to validate the indicated conception of fundamental moral requirements by articulating the notion of
an ideally rational agent. And in the end, this will permit him ". . . to advance an argument proceeding from a formal conception of practical reason to the conclusion that at least some basic moral considerations are overriding reasons for any agent who exists in the company of others". (202-3)

What conception of practical rationality is at issue? Darwall first introduces the notion of a normative system. It is defined in terms of certain essential elements that comprise it, including, inter alia, norms, agents subject to those norms, the behavioral objects of the norms, etc. (203) Perhaps the central elements here are norms: "Norms are the standards or principles that those to whom they apply, the subjects of the normative system, use to appraise and guide their conduct. This element is crucial. Norms are guides that subjects can themselves apply to regulate their own conduct." (203) Moreover, our notion of practical reason is normative in that it implicitly refers to a system of norms. (204) Appraisals of subjects as rational also make implicit reference at least to,

... the idea of standards that underly and justify them: the norms to which rational creatures as such are subject. In identifying ourselves as rational creatures in this sense we implicitly refer to our status as subject to rational norms. We suppose ourselves capable of making rational appraisals and of guiding our behavior by them. (204)

On Darwall's view, all normative appraisals, including the appraisal (or characterization) of something as rational, make implicit reference to norms. (207) Yet Darwall reminds us that this does not guarantee that all reasons are universal, that is, that any fact that is a reason for one rational agent is so for all. "At the very minimum, however, the normative aspect of reasons for acting requires this much: all agents ought (rationally) to act as there is, all things considered, reason for them to act." (208) Yet, ". . . there can be no explanation of why a person ought rationally to act on reasons unless something's being a reason relates it to a norm of rationality that applies to all agents". (208) "In claiming that some particular principle has the status of a principle of rationality, we claim that it is a principle that is normative for all rational agents: one they ought rationally to follow." (208) Finally, even though on the internalist account ". . . a fact's being a reason is not grounded in universal reason-specifying principles, it is still anchored in a universal principle: namely, for any agent, if he would be motivated to prefer an act by his rational consideration of a fact about it, then that fact is a reason for him to act". (210) Indeed,

The normative character of reasons for acting explains the initial openness of the question of what facts are reasons for a person to act. With respect to any fact we may at least sensibly raise the question whether it is a reason. That no substantive considerations are analytic of the notion of a reason to act is a special case of the logical the gap between 'ought' and 'is'. (202, footnote omitted.)
11. In the Temple of ISIS

The central arguments of Darwall's book are to be found at the end of Chapter 14 and in Chapter 15. These he takes to establish a number of paramount and related claims we have adumbrated in earlier sections of this paper, viz., that the initial account (here the revised initial account) is incomplete as a theory of practical rationality or of reasons, that the "common sense" (Kantian!) conception of morality can be validated, and indeed shown to flow largely from formal considerations having to do with the very nature of practical reason, and that the informed rational egoist can be confronted and defeated. To these claims I respectfully demur.

My approach will be twofold. I shall first examine in detail one part of Darwall's overall argument. If it falls, an essential premise in the larger argument also falls. I have chosen this specific argument because it most glaringly trades on and exhibits the sort of deep error which the revised initial account would counsel against. The moves Darwall attempts to make in this argument constitute the very nerve, as we shall see, of his departure from internalism. The second part of my approach will be to provide, at least in outline form, a simpler and more coherent explanation of the principal matters addressed by Darwall, including the nature of practical rationality.

Darwall's opening gambit is to characterize what he calls a "rationally self-identified (and self-critical) subject of the RNS", or ISIS. Preliminarily, such an agent recognizes that he is a subject of the system of rational norms or principles, the RNS. (211) He takes, and is capable of taking, the internal point of view (somewhat in H. L. A. Hart's sense) toward rational norms. He is moved to act on such norms, or to do what they require, just because they apply to him and he accepts their very existence as justification for his motivation and conduct. (212)

... The agent who is internally self-identified as subject of the RNS is one for whom its norms have motivational and attitudinal weight. He is disposed to ask which course of action is best supported by reasons, to consider matters in certain ways because he believes it rational to do so, and to act on his judgment of what, all things considered, it is best to do. Although his reasons will themselves often provide him with motivation, especially when he judges their rationality for himself, his judgment of them as reasons is the controlling factor. (213-14)

But what about the norms of the RNS? Precisely to what does the ISIS take himself to be subject? As we have seen, rational norms, in this sense, apply to, are valid for or binding on, all rational agents (or agents capable of rationality). But more than this, and independently the concept of an ISIS, "It is part of the very idea of the RNS that its norms are finally authoritative in settling questions of what to do". (215) Indeed,

Once we have established that a line of conduct is recommended [or required] by norms of the RNS, however, we put to rest the question of why we should do it. We cannot sig-
significantly ask whether there is [perhaps overriding] reason to do what there is [overriding] reason to do.

... it is only in the case of rational norms that this question is unmeaning. With respect to any other norms we may sensibly ask why (that is, for what reason) we should do what they require of us. Only with respect of those norms in terms of which reasons are themselves understood, conceived as such, can we not meaningfully ask why we should follow them. (215-16)

Hence, because an ISIS would be aware of the foregoing considerations, "In internalizing a conception of the RNS as constituted by norms that are finally authoritative on matters of practice, he regards his own commitment to follow such norms to be unconditional". (216)

Now, whether as a further articulation of the concept of an ISIS, or as a logical consequence of the foregoing claim, or perhaps both, we find that an ISIS has some very interesting properties which I shall here summarize. What not only explains his action but also motivates an ISIS to act on a certain principle is precisely (and apparently only) that he judges it to be a rational norm, that is, by definition, a norm on which any agent ought rationally to act. (212-213) Furthermore, an ISIS regards his action on a rational norm to be justified by its being a rational norm, i.e., one on which all agents ought rationally to act. In short, "an ISIS is concerned first and fundamentally to act rationally" (228), that is, to act on rational norms and to do what they require.

Certainly an ISIS will have an overriding or dominant desire that he act on rational principles and norms. But the judgment that a principle or norm is an element of the RNS can only be made from the impersonal intersubjective standpoint of a subject of rational norms or of a member of the community of rational beings. (211, 227) Hence, what justifies and motivates an ISIS's overriding preference or desire that he act on a certain principle is, and only is, his judgment that that principle is one on which all agents ought rationally to act. (227) Hence, his personal desire to act on rational norms is derivative from his impersonal judgment that the norms in question are norms on which all agents ought rationally to act. But since the latter desire or preference is impersonal, and overriding or "controlling", the ISIS will equally prefer that any agent act on the norms in question. Thus,

... while an ISIS's action on P is indeed an expression of her (personal) desire to act rationally, she cannot see it as simply so from her own viewpoint. That would rob her internalized conception of rationality of its normativity.

To understand this, consider how such an agent would ordinarily regard acts she sees as simply expressions of her desire and not as further justified. While she takes action on her desires to be presumptively rational as per our initial account, she regards the reasons to promote any particular desired state to be contingent on her having the desire, or on its being the case that she would have it were she adequately informed.
In order to regard action on a desire to act rationally similarly, she would have to think that the rational warrant for acting rationally is conditional on her having the desire to be rational; or on its being the case that she would so desire were she adequately informed. But this conflicts with what she internalizes in accepting her status as subject of the RNS. In accepting that role she accepts the idea that norms such as P ought rationally to be acted on by all agents, including herself, simply because they, and she, are subject to them. (224-25)

The principal conclusions about the character of an ISIS to be drawn from the foregoing considerations are formulated by Darwall as follows:

It follows, therefore, that a necessary condition of an ISIS's regarding P to be a rational principle is that he impersonally prefer [and we must say overridingly prefer] that all agents act on P. Moreover, since his preference that all agents act on P is grounded [solely] in P's being a rational principle, it will not be conditional on anything not relevant to the latter judgment. This means that in addition to being impersonal, the preference will not be personally based. We may say, then, that if an ISIS takes P to be a rational principle, then he prefers from the impartial standpoint of an arbitrary rational agent that all agents act on P. This is the motivational consequence of the internal acceptance of rational norms. (226)

Bearing in mind the special (and overriding) motivational and normative character of the ISIS, consider how such an agent would have to regard, according to Darwall, pursuit of self-interest, here understood as those preferences the agent would have under conditions of adequate consideration per the revised initial account: "Insofar as she is an ISIS, any preference she has for promoting her own interest will be [solely?] motivated by her judgment that any agent ought rationally to do so. Since the latter judgment is impersonal, it will [overridingly] motivate the impersonal preference that all agents promote their own interests." (228)

In sum, an ISIS is overridingly disposed or motivated to perform actions because those actions are required or recommended by rational norms, those the ISIS takes to be applicable to all rational agents as such. Yet, what turns the ISIS into more than an idle curiosity in Darwall's conceptual bestiary is that, as Darwall puts it, "We may assume that... an ideally rational agent is an ISIS of the RNS". (219) Moreover, the rational norms which are elements of the RNS are norms, and by definition, a norm must be able to function as a guide for those subject to it. (220) In light of this, Darwall formulates the "guidance principle": "A principle is a rational principle only if an ISIS of the RNS could regard it as such." (221, footnote omitted) It follows, in turn, that, "If a principle's having a feature is necessary for it to be regarded as a rational principle by an ISIS, then its having that feature is also a necessary condition of its actually being a rational principle". (221)
Now recall that the ISIS can only (or more accurately, need only)
judge that a principle is a rational one from the impartial perspective of
a rational agent as such, without any information about his personal in­
terests, and so forth. This, Darwall will argue, implies that a rational
principle is one that an ISIS would impartially choose or prefer all
agents to act on from behind a thick veil of ignorance a la Rawls. He
will further argue that a principle that exclusively counsels pursuit of
informed individual utility, for example, Tw, would not be so chosen, but
that a different principle, or perhaps a range of alternative principles,
would.

There is not space to examine the claims just enumerated and the
arguments for them. I think, however, that prior to reaching the
question of what principles would be chosen from behind a thick veil of
ignorance, Darwall has already committed a fundamental error. To see
how this is so, let us suppose, for the sake of argument, that there is a
norm or principle N, not identical to Tw, which would be impartially
chosen by an ISIS from behind the thick veil. Let us further suppose
that there is an action A and an agent S such that (i) S oughtw (i.e.,
ought overriding per Tw) to perform A, and (ii) N requires that S not
perform A. Presumably, the very coherence of the dispute between the
informed rational egoist, also an advocate of a self-centered theory of
practical rationality, on the one hand, and those who would claim that
fundamental moral considerations and constraints can and do override
informed self-interest, depends upon this possibility. Our question, then,
is, what is S to do?

Let us begin with a somewhat tendentious counter-argument. How
could Tw, understood as incorporating all valid principles of relative
rationality, be regarded as anything other than a complete account of
practical rationality and as being finally authoritative on matters of
what to do? Just to read and understand it suggests the inconceivabi­
ity of the opposite view. And why? Consider Mr. S. He is deliberating
about whether or not to do A. Let us, on the assumptions made in the
last paragraph, formulate the following reason for him: "That A is an
action which is inconsistent with that action that is required by rational
norms, those principles which, inter alia, would be impartially chosen by
an arbitrary rational agent as principles on which all rational agents
ought to act." Indeed, we might flesh out this complex reason with all of
Darwall's characterizations of and arguments regarding the ISIS. Now,
either this reason will sufficiently, perhaps overridingly, affect S's mo­
tivation or not. If it does, then it will, perhaps by itself or with other
reasons, constitute under Tw a minimal L-set holding with respect to
actions A and not-A. In that case, adequate consideration will lead S to
choose to prefer his doing A or it won't. If it does not, then, under Tw,
it is false that S overridingly ought to do A. But on the supposition
given, that is impossible, for we earlier assumed that it was true that S
oughtw to do A. If we drop the latter assumption, and give up the pos­
sibility that there could be an action A and agent S of the type de­
scribed, then we, or Darwall, must give up the incompleteness thesis, for
now we may view Darwall's account of the ISIS as constituting merely a
specification of a very complex type of reason which may hold with re­
spect to a given action and of which Darwall is making the substantive
claim that this reason, where applicable, would be of overriding motiva­
tional effect for all rational agents. But whether or not this reason itself
would have that sort of effect for all agents under conditions of ade­
quate consideration has not been established, and we may, in any case,
find that there are far fewer rational agents, indeed perhaps none, in the actual world.

So let us assume that S would still choose to do A, that action required under \( T_w \), even upon adequately considering the complex reason formulated above. Here, faced with the same choice, an ISIS would not choose to do A. Thus, since an ISIS is an "ideally rational agent", then S, in choosing to do or in doing A, is doing something irrational, or perhaps, we must conclude that S is not subject to the full RNS, since perhaps he is unable to conform his conduct and motivation to its requirements. In that case, again, we might be led to say that S is somehow irrational.

But now S is likely to become somewhat annoyed at this charge. Defiantly he may ask, "Why should I be rational?" Darwall, or a convenient ISIS lurking about, would doubtless respond that S just asked a meaningless question. By definition, what it is rational to do is finally authoritative. But why is it finally authoritative for S? S, we shall assume, is capable of conforming, and willing to conform, his actions to principles of relative rationality. It would be incoherent of him to do otherwise. S is further willing to acknowledge the "authority" of actions required under \( T_w \), for S realizes that if he does not at least attempt to do what he ought to do, then either he is being incoherent or he is ignorant of some fact, some aspect of reality which, were he but sufficiently to consider it, would lead him to choose a different course of action. However, what S cannot quite fathom is why he should further want to act as the ISIS does. S is therefore willing to buy part of Darwall's conception of practical rationality, nothing less and nothing more than that contained in the revised initial account, but is unwilling to accept the rest of it. So our question now is, why is an ISIS ideally rational, whereas S is only partly so? Why should S, or anyone, want to be or stand under a requirement to be an ISIS?

There is always the temptation to build into one's metatheory of the evaluative or the normative substantive judgments about what is good or to be done. I am inclined to think that Darwall has succumbed to precisely this temptation. To see how, consider what is different about the ISIS and S. The ISIS has a certain kind of motivation. He overridingly values his acting on principles on which all agents ought rationally and overridingly to act. Moreover, the ISIS, even under conditions of adequate consideration, may have other desires, preferences or interests, but these pale in comparison with his desire and motivation to be rational. The latter is, as Darwall says, "controlling". The remainder of his interests and intrinsic preferences are, almost, minor accouterments to his personality. But more than this, and crucially, his overriding motivation to perform or prefer any action by any agent is derived solely from his judgment that it is an instance of rational action, and this is to say that it is required by principles on which, as the ISIS judges it, all rational agents ought overridingly and rationally to act.

But precisely what have we just described? Is not the ISIS, the "ideally rational agent", one who overridingly values, above all else, rational action, whenever, wherever, and in whomever it occurs, as such? Because the judgment which by definition motivates an ISIS must be so sparsely impersonal, the ISIS cannot, for example, desire in whole or in part to be rational because that will further his informed preferences,
whatever they might be. No, he overridingly prefers rational action, on principles applying to and the same for all rational agents, per se, and under nothing more than that description, no matter how much the realization of that preference in particular cases might conflict with satisfaction of all or any of his other actual or informed preferences.

Now, however, we may restate our immediate question with a bit more force: Why is an agent with such a specific overriding intrinsic preference, one that will indeed commit him to overridingly valuing other agents acting rationally to pursue their interests no matter how much these conflicts with his own informed interests, to be taken as definitive of rationality? Is that not, rather, simply to build into the notion of rationality a substantive, but indirect value judgment? Why could it not be the case that an informed agent, who is in no respect himself irrational, have other intrinsic preferences that would override his preference, if any, that others act always, relative to their interests, rationally? If that preference, without more, can be assumed to be definitive of rationality, then why not an overriding preference for cultivating roses? But in that case, 'rational' has simply become a substantive term of evaluative or normative appraisal.44

But surely, Darwall will respond, the intrinsic preference built into the ISIS is special. In this context, I do not see that it is. It is a preference that, no matter what, all agents, including all other agents, act rationally. Why is my rationality dependent upon having that preference, especially when I have other rationally qualified preferences which in practice may conflict with this one? It is not here sufficient to respond to that rational norms are finally authoritative on the matter of what to do, indeed on the matter of what anyone is to do. From the "perspective" of any given agent, they may be, but from that it does not follow that I must rationally prefer another's doing what he rationally ought. To assume so is simply to assume the universal overriding normative status of the substantive intrinsic preference Darwall has built into the ISIS.55 What is rational for one to prefer may not be, on the level of substantive description, what is rational for another.

It might seem that this violates a quite sensible constraint Darwall accepts, viz., that the appraisal of something, anything, as "rational" presupposes a universal standard or principle applicable to all rational things, including agents. It does not. This constraint is no different than the like constraint regarding the appraisal of something as "red" or "square". One simply must have a set of general conditions defining the relevant predicate. But in that case, the revised initial account can constitute just such a universal principle or definition of practical rationality. As such, it can even be "normative" in that cognitively rational agents can "guide" their appraisals of actions, agents, whatever, as rational by it.56 What does not follow is that if a given agent correctly judges, for example, that another agent's impending action would be rational, then he is somehow rationally required overridingly to prefer that other agent's rational action. This is perhaps what Darwall had in mind by his insistent claim that practical rationality has an essential normative aspect, that is, that a condition of rationality for any agent is that, if he judges something to be rational for or relative to another agent, he is eo ipso committed to a motivationally positive, and as it turns out overriding, evaluation of it.57 But this simply assumes that 'rational' is, by its very content, a substantive term of normative appraisal. It may not be, but rather may be a theoretic term capable of
reductive theoretic analysis. If it is, that is not inconsistent with our positively valuing actions, in part or in whole, just because they are rational. Neither does it guarantee this.

I said in Section 6 of this paper that Tw, and indeed the entire revised initial account, entails no substantive value judgment whatever. It does not even entail that we oughtw to use the notion of the overriding 'ought' defined in Tw. Neither does it entail the negation of that claim. Now for Darwall, the phrase, 'rationally ought', is similar to the phrase, 'overridingly ought', explicated in Tw, at least in the respect that where it applies to an agent and action, it is "finally authoritative". But cannot one ask, regarding any particular conception either of standards or requirements that are claimed to be ultimate in practical matters, why they are ultimate? In asking this question, one need not be asking to be given a practical or motivating reason for, but rather a theoretical explanation of, the claimed ultimacy. In the case of Tw, there is a quite simple explanation of why what one oughtw to do is overriding or ultimate: Tw subsumes all possible types of considerations, facts, arguments, "perspectives", and so forth, within the limits of internalism. What, on the other hand, is the corresponding explanation for the ultimacy of that part of Darwall's conception of practical rationality which goes beyond (and might be inconsistent with the requirements of) the revised initial account? As I have said, it is the bare claim that rationality requires that one overridingly value any and all instances of rationality, whether in oneself or in another. However, even if our notion of practical rationality involved such a requirement, what prevents an agent simply from opting out of that part of the concept? All right, he might say, I'll be irrational in that respect. The charge of irrationality no longer has the claim on him that had under the revised initial account. We can no longer charge the reluctant agent, as we could in the latter case, with inferential incoherence or ignorance of matters that would be relevant to him.

Another way to understand what Darwall attempts to do in the argument we are considering is this. He wants to insure that every reasonable construal of the question, "Why ought I to do what I rationally ought to do?", is meaningless as a practical question about what to do. On the revised initial account, however, there are at least two ways to interpret this question, one on which it is, as a practical question, pointless, and is rather, a request for a theoretic explanation, and another on which it is not at all meaningless as a substantive question relating to practical matters. On the first, if an agent grants that it is true that he oughtw to do A, he cannot sensibly be raising the question of whether there are motivationally sufficient reasons for him to do A. He has granted that already and in principle we can provide them if he will but sit down and listen. On the other hand, the agent may be asking to be given those reasons, substantively specified, that will overridingly determine his motivation. Let us say that, so far, all he knows about the action in question is that he oughtw to perform it. That fact alone may or may not be motivating in itself, but at least its truth guarantees that there are motivating facts there to be discovered by him. "Why ought I to do what I oughtw to do?" may be viewed as a request to specify the substantive reasons that its truth guarantees, by an agent who is not (impersonally or otherwise) motivated solely by that fact, as described. The request that one be given such substantive reasons, without which it would not be true that one oughtw to perform a certain action, is not a meaningless request.
What Darwall has done is to incorporate in the ISIS the substantive and overriding motivation which would make the fact that one rationally or overridingly ought to do A itself a sufficiently motivating reason to do A, without more. For the ISIS, then, once he knows that fact about A, the request to specify the further facts, or substantive reasons, about A, is senseless. It might seem that one could view this preference as impersonal and so wind up with the conclusion that the ideally rational agent overridingly prefers rational action as such, no matter what, because, for such an agent, the underlying substantive reasons in any case are irrelevant or always of motivationally secondary importance.

We can perhaps tie the foregoing matters together by considering a final counter argument. Let us suppose that we have an agent S (one of whom the principles of relative rationality are fully descriptive) and an action A which he might perform or not. Either S has achieved adequate consideration (per Tw) with respect to A, or he has not. Suppose first that he has and has chosen to perform A. Now, either an ISIS would, upon rationally considering the matter, perform A or not. In case the ISIS would perform A, there is no conflict between the requirements of Tw and the RNS. Again, as I argued earlier, if this holds true for all agents S and all actions A, then the assumption with which Darwall began, that there could be conflict between self-centered theories of practical reason and moral requirements, derived in part as requirements of the RNS, is false. So we shall suppose that the ISIS would elect not to perform, or prefer that S not perform, A. By hypothesis, no further consideration of anything will alter S's choice. As to him, the requirements of the RNS are a dead letter. Hence, in cases of normative conflict, and whatever Darwall or anyone else might derive from the choices of an ISIS behind a thick veil of ignorance, they are, as to any action and any agent under conditions of adequate consideration, practically irrelevant.

Regarding matters of actual practice, then, Darwall must take himself to be addressing cases in which the agent has not achieved adequate consideration (per Tw) with respect to the action in question. In other words, Darwall's only relevant audience is the class of those agents who are relatively ignorant of some motivationally relevant truth. Let us suppose that S is such an agent. Again, if what an ISIS would do and what S would do were he to achieve adequate consideration with respect to the action in question coincide, there is no conflict, and in practical effect, no difference between the requirements of Tw and the RNS. Let us suppose then that S would perform A under conditions of adequate consideration, but that the ISIS counsels not performing A. What practical argument might the ISIS here make to S? Let us suppose that the ISIS tells S, and S comes to believe, that not doing A is required by the RNS. We may further assume, to make matters interesting, that S carefully considers this fact, and the arguments supporting it, in all their ramifications, diligently and vividly.

Regarding S, there are still again two alternatives. Either he is thereby (now) motivated not to do A, or he is not. If he is, then it is essential that he remain ignorant of at least some facts about the world that would be relevant to him were he aware of them. Otherwise, he will achieve adequate consideration and, by hypothesis, will choose to do A. On the other hand, if, as is certainly possible, S is not motivated not to
perform A, what is the poor ISIS to do? There are only two possibilities, short of coercion. The ISIS might reveal just enough of the truth so far unknown to S to change the latter’s motivation so that he will then choose not to perform A. Again, the ISIS must not reveal all of it. But there might not be any such partial truth to reveal short of bringing S into conditions of adequate consideration with respect to A. In the latter case, the ISIS will have to, what shall we say here, feed S some false information in order to achieve the desired result regarding S’s motivation.

It seems, assuming there is any conflict between the requirements of Tw and the RNS (an assumption which itself may be incoherent), that in actual practice, and with regard only to relevantly ignorant agents, those agents, or the appropriate "guardians" of the requirements of the RNS, will at a minimum have to regard some motivationally relevant truth as intrinsically off limits, i.e., even in the context of a potentially global consideration of all or enough such truth. At the worst, such agents will have to receive, and appropriate guardians perhaps propagate, falsehoods in the very context of practical (rational?) deliberation. That these measures are or could be, absent special circumstances, normatively acceptable, I call the doctrine of Poisonous Truth. That Darwall is stuck with it is beyond question.

But what, finally, of practical rationality and morality under the revised initial account? Practical rationality may be thought of as having a formal component and a substantive component. The former may in turn be thought of as a function and the latter as a type of defined input to that function. The function is the operation of valid deductive inferences, epistemology and the principles of "relative rationality", including the principles of decision theory under risk. The substantive component, the input, are represented facts that would tend to affect motivation in the context of (actual or hypothetical) deliberations up to the limit that consideration of no further such facts will affect one’s motivation or choice. There are, correspondingly, two principal ways in which an agent may act irrationally. One involves performing an action where the underlying preference essentially depends on some inferential or conceptual incoherence, such that, were it corrected, the preference or choice would change. The other is for the agent to grant and indeed know that he ought to perform an action A, but still intend not to do A, and in particular, to be unwilling to learn of or consider the motivationally sufficient reasons he has admitted exist. He is irrational here, not in being logically incoherent, but in acting in a way he knows is not sufficiently premised on justificatory reasons.

Curiously, there is an asymmetry between the two sorts of irrationality. Irrationality of the first type can only be charged against an agent when he doesn’t know that he is engaging in it, for example, when the reasons determining or essential to his choice are implicitly contradictory. Once the contradiction is brought to his attention, and he understands its basis, he can no longer maintain the contradictory position. He may say that he does, but we have now lost all basis for saying either that he believes that p or that not-p at the same time. The second sort of irrationality can only hold of an agent if he is aware of it. In this case he must believe or be aware that he has good reason to know that he is not doing what he ought to do. An agent’s action is not properly characterized as "irrational" just because it is not in fact
based upon sufficient reasons, else there would be no difference between being irrational in this regard and being mistaken or uninformed.

But what about the stalwartly incontinent agent who is irrational in this respect? What about him? If he persists, he persists. Neither the present account nor Darwall's has any foolproof or magical way of talking him into proper action if he won't listen or consider relevant facts and arguments.

What then of "fundamental moral requirements", whatever they are? In particular, what if one wants to hold that these requirements are overriding, universally valid, and at the same time also wants to adhere to the constraints of internalism and avoid the Doctrine of Poisonous Truth? Does the revised initial account provide a way to make such a claim? Yes, and as suggested in Section 6 of this paper, there are several alternatives. First, such requirements may be viewed as identical with the requirements of Tw for any agent. In that case, however, the specific requirements of morality may be different for similarly situated agents. To remove that possibility, one need only add the assertion that for all rational agents similarly situated, the specific requirements of Tw will be the same. One's claim has now become in part empirical, and at best, given our current state of knowledge on these matters, epistemically tenuous. The third and last alternative is this. One may hold that the word 'moral' introduces a substantively, though perhaps not precisely, specifiable restriction on what is to count as a morally relevant consideration. Let this aggregate of specifications be designated, "M". Then, all that one need claim is that for all rational agents similarly situated, adequate consideration of enough M-type facts will in fact constitute adequate consideration of a mL-set such that the motivation thereby produced and the actual choice induced will be both overriding and specifically the same for all agents and actional alternatives to which M-type considerations are relevant. That too is in part an empirical claim, one even more difficult to hold and assert than that indicated on the second alternative. In any case, these are the only choices satisfying the four-fold requirement set forth above.

Of course, one might, despite the foregoing considerations, want to maintain some form of neo-rationalist metaethics, and this one might attempt to do by rejecting internalism. But the full story why one cannot reject it will have to await another day.

ENDNOTES

1 Cornell University Press, Ithaca, New York, 1983. Parenthetical references are to pages in this work. Unless otherwise indicated, italicized expressions are in the original.

2 My paper is primarily a critical work. I therefore do not expound or tout the many parts of this book with which I agree, for example, the excellent discussion and criticism in Chapter 10 of Thomas Nagel's, The Possibility of Altruism (Princeton: Princeton University Press, 1970).

3 I am convinced that a genuinely adequate metaethical theory is available. My larger concern here is to promote that theory, in this case negatively, by indicating how Darwall has misconceived and misapplied

Certainly, not all specimens in the neo-rationalist menagerie are of this type. For example, see Alan Gewirth, *Reason and Morality* (Chicago: University of Chicago Press, 1978).

Principles of relative rationality would include, for example, the axioms of decision theory and Kant’s principle of the Hypothetical Imperative. "The sense in which principles of relative rationality are uncontestable is as principles of coherence." (15) But for Darwall, such principles cannot constitute the whole of practical rationality, for there is a "... crucial distinction between what it is for an action to be consistent or coherent with certain ends or preferences, assumed as fixed, and what it is for ends or preferences to give an agent reasons to act on them. To mark this difference, we may call the principles in question principles of relative rationality, since they tell us what it is rational to do relative to holding the rationality of certain other conditions fixed." (15) I am quite in agreement with Darwall that decision theory and the Hypothetical Imperative alone cannot be all there is to practical rationality. For example, and I think Darwall would agree, one needs a sound analytic definition of what a substantive reason is. And it is certainly true, that the actual, and even intrinsic, ends or preferences of agents can be subject to rational qualification and criticism. What is not clear is that the additional theoretic machinery required to express such definitions and ground such intuitions will not turn on matters of coherence. There must at some point be a demarcation between those elements properly to be included in a general theory of practical rationality and substantive claims about what, in a specific context, it is rational for a particular agent to do.

Likewise, decision theory itself does not make claims about what actions, relative to a given agent or class of agents, will in fact maximize net expected utility. To make the latter sort of judgment, often difficult indeed, one would have to supply complex factual information, empirically ascertained, about the actual preferences and preference orderings of the relevant agent as inputs to the functions of decision theory. But again, the very definition of what a preference or preference ordering is, which might properly be expressed within an adequate decision theory, would turn on proper theoretic and conceptual analysis, on matters of coherence. And the resulting definition, and derivations therefrom, would also be a matter of conceptual and theoretical coherence. The relationship between a general theory of practical rationality, claims about the actual substantive reasons there are for agents to act and the theoretic definition of what a sound reason is may overall be structurally similar. And of course, it is not a proper part of general decision theory what the aggregate demand curve for crude oil is, nor whether the average American in fact prefers crude oil to jelly beans.

In Chapter 6 of *Impartial Reason*, Darwall undertakes a rather elaborate argument to show that decision theory does not constitute a
complete theory of practical rationality. Although I agree with this conclusion, the argument he presents is curious. In brief, decision theory assumes, according to Darwall, that preference orderings are transitive, that is, if one prefers A to B and prefers B to C, it follows that one prefers A to C. But, asks Darwall, what is incoherent or irrational about intransitive preference orderings (72), that is, where an agent prefers A to B, B to C, but prefers C to A? Decision theory therefore merely assumes that intransitive preferences are "contrary to reason" (67). "It may, of course, be replied that either 'preference' or 'rational preference' is simply partially defined by the transitivity condition, but that is hardly satisfying. It does not make it at all intelligible why transitivity of preference is a condition of its rationality" (71). Further, as Darwall points out, someone who acts on intransitive preferences over time can be turned into a "money pump" so that he winds up where he has started but has given up some further thing that he valued, not identical to A, B or C, and gotten nothing he valued more or as much in return.

The way out is to see that preferences and indeed entire sets of preferences are criticizable in terms of reasons. And in what such criticism consists is not accounted for in decision theory. "What does appear contrary to reason, then, is a set of intransitive judgments about the support of reason, all things considered: that, on the whole, A is better than B, B better than C, and C better than A" (70).

Two things are troubling about this argument. First, the money pump scenario, to have any application here, presupposes that there are some further things that the agent also de facto values other than A, B or C. It is not at all clear, given this, why the decision theorist cannot appeal to these further preferences and valuations, without also appealing to the notion of "a reason" and derive an overall preference that the agent not act on any of the previous three preferences over time, given that he already has either A or B or C. If this can be established, and I am not sure that it can, it is not open to Darwall to reply merely that it is a contingent fact that the agent has any preferences at all other than those in the original intransitive ordering, for in that case, the money pump scenario would never arise for such an agent.

But the more fundamental problem with the argument is this. Decision theory does not assume that intransitive preferences are irrational, if what is meant by this is that an agent who has intransitive preferences and acts on them over time is acting irrationally or incoherently. Indeed, there may be nothing irrational about a series of choices over time which express sequentially intransitive preferences. For decision theory does not rule out the possibility that an agent may systematically and cyclically change his mind. Today, if I have C, I may prefer giving it up for B and a little of something else. Tomorrow, if I have B, I may prefer giving up B and a little of something else to get A, and finally, the next day, I may prefer giving up A and a little of something to get C, and so on. I may or may not be acting irrationally. But Darwall's solution is subject to the same problem. Today, I may find C, all things considered, not as good as B, and will give it up. The next day I may find that B is, all things considered, not as good as A, and give up B and a premium, and so on. All that decision theory assumes is that preference orderings are instantaneously transitive. But what about an agent who has, at the same time, intransitive preferences? Putting aside the matter of his changing his mind over time, is his current preference
ordering irrational or not? The proper response is, I think, that the question does not properly arise. For it is not that an agent with intransitive preferences at a single time is irrational, but rather, that such an agent is a logical impossibility, given the concept of having a preference at a given time, or so I would argue. And it is a possibility that Darwall does not take seriously enough. (Oddly enough, Darwall indicates his awareness of this option much later in the book (213), but does not see its relevance to the argument of chapter 6.)

This account obviously follows closely that of R.B. Brandt, op. cit., and indeed, Darwall adds the requirement that one’s desires or preferences not arise or be maintained owing to "accidental association or stimulus-misgeneralization". "Were it the case that a preference would vanish, or its strength be greatly diminished, if a person [here quoting Brandt] 'repeatedly brought to mind, with full belief and maximal vividness, all the knowable facts that would tend either to weaken or to strengthen the desire or aversion,' then it ought to be discounted in an ideally rational decision [reference omitted]." (96)

I might add that Darwall's failure to notice the rather obvious correlation between the degree of normative justification for a preference and the extent of one's consideration of motivationally relevant reality is all the more curious when we recall that he earlier touted the initial account because it would reflect not only the motivational, but also the "normative" aspect of reasons, and it would do this by fleshing out "an ideal of rational consideration". (81) But it seems to me that if anything is "ideal" about the process of rational consideration as Darwall has explicated it, it inheres in the sort of correlation we have just mentioned. Darwall's emphasis, however, is almost single-mindedly on matters involving the resolution of inner conflicts and "integrations". And this seems, without more, insufficiently ideal. Suppose, for example, that whenever two of my intrinsic preferences were in conflict, I could attach an "integrator" device to myself, and by pressing a button, could weaken one of the preferences just enough to allow resolute decision and action. Would resort to such a machine also be ideally rational? Would the resulting dominant intrinsic preference thereby be normatively justified? Not obviously.

Interestingly enough, Darwall advances a similar criticism against Thomas Nagel for simply assuming that it is part of the very content of one's judgment when one makes a normative judgment that one is in fact moved to act, and of course, R.M. Hare notwithstanding, that is an odd sort of content for any judgment to have. See Darwall, pp. 127-31, and infra, Section 7.

Bear in mind that, per the initial account, the fact that there are no reasons against acting on a desire is equivalent to the fact that rational consideration of other facts would of itself generate no motivation against acting on that desire.

Note that none of these observations is inconsistent with Darwall's claim, nor his argument for it, that "... it is possible for a person to be moved by awareness of some consideration, without that being explained by a prior desire ..." (39)

I am somewhat loath to call what follows a "revision" of Darwall's initial account, and this for the reason that, in temporal sequence, Dar-
wall followed the work of Brandt and Falk in formulating the initial account.

13 A definition of objective value, or what it is for a thing to be, at least to some extent, good objectively and without qualification, can be developed along similar lines. Here, a significant difference in the requirements specified in Tw is that adequate consideration by a subject need cause only some, perhaps partial, positive response toward the object of evaluation or toward things of its kind.

14 Hence, Cognitive Naturalism in general, and definition Tw in particular, satisfy Thomas Nagel's requirement, specified in The Possibility Of Altruism, that any purported fact about the world must be capable at least in principle of being judged or represented (though of course not necessarily known) by any agent. We all in fact inhabit one world and share ultimately one reality. (But it does not follow from this that we may all judge the same fact using the same form of words. See infra, this section.)

15 I here say "more or less extensive" for the following reason. One can begin to go some way toward accounting for qualified evaluative and normative judgments by appealing to the theoretic device of a substantively delimited range of or restriction on the facts and conditions viewed as input to the hypothetical process of adequate consideration. For example, to judge that something is a good carving knife or a good method of torture is to make a qualified evaluative judgment and this in turn may mean that the qualifying expressions, 'carving knife' or 'method of torture', perhaps together with other factors implicit in the context of judgment or utterance, may serve to substantively specify, in this case possibly with reference to a specific function or fixed hypothetical end, a range of facts and corresponding restriction on facts which may be said to be "relevant" to the judgment in question. Thus for example, to judge that X is a good carving knife may in part be to judge whether adequate consideration of a minimal evaluatively complete set of "relevant" conditions holding with respect to X would generate or not extinguish some partial positive response toward X (or things of its kind) etc. Here "relevant" indicates facts or conditions tending to show whether or not X, given that it is a knife, would efficiently or effectively serve the "carving function". Nor need this mode of analysis be limited to range determination and restriction by way of functions or assumed purposes. For example, to claim that a thing is beautiful may be to claim that adequate consideration of its aesthetic or aesthetically relevant properties would generate or not extinguish some positive response toward it. To cash out the qualifier, 'aesthetically relevant', one might say that such features are those which can be noninferentially exhibited or manifested in one or more sensory modalities. What distinguishes the overriding 'ought', 'oughtw', and the unqualified judgment of goodness, is that they involve no such restriction on the kind of fact, property or condition that might figure as an element in an evaluatively complete set.

16 I should also point out here that definition Tw is not intended to express an adequate, or indeed any, semantic analysis of the word 'ought' in English. That I think has already been done fairly well by Roger Wertheimer in, The Significance of Sense (Ithaca: Cornell University Press, 1972). It is rather intended to express a "real", or theoretical-
ly adequate, definition of what it is for it to be the case that someone ought overridingly to do something.

17 Hence, the semantic structure of 'oughtw' is not entirely formulable in terms of truth conditions. Such judgments clearly can nevertheless have truth values. Moreover, it may be possible to capture the indicated pragmatic condition in terms of a further truth condition by appropriate reference to the relevant speaker and token utterance expressing his momentary judgment.

Another point. I have already said that P's judgment that another, S, oughtw to do something, A, logically entails that under conditions of adequate consideration, both P and S would prefer that S do A. But suppose that under conditions of adequate consideration, P would but S would not prefer that S do A. I have suggested that it would be semantically deviant for P, without more, to express his judgment of this fact by asserting a token of the sentence type 'S oughtw to do A.' How then might P felicitously express his judgment of this fact? By using a sentence token of the type: 'It would overall be a good thing if S did A.', or something similar. See infra, text of note 57.

18 Though I cannot fully explain this here, what, in pre-theoretic terms, this claim comes to is that a valid reason is some fact which under conditions of adequate consideration, could in some instance "tip" the motivational scales overall for doing a certain action. A sound negative reason, a reason against performing that action, is a fact that could tip a relevant agent's motivational scales the other way.

19 Indeed, on Tw, for it to be the case that S oughtw to do A, it is not even necessary that S would choose to do A were he to consider all relevant facts, where a "relevant fact" is understood to be some consideration that would affect S's motivation vis a vis his choosing A upon being "vividly and dispassionately" considered by S. That requirement is too strong. All that is really necessary is that S consider enough such facts such that adequate consideration of any further ones would not alter his choice.

20 Recall that A must be an action within the agent's power to perform. If at the time of performance, the agent is, for example, seized by an unexpected and total motor paralysis, this condition is not satisfied.

21 And certainly, regarding ordinary language, it is doubtful that this is so. I find nothing semantically deviant about the following utterance type: 'In times of extreme crisis, great men may have to lay morality aside for the good of the nation,' this uttered by way of justifying the actions of some contextually indicated "great man".

22 See note 15, supra.

23 Abstracting, of course, from considerations having to do with the shortness of time and the finitude of patience, paper and ink. Nor can Darwall's difficulty in this context be ameliorated on the grounds that he is working with a narrower reading of the word 'fact'. First, that is simply inconsistent with his own exposition of the initial account, especially that point in it where he claims that the rational consideration that would result, in a normatively justified preference, one with the strongest support of reasons, must include all facts the consideration of
which would tend to effect the agent's motivation. Second, in the work of Brandt, Falk and myself, of which Darwall was aware, no suitably narrow reading of 'fact' or 'fact about' will be found.

24 See Darwall, pp. 118, 121–22, 130–31; and Nagel, op cit., p. 90.

25 This may seem a bit confusing, and it is. Indeed, until we have an adequate semantic analysis of virtually every morpheme in English, it may be impossible to tell whether a given reason, canonically formulated, is subjective or objective. For example, consider the following reason-type: that X is red. Is this subjective or objective? Until we know whether the semantically nondeviant predicative use of the word, 'red', somehow involves implicit reference to the speaker then and there using it, one cannot tell.

26 Note that on a slightly deeper level, Darwall may appear to be right that personal intrinsic preferences can only be motivated by subjective considerations. And why? Because, on the initial account, reasons are reasons for acting; they must be considerations or facts about an action A. Hence, trivially, where the intrinsic preference is personal and is quite simply, "to perform action A", the formulation of A will contain a free agent variable referring to the subject, for example, 'I' or 'me', etc. But in that case, if any fact or consideration about A must implicitly or explicitly refer to A, then any such fact or consideration, reason, will be subjective. At least this conclusion will follow if we must precisely interpret the formulation of the preference in question and "facts about A" as indicated. (Interestingly, a similar consequence does not obviously hold in the case of impersonal preferences. I see no reason why only subjective considerations could not motivate an intrinsic impersonal preference.)

Unfortunately, the way out sketched above regarding intrinsic personal preferences is not really available to Darwall. For later, throughout the rest of Chapter 11 and Chapter 12, he talks about the value of things and states of affairs, and the considerations (reasons) supporting judgments of their value. But in that case, an intrinsic personal preference for a state of affairs will not guarantee that all or any of the considerations motivating that preference are properly speaking, subjective. Hence in general, the claim about the logical distinctness of the elements occurring in both definitions stands.

27 Darwall quite rightly notes the following: "Altruistic preferences may be either personal or impersonal depending on how they are understood. A preference that others flourish is personal, since we cannot impersonally judge the otherness of persons. A preference that people, or that Tom, flourish, however, is impersonal. Impersonal preferences are by no means limited to altruistic ones. General malevolence, or the specific preference that Harry suffer, is every bit as impersonal (in the sense of which we are concerned) as benevolence, whether general or particular." (133–34)

28 For some reason, Darwall neglects to formulate the rather obvious parallel claims regarding impersonal preferences and objective considerations, viz., that they "get no grip" on the world seen from a "personal perspective".
Two more less technical points. First, on the given definition, an impersonally basable preference might turn out to be a personal preference. Darwall just assumes that these preferences will always be, must be, impersonal ones. Second, there is just no such thing as considering, actually or hypothetically, all objective considerations "relevant to p's being the case", and this for reasons discussed in Section 5 above. So modifications will be required in Darwall's second definition, but let us suppose that they can be made along the lines suggested in our revised initial account.

However, I may find another agent's preference more understandable, more intelligible in some cases where it is based upon considerations that are subjective relative to him, e.g., "that X would harm me" [him], that parallel my own, "that X would harm me" [Kelley], than it would be were his preference based on objective considerations which are not also reasons for me.

Except perhaps in the following trivial and irrelevant respect. Thus one could say that there is an objective reason, say 011, shared by all five members of this community, where 011 is simply the conjunction of 01 - 09. But of course in that sense, every fact in the universe, or every representable fact subject to the operation of conjunction is an objective reason for any agent whatever if any fact is, or rather, there is relative to any agent, preference or action, only one objective reason, the aggregate of "objectively" representable reality.

How Darwall might avoid this unwelcome consequence is not entirely clear, yet there may be a way. Suppose that X is a state of affairs and has intersubjective value relative to the hypothetical community specified above. There may be no shared objective reasons among the members of S motivating a preference for X. But note that Darwall claimed that there would be shared objective reasons for acting. Hence, since by hypothesis each member of S prefers X, this preference itself would be shared by all members of S. Hence, because of this shared resulting preference, rationally qualified as it is, can one not say that each member of S has common objective reason to act so as to promote the state of affairs indicated by "X"? The trouble is that this way out of the difficulty does not easily comport with Darwall's characterization either of what it is for a consideration to be "objective" or to constitute a "reason". For as we have seen, reasons must be considerations, and in turn, considerations must be specifications of facts, things that could be expressed in 'that . . .' clauses. But what fact could here constitute a common reason for all the five members of S? For recall, reasons must be facts that would motivate an agent to prefer something or, here making a charitable expansion, to act. And surely, Darwall would not here wish to renounce, sub silentio, his earlier claim that, "In our initial account, the reasons that there are for agents to act are not given by the fact of desire or preference itself but by facts that motivate preference when dispassionately considered." (107)

One suggestion we can immediately rule out is that the common fact is really five different facts, viz., for A that he has a preference for X; for B that he [B] has a preference for X; and so on. For although these facts might move the members of S to action, they are not the same fact and neither are they objective. Then what is left? Perhaps the single fact that all members of S prefer X. But are we to understand here that what would really motivate the members of S to act so as to
promote the state of affairs indicated by "X" is dispassionate consideration of that fact, and not the motivation inherent in each of their respective original preferences for X? Not only does such a ploy seem gratuitous, but it might be incorrect for reasons discussed earlier in Section 4. This is to say that rational or dispassionate consideration of the fact that all members of S prefer X might not add any or enough additional motivation for any given member of S to that motivation he would already have in having the original preference. But let us suppose that these problems can be solved. Suppose, therefore, that it is plausible that the fact that would motivate action to promote X is the putatively objective fact that all members of S prefer X. Is that fact an objective reason for say, A, to act? It could be, but one doesn't see right off how it would have to be or even likely would be such a fact. For it seems that until any member of S himself realized that he was a member of S, the fact that all members of S prefer X may well have no motivational relevance to him at all. But if we must add the additional information just indicated to the so called "objective" consideration in question, we turn it into a subjective consideration. Finally, even were this not so, the artificiality of Darwall's notion is again manifest. For if shared objective reasons for acting can be guaranteed merely on the grounds that two or more persons share an objectively specifiable preference, then shared objective reasons might also be guaranteed if we eliminate the requirement that the preference must be motivated only by, at one remove, objective considerations.

33 As will shortly become apparent, Darwall likely has very much larger "communities" principally in mind.

34 Moreover, this conclusion cannot be avoided by claiming, as Darwall comes close to claiming (166–67), that the intersubjective values which hold relative to a community are partially definitive of that community. If that were true in general, then to specify a value for the relevant variable 'S' in any claim of intersubjective value, one would first have to determine whether that claim is true. In consequence, Darwall's proposed analysis of intersubjective value would be hopelessly circular.

35 See Sections 3 and 7, supra.

36 See Section 5, pp. 21–2, supra.

37 In this regard, see especially subdefinition (c), p. 22, supra.

38 Here I say "seem" because it is not entirely clear, especially regarding thesis II below, that Darwall wishes to make just these claims and takes himself to be providing arguments to support them in this Chapter. This is because at the beginning of the next Chapter, 13, he raises anew the question of whether taking these key attitudes or there being objective reasons to act holds for all rational agents. This suggests that either he really does not take himself to have established II in Chapter 12 or that he is going to offer a further argument to support it in Chapters 13–15. See also footnote 44, infra.

39 His detailed comments bear this out. We are sometimes "loath to suppose" that things do not have this sort of value (139); taking key attitudes involves in part "judgments" of intersubjective value (149); they are "implicitly assumed" or what "we believe" (150, 153); or that
things have intersubjective value is something "we suppose" in certain cases (155, 161). Nor, regarding thesis II, can Darwall simply mean that cognitive rationality is in part a function of having some concept of intersubjective value in the (here trivial) sense that such rationality is likewise in part a function of having a concept of number or of peat moss.

There is here no space to examine each of these arguments in detail. On my view, some seem acceptable once relevant qualifications are made explicit; others are rather hard to follow, and still others seem less than cogent.

Indeed, one wonders why it cannot be the case that I might respond positively to my courage, and yet overall respond negatively to yours, especially if your courage is negatively related to satisfaction of other intrinsic interests or desires that I might have. This specific point takes on no little measure of importance when we consider that the motivation, actual or hypothetical, inherent in preferences or which would result from the consideration of reasons must be subject to aggregate summation before its relevance to an agent's action, overt or internal, can be determined. As we shall see, the truth of a claim that there is objective reason to promote something, entails absolutely nothing about how agents, in actual or hypothetical circumstances, will behave or act and this because the projected partial motivation ingredient in such a claim may be summed out or overridden long before internal processing of information, coded either in terms of truth or compliance conditions, results in deliberate action. See infra, this Section.

Especially when we note that it is more than just abstract general preferences that are in question, but preference orderings and potential resolutions among conflicting abstract preferences across individuals and groups.

I do not wish to appear to be making a substantive case for an extreme egoism, for that too is an empirical matter. I am inclined to think that in fact human beings are and would be, under conditions of adequate consideration, altruistic to some extent, but not completely.

And it is, therefore, premature to suggest that we now see or would see the world evaluatively from a completely or absolutely intersubjective point of view. Not at all. Sometimes we might, but I am inclined to think that such occasions are likely to be rare indeed. Suppose that one Siberian nomad or even one Andromodean, upon adequately considering the matter, would not prefer the welfare of my loved ones. Would that tend to make their welfare, which I prefer, a matter of any less importance to me? Not likely. But suppose then that we enlarge the divergent group to include say, all the Siberians and half the Japanese, or all the Siberians, half the Japanese and 70% of Chicago suburbanites? No problem for me yet. Of course, at some point, as the divergent group begins to increase, I may begin to question my original supposition that I have adequately considered the matter, whether in representing and appreciating objective considerations, or subjective ones, or both. Yet, even if I do conclude that I have not adequately considered the matter, I may be wrong.

Darwall might respond that being "meaningful", which he characterizes as being "taken to be more important or serious", has nothing to
do with motivation or dispositions to act, and hence, that my counter­argument misses the point in that showing that an agent's preference could remain even though, as that agent knows, it is shared by no one else, does not establish that he views either that preference or its ob­ject as "meaningful". On the other hand, it seems misleading to contrast, without more, someone's merely "liking" something with his viewing it as meaningful, significant or important. For suppose, as one coherently can, that an agent has a deep, pervasive and abiding positive attitude to­ward a thing even though he correctly believes that it is not shared by others. It seems rather odd to conclude that this attitude does not, in­deed could not, lend meaning to his life. But even if we implausibly grant that it could not, so what? That it could not is a fact utterly in­relevant to the behavior and motivational dispositions of this subject.

At the beginning of the next Chapter in his book, Darwall says the following:

We may view Part III [which includes Chapters 10, 11 and 12] as having established three significant propositions. First, objective reasons exist for any being who is capable of being motivated from an impersonal standpoint. Second, we possess a concept of intersubjective value, a rough ac­count of which shows there to be objective reason for any member of a community to promote what has intersubjective value. And third, any being who can bear his own respect, esteem, or spontaneous concern, or see his life as signifi­cant, must also be capable of impersonal motivation. For any such being, therefore, there will be objective reasons to act.

(171)

Regarding these claims, the first is simply a consequence of Darwall's definitions (unproblematically interpreted). As stated, the second is problematic, as I have shown, but could perhaps be revamped so that it too is a consequence of relevant definitions. As for the third, it is also problematic and those ingredient claims which are correct are subject to a number of important qualifications. However, the point I wish to make here is that the text of Chapter 12 certainly gives the impression of claiming more than the three propositions just mentioned and of extrap­olating them further than this relatively innocuous summary would sug­gest. I therefore do not think I can be fairly accused of over arguing contrary views.

This is generally presented in Part IV of his book, which includes Chapters 13–16. There is certainly no space here to consider all the issues raised therein. We shall focus only on what is essential to Dar­wall's own argument, and not even all of that.

Assessing this claim will be all the more difficult because Darwall tells us very little about the fundamental requirements of morality he will seek to justify. They involve a prohibition against wanton cruelty (172), and what he calls a "conjunctive principle" of great abstraction and, as I shall later argue, some rather surprising omissions and per­missions. See infra, text of footnote 53.

See pages 176–98. Overall, I think Darwall's criticisms are correct, only that he does not go far enough in questioning the suppositions upon which these approaches are based.
49 Note that Darwall does not profess to abandon the initial account, but rather, as we have said, supplement it. The internalist definition of a reason presumably will form some part, but not the whole, of the concept of practical rationality.

50 Darwall's specific argument for this conclusion is itself singularly unconvincing. It rests on the premise that whatever one can apply the adjective 'rational' to is thereby implicitly referred to by the concept of rationality. Since a norm can be a rational norm, it follows that our concept of practical rationality implicitly refers to norms. On that reason alone, our concept of practical rationality would also implicitly refer to institutions, plans, ways to exercise, etc. The general form of the argument can be seen to be invalid by considering the concept of redness. Does that concept implicitly "refer" to Ferraris owing to the fact we can sensibly speak of a red Ferrari? Some of what Darwall will later say suggests that there may be nothing more to this claim, at least as regards norms, than that the predicate 'is rational' must refer to universal norms or standards in the sense that it must have some conditions semantically associated with it. But that somewhat trivial and quite correct claim will seem much too weak an interpretation of Darwall's view.

51 See specifically Chapter 14, Sections 6-9, and Chapter 15, pages 218-39.

52 To say that this veil is "thick" may be something of an understatement.

Whereas [Rawls'] original position is impartial with respect to human beings, let us imagine that our hypothetical chooser is ignorant of every fact about himself except whatever is involved in his being a rational agent. We may assume, therefore, that he knows that he has preferences, the capacity to revise his preferences in the light of information and experience, and, as per our initial account, the disposition rationally to realize his informed preferences, other things equal [footnote omitted] . . .

Actually, it is not sufficient even for Rawls's purpose that the veil be simply one of ignorance. After all, a person may be ignorant of desires and interests while these still affect choices. [footnote omitted] What the veil of ignorance is supposed to accomplish is not just an abstraction from information regarding individual preferences and interests but from their motive power as well. So suppose that those behind our thicker veil are both ignorant of any preferences they may have that are not common to any rational agent as such and immune to their motive force. This means that there is in effect only one chooser behind the veil: an arbitrary rational agent. (230-31)

Of course, in this case, it is not even clear that such an agent would have any choice or preference at all, but Darwall handles this by assuming that the agent at least prefers to pursue his informed preferences, whatever they are, within the constraints imposed by the RNS, whatever they will be chosen to be.
It might be worthwhile to summarize very briefly Darwall's argument for the conclusion, just mentioned, that Tw would not be chosen as a rational principle. Rational principles must satisfy the requirement (taken over from Gauthier) of "self-support", i.e., that it would be rational on such a principle to choose to operate with it. (198, 229) But recall that an ISIS is ideally rational and that its judgment that a principle is rational (or choice of a principle as rational) would be made from the impartial perspective of such agents. "The issue before us, therefore, is whether, were we to choose from the impersonal, intersubjective standpoint of a rational agent (considered as such), it would be rational, according to the principle of maximizing individual utility, to choose that all agents act on that principle." (230) That standpoint is the veil of ignorance already described. Thus the question is, "... would it maximize the expected utility of a person choosing behind a thick veil of ignorance to choose that all agents act on the un constrained principle of individual utility? It seems clear that it would not." (231, footnote omitted) And why? Because there are "Hobbes Situations" occurring in life (an instance of which is the single interaction prisoner's dilemma). In such situations, everyone's acting to maximize their informed individual preferences according to decision theoretic principles of relative rationality would result in their realizing less net utility than if everyone "constrained" his pursuit of informed individual utility by principles of "morally right conduct". (232) The principle which would, however, be so chosen would be the following:

If there are no other agents with whom cooperation or conflict is possible, then maximize informed individual utility.

If there are other agents with whom cooperation or conflict is possible, then constrained pursuit of individual utility by requirements of morally right conduct (for example, by the requirement of nonmaleficence). (232)

There are many questions to raise about this argument and many qualifications which Darwall himself makes. For example, he seems to admit in a long footnote that, under conditions of "perfect information", for example, including knowledge of the deep motivations of and principles on which other agents act, that pursuit of informed individual utility may "... dictate a decision to follow moral requirements ..." if others do so as well (193). (For different reasons, I shall shortly argue that Darwall can only be addressing agents who have not satisfied the conditions of adequate consideration with respect to the relevant action or range of actions in question.) Moreover, if others do not constrain their behavior by principles of morally right conduct, neither is one required to do so either. (243) In that case, however, which must foresee repeated interactions over time, we are no longer in the single interaction prisoners dilemma. Notice also that the derivation of the conjunctive principle likewise rules out certain preferences, for example, the overriding intrinsic preference to be uncooperative or maleficent. If an agent had such a preference, then either he cannot maximize his utility by constraining his maleficence or we must say that in his case there is no relevant possibility of "cooperation or conflict".

Finally, and as perhaps a definitional gloss on the relevant notions of "cooperation" and "conflict", Darwall seems to admit that if a particular agent had "an extraordinary power advantage" (232) over other agents, then pursuit of informed individual utility might always be
the rational course. Let me make two observations regarding this remark-
able claim. The first is that a partial functional analog of such an ad-
vantage might be constituted by relative advantages in power or in-
telligence and that given certain preferences one had, everyone's con-
straining individual utility might not make such agents better off than
if no one sought to do so. This would have to be determined on a case
by case basis. It is hard to see, therefore, how any simple principle
would be chosen from behind Darwall's thick veil of ignorance. The sec-
ond point is this. Darwall claims to show that fundamental moral re-
quirements, universally valid for all rational agents, are reflected in the
conjunctive principle. But suppose that I had such extraordinary powers
that I could, by threat, force people to yield whatever I wished and was
immune from retaliatory harm. I suppose for me, at least relative to the
human agents we know of, there would be no "agents with whom coopera-
tion or conflict is possible". Hence, once the veil was lifted, I could
selectively torture and brutalize other persons as I chose. This is, to
say the least, a rather extraordinary conception of the "fundamental
moral requirements" and corresponding permissions applicable to me.

54 In a footnote found in the passage where he describes the thick
veil of ignorance, Darwall says of the rational agents therein choosing
principles, "... we may suppose that such beings are also rational in
the sense that they are willing, and able, to constrain their pursuit of
informed preference by principles that would be chosen". (231) As ap-
plied to the ISIS, this follows by definition. As applied to ordinary ra-
tional agents, this is an empirical assumption that all such agents will
value one thing over everything else.

55 Notice that this particular intrinsic preference definitive of the
ISIS appears to be the one exception Darwall allows of a justified pref-
ERENCE which is justified, but not because it was produced by, and is
therefore a response to, rational consideration of substantive reasons.
Here also, as far as I can understand, is the respect in which Darwall
regards the conception of the practical rationality he develops to be de-
ontological: it does not depend on the informed responses of any being,
ideally rational or not. Here, the overriding preference of the ISIS for
anyone's rational action is not a response to anything. If that is all that
is involved in deontology, then the revised initial account is in part de-
ontological. But whether a response or not, the contingent character of
the defining motivation of an ISIS for all other thinking, deliberating,
knowing and acting agents is not removed; the issue is only recast. As
to them, it is surely a contingent matter whether they have the over-
riding preference of the ISIS.

56 This only shows that an account of the relevant sense of 'norma-
tive' solely in terms of the notion of "guidance" is worthless.

57 This claim, and several made in the preceding paragraph, may
appear to contradict my claim, in Section 5, paragraph six, that when
one person correctly judges that another ought overridingly to perform
a certain action, and that judgment is to be conceived as being overrid-
ingly normative for the person who made it, then he is logically com-
mitted to the further claim that he would prefer upon adequately con-
sidering the matter that the other agent perform the action in question.
This does not entail, however, that the second condition indicated in the
antecedent of the foregoing conditional cannot be negated. In the pre-
sent context, we are concerned with one person's judging or granting
that another person's action is rational. If it is either the case that the term 'rational' is entirely a descriptive term or has enough descriptive content so that it can coherently be used even though any similar conditions making it normative for the agent using it are negated, then its application by one agent to another's action need not commit the former on pain of his own irrationality to preferring such action. Thus S may, without oddity, judge that it is rational for K to perform a certain action, but not rational for him, S, to prefer that K do so.

It is important to understand that the fact or information avoidance at issue here is of a very special type. It is the avoidance of motivationally relevant information which might, at the margin, determine one's decision in the context of adequate consideration of enough motivationally relevant fact such that no further would make a difference to that decision. This is not to be confused with the sort of case in which one might avoid, consistently with Tw, certain motivating information because one has indirect reason to know that learning of it would, for example, precipitate action, prior to one's achieving adequate consideration of the alternatives in question, that would be inconsistent with the choice one would make had such consideration been achieved. Thus, I might have good reason to think that, at the time, had I more information, for example, about the precise details about how my child was murdered, I might thereby be moved to do the wrong thing; that is, what I ought not to do. I might rationally conclude that I ought to remain ignorant of those details. But such a judgment can only be sustained if somehow I have reason to know that my present inclination, say not to take the law into my own hands, is the same as that which I would have on adequately considering the matter and that the response I would have if I acquired the additional information is contrary to this. To formulate a relevant principle,

The Principle of the Motivic Sufficiency of Inadequate Information: One may (indeed ought to) avoid acquiring or further considering information if the effect thereof will generate attempted action foreclosing consideration prior to its becoming adequate with regard to that action, where the choice generated is contrary to that choice that would result under adequate consideration.

Indeed there are a number of principles of information avoidance consistent with the revised initial account and I have formulated these elsewhere. See my, The Foundations of Value Theory, forthcoming. This is, however, not the sort of fact avoidance Darwall is committed to, namely, remaining ignorant of motivationally relevant facts that would incline one to act as one would choose to act had one in a unified and global way confronted all or enough motivationally relevant fact. On that showing, some truth is even from a god's eye view, intrinsically poisonous or inimical for doing the "proper" thing.

It should be obvious that the sort of difficulty here facing Darwall can be generalized, with slight modifications, to cover any sort of account that claims overriding normative status for any principles on the ground that they would be rationally chosen under conditions of hypothetical or imposed ignorance, for example, that presented by Rawls in A Theory Of Justice (Cambridge: Harvard University Press, 1971). Thus, his assumption that members of a society will be able to conform
their actions to the principles that would be chosen in the original po-
sition may itself cover a multitude of theoretic sins.

60 Here I can provide only the briefest and most general sketch.

61 Note that, as Darwall rightly insists in his Chapter 6, the present account does not take actual intrinsic preferences as given or as the ultimate point of reference within the process of normative evaluation.