National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

## NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.



## AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.



Canada

UNIVERSITY OF ALBERTA

# ON REPLIES TO THE MORAL SCEPTIC

BY

AIDAN DONOHOE



A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Arts.

DEPARTMENT OF PHILOSOPHY

Edmonton, Alberta
Spring, 1993

ISBN 0-315-82008-X

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: AIDAN DONOHOE

TITLE OF THESIS: ON REPLIES TO THE MORAL SCEPTIC

DEGREE: MASTER OF ARTS

YEAR THIS DEGREE GRANTED: 1993



5818 52nd. Ave.
Ponoka, Alberta
T4J 1K4

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled ON REPLIES TO THE MORAL SCEPTIC submitted by AIDAN DONOHOE in partial fulfilment of the requirements for the degree of MASTER OF ARTS.

Wesley E. Cooper

Bruce A. Hunter

J. Cameron MacKenzie

Don Carmichael

Sept. 23/93.

## ABSTRACT

Whether moral reasons provide good reasons is an old question but still a live one. Invariably, answers to it have tendered some position respecting the relation between moral conduct and self-interest. Plato, for example, saw such conduct as according with the individual's "true interest", though the individual himself might not realize that it does. There are modern variants of this kind of reconciliation, e.g., those of Mackie, Kavka, and Gauthier. One distinctive version is proffered by Parfit where, against the prevailing inclination, the direction of reduction is from prudential reasons to moral ones.

Others hold that moral and egoistic reasons are, or can be, fundamentally opposed. How then to decide between them? One might adopt one of a number of views. One might hold to a rational egoism which champions reasons of self-interest in all cases where moral considerations and egoistic ones give conflicting direction. Or one might hold the contrary position that moral reasons are overriding, either always or, at least, usually. One might, that is, subscribe to one of a number of different views, e.g., that of Baier, or Falk, or Gewirth, or Darwall, that have in common the conclusion that moral reasons defeat prudential ones. Or, following Prichard, one might see the two kinds of reasons as incommensurable.

My concern here is to examine the adequacy of a representative sample of arguments that, in one way or another, defend the moral point of view. I should say,

perhaps, "seek to defend", as my conclusion is the (for me) disheartening one that none of these attempts are successful.

Although I remark on this in the text I wish to emphasize here that the prudential response which springs from Mackie's attempt to (partially) reconcile the moral and prudential viewpoints is an extension of that attempt which Mackie would not have endorsed. In fact it fails, in my view, precisely in trying to overreach his conclusions. Though I attribute it to no one, the response is not fictional but is, rather, a natural one and of respectable lineage.

I have too to acknowledge that the outline of Gauthier's view, which informs the "rational/prudential response", is culled from Darwall's brief description and commentary on that position as it is presented in Gauthier, 1975 and 1986. Though perhaps inexcusable for a scholar, taking the short route to someone's view may be permitted for one who is concerned to examine a view. Of course, should Darwall have not mentioned a feature of Gauthier's position that bears crucially on my criticism of it, the fault is entirely mine. I am, however, confident that such is not the case.

I find fault with each of the theses presented in this essay. Nonetheless, I have learned from each proponent and I hope some of this will show in what follows.

What follows begins with a dialogue. Although the conversational form is soon discontinued, I wish the reader to see the dialogue as carrying on uninterrupted; a dialogue where, if my arguments are sound, the moral apologist is forced to adopt new strategies, to appeal to different grounds, in his attempt to vindicate moral conduct.

TABLE OF CONTENTS

# PART I

# THE PROBLEM
# and
# FIVE ATTEMPTS
# AT A SOLUTION

CHAPTER 1

## SETTING THE PROBLEM

1. X: Why should I help someone else, or otherwise behave morally, when it's not in my own interest to do so?

Y: Because it is in our mutual interest for each person so to behave.

X: I know that, but what do I care about that, as long as it's not in my interest so to behave?

Y: There are rules overruling the dictates of self-interest such that it is in the interest of all alike that everyone follow them. Others have followed and others still do follow these rules and your current well-being is, possibly in large part, consequent on those facts. In refusing to do your share in return you may well be branded as unreasonable and a parasite.

X: I grant all that but I'm still not going to do what's against my interest.

Y: I'm not saying you will, I'm saying you should.

X: But you still haven't given me an acceptable reason why I should.

Such, in essentials, is the imagined conversation -- presented by John Hospers -- between a moralist and one who is seen as both incorrigible egoist and moral sceptic.1 What further can be said, Hospers asks, in this "maddening situation"? The fellow accepts Baier's characterization of a morality and, seemingly, also it's justification, and yet he persists in his demand for a reason, in situations where others are self-denying, that he be moral also. There is indeed such a reason, Hospers holds: we should be

moral simply because it is right; or with respect to a right act -- "...doesn't the rightness provide a sufficient reason for performing the act?"2

But -- as might be guessed -- the questioner is still not satisfied. Here, Hospers maintains, in admitting certain acts as right, yet continuing to ask why he should perform them, the sceptic is making an impossible demand. He is, it seems, refusing to accept anything but a reason of self-interest for doing what is, ex hypothesi, against his interest; and in persisting with this self-contradictory request he reveals himself as immune to rational argument and as, consequently, an obvious candidate for quite another form of intervention.

2. But is Hospers right? Must the questioner be seen as a recalcitrant, and irrational, egoist? Perhaps we could try an alternative construal of the dialogue before relegating him, as Hospers would, to the psychiatrist's couch. Rather than asking for a reason of self-interest for performing an act that is against his interest, the sceptic, in saying "I know this act is right but now tell me why I ought to do it", could well be seen as looking for something that goes beyond the fact that the act is the right one to do. He may want to know how its rightness supplies him a reason and, if it does, how he might compare that reason with the reason of self-interest against his doing the act.

That is, his demand can be given these roughly equivalent expressions. It may be that 1) he is asking for an explication of the force of the moral "ought"; 2) he is inquiring into the status of morality, of moral requirements and prohibitions; he is questioning, or asking to be shown, the authenticity of moral reasoning; or, 3) though he is asking still a normative question he is not asking an ethical one; that is, the question

"for what good reason should I be moral?" may be, pace Hospers, a question posed from outside of, not from within, morality.

3. In asking for a vindication of moral reasons in terms of something more fundamental it is not necessarily the case that what the questioner sees as more fundamental are reasons of self-interest.3 Moral scepticism need not reduce to a kind of refractory egoism. The sceptic could also have inquired (though he doesn't) into the status of prudential reasons; as to the force of the prudential "ought". The question "Why should I protect or advance self-interest?" is one put from outside of, not from within, the "science of prudence" and as such, like the meta-ethical question, does not allow of a question-begging answer. Granted this question is not often asked -- and should our questioner be open to criticism in this regard he is in very good company, both late and present -- it being generally assumed that reasons of prudence stand on their own feet. That is, "that it benefits, or will benefit, me" is usually simply accepted to be a reason for the subject (if not necessarily an overriding one) whereas that it benefits/will benefit others/another seems less impregnably a self-evident reason.

I am not saying that this assumption has gone unnoticed or unchallenged. Mackie, for example, notes that "peculiarly prudential reasons", that is, reasons provided by a desire which the agent knows that he will have but which currently he lacks (lacking also a present desire that future desires be fulfilled), lean on a "...concept of the identity of a person through time",4 a concept which, he holds, functions like an institution, personal identity not being "...absolute, as it is believed to be".5 So just as, when one admits moral reasons -- which do not depend on a present desire, e.g. "(that) I promised" or "(that) it will help those people" -- one must, to have a clear grasp of what

is going on, recognize that one is speaking within the institution of morality, so it is when one claims prudential reasons; the claim must be qualified in like manner.

This qualification with respect to moral or prudential reasons does not amount to a denial that there are such reasons. Rather it is to say that to claim such reasons is to speak within the respective institutions6 and to say that we are not logically constrained so to speak.7

4. Parfit, too, not only questions the self-evidence of prudential reasons but employs a strategy of (so to speak) squeezing the (classical) theory of self-interest between Present-aim and moral theories.8 That is, he tries to show that the arguments which champion a theory of prudence as the rationally correct theory vis-`a-vis some Present-aim theory leave it vulnerable to the challenge of a moral theory to that title and, conversely, should attention be turned to defending the Self-interest theory against the moral view, that it is in like fashion vulnerable to a Present-aim theory. The prudential egoist (he tries to show), in arguing for the rational eminence of a bias in one's own favour that is temporally neutral, himself generates a vis a tergo which drives him, if he is to be consistent, into the moral camp. This attack is then reinforced by a difficult (but perhaps potent) argument which is designed to annul the metaphysics of personal identity through time and which threatens thereby the underpinnings to any prudential theory of rationality.

5. One could, however, with respect to the range of reasons it is thought to exclude, overrate this challenge, most obviously in the form in which it is given by Mackie. For Mackie there are three kinds of reasons; the two previously mentioned, both

of which are qualified in the manner indicated, and a third delivered without qualification. A reason of this third sort is generated by a "...desire, or purpose, or ideal that [an agent] now has..".9 Now it is very generally the case that people have the desire that their lives go well over time; that is, for most people, such a desire, though it may well not always be front and centre, is almost always present. In such cases what we ordinarily conceive as reasons of prudence go through unchallenged.

In fact Mackie's "peculiarly prudential reason" is rather a rare bird as we can see when we look at its description; to wit, the exclusion of any relevant present desires. My use of "prudence", by contrast, does not intend this refined sense of the term.

6. The case with Parfit is rather more complex and though I wish to avoid the complications introduced by his account it is necessary to say a few words in order to do so According to Parfit the Instrumentalist theory (the Instrumentalist version of a Present-aim theory of rationality) makes this claim: "What each of us has most reason to do is whatever would best fulfil his present desires".10 Mackie's unqualified acceptance of the third kind of reason implies as correct a watered-down version of this theory, i.e. one where the word "most" is deleted from the claim. But as Parfit suggests -- I think rightly -- the Instrumentalist theory (and I believe this applies to the weaker version also) must give way to a Deliberative version, which in turn must give way to some Critical version of the Present-aim theory. But in many such versions prudence, both in Mackie's restricted sense and in the sense in which we unreflectively regard it, can survive. The challenge (here) to prudential concern is one which threatens to demote and not demolish it as a source of reasons. But such concern and the reasons it

generates now stand on a different footing, they being no longer justified by the (now) discredited strong claims of the Self-Interest theory.

Again, as with Mackie, I would like to gloss over such possible "refinements", e.g. by proclaiming my use of "prudence" etc. as being indifferent as to the source of justification. But such a move, which is at least plausible in the earlier instance, is insupportable here. Parfit's doctrine is deeply revisionist and, should it go through, would, by dint of dissolving the contrast between prudence and altruism, render superfluous much of the discussion in this paper. I will then, perforce, take advantage of the fact that the jury is out on Parfit and is likely to remain out for some time. His conclusions put "on hold" in this way, I can affirm prudential reasons/reasons of self-interest as (roughly) equivalent to self-regarding reasons, and should these be seen as rooted in a Critical Present-Aim theory I can assume it to be one which preserves the opposition between such reasons and other-regarding ones, should the latter exist.

7. The "prudential assumption", then, has not gone unnoticed nor unquestioned. But this is not to say that it is not a natural or easy assumption to make. Mackie, I think, would have agreed that it is more than a little difficult to talk outside of the institution supported by -- and supporting -- our concept of personal identity, and well-nigh impossible to opt out of it for any length of time; much more difficult and more nearly impossible, I should think, than similar behaviour with respect to the moral institution. Indeed many of the reasons that go through unimpeded on his account on the basis of present desire would seem to arise from the conviction, on the part of those who possess such desires, that personal identity is indeed "absolute".11 That is to say, many present desires (and the associated reasons) can be seen as issuing from that conviction. The

notion of personal identity is simply massively entrenched in our language and consciousness. Parfit, too, sees himself as bucking a theory (the Self-Interest theory) which "...has been believed by most people for more than two millenia".12

8. Furthermore, important attempts, both past and recent, to justify moral reasons (and the behaviour consonant with such reasons) have depended on the authority or legitimacy of reasons of self-interest. Hobbes is, of course, famous in this regard, while the works of Gauthier and Kavka, for example, bear modern witness to the power of self-interest as an engine of motivation and justification. Even Mackie who, as noted, doesn't swallow prudence whole, explains and justifies both morality as a public institution and individual moral behaviour on largely prudential grounds.

This is not to say or suggest that such writers must be blind to the apparent tension between moral reasons and those stemming from self-interest. Sidgwick13 saw "...an unresolvable tension between moral reason and the rationality of self-interest..". Baier holds that "...following the rules of a morality is not of course identical with following self-interest. If it were, there could be no conflict between a morality and self-interest and no point in having rules overriding self-interest.."14 The works of those just mentioned are of interest and importance precisely because they take into account this prima facie tension and conflict and try to resolve it. But in each case the direction of justification is from prudence to morality and not vice versa.15

Nor is the "prudential assumption" or, to be more cautious, the reflective acceptance of the force of prudential reasons, confined to those influenced by Hobbes. R.M. Hare, for example, who is not obviously within the Hobbesian tradition, holds that "...morality, at least for the utilitarian, can only be founded on prudence, which has then

to be universalized".16 Indeed (and this is simply to echo Hare) without such acceptance it is virtually impossible to see how utilitarianism could get off the ground.17

9. We can now get back to Hospers' questioner. Clearly he is some kind of moral sceptic but just as clearly he takes considerations of personal well-being as providing him with reasons. That is, he finds it natural to operate within the institution of prudence. I am tempted to say that we can see him as one who has not read Parfit or, if he has, is not convinced by what he read. But in light of what has been said some caution is needed here. Parfit's main target is what he labels the S-theory, a theory which claims that the "...rational agent must be biased in his own favour.." regardless of his desires in the matter and regardless of the cost to others.18 Clearly, if S is the correct theory, reasons of prudence stand on their own feet. But suppose that not S but some other theory is shown to be the correct theory of rationality and that this theory is, or endorses, some kind of moral theory. Reasons of prudence can still exist but now because (great) imprudence is immoral.19

It is in this sense that the questioner is non-Parfitian. For him putative moral reasons stand over against prudential ones and profess to compete with them and he would see it as not only unnecessary but as incongruous that he be offered moral reasons for prudent behaviour. But -- as I have stressed -- this is not to say that he is an incorrigible egoist. He wants it to be shown -- and he is open to being shown -- that there is a real competition here, that the moral opponent has some punch; assuming all the while that prudence has a formidable one.

There might seem to be a problem here. In assuming that considerations of prudence have independent force mustn't he be an S-theorist? But given that an S-

theorist holds that an agent must be biased in his own favour is there any real sense in which it can be said that such a one is "open to being shown...etc"? So given the claim that our questioner is "open to being shown ..." he can't be an S-theorist. How then can he hold that reasons of prudence stand on their own feet? In response it could be said that holding to the independence of prudential reasons does not commit one to holding such reasons as always overriding. It is not immediately clear, though, that this response is adequate. One would need to come up with an argument to show that defeasibility does not undermine independence and this our questioner might not be able to do.

What I take to be the most plausible construal of the conversation avoids this difficulty, suggesting that the problem arises from too strong an interpretation of the sceptic's position. The questioner says in effect: As a matter of fact I have no sceptical worries about considerations respecting my current and future well-being giving me reasons to act; it is simply prima-facie so very plausible that they do that I'm not in a sweat about it. On the other hand I'm not at all convinced that moral considerations do; they are not in same way plausible nor, so far as I can see, in any other way. And that is as far as my cogitations have taken me.

10. Now despite the expressed confidence with respect to the existence of reasons of self-interest there is no intransigent egoism here. Nor is the sceptic making the "impossible demand". He is not asking that it be shown that, in the situation in which he sacrifices his interest, in precisely that same situation he enhances it. But to say both of these things is not to say that a more comprehensive prudential response is excluded. Hospers has, however, prior to the presentation of the dialogue, considered a response

of this kind -- Plato's -- and rejected it. (I shall soon briefly describe two other versions of this response -- one of which is related to that of Plato's, the other not -- and indicate why I think them inadequate.) Given this context we can regard the prudential option as foreclosed. And since Hospers' moral option is also precluded, the sceptic (rightly) refusing to accept a moral response -- he is not putting a moral question -- there is a clear invitation here to go beyond (or behind) the moral and the prudential to something (and it is at this stage an open question if there is such) more fundamental, and (if there is such) to see whether and in what manner an appeal to this level vindicates prudence or morality to the exclusion of the other; or, if both are vindicated, to see how a decision can be reached in those cases -- and these are, of course, the crucial ones -- where the contending viewpoints provide conflicting direction. In short, and contrary to Hospers' suggestion that he is not one who "...can be appealed to by rational considerations ",20 the questioner can be seen as asking of his mentor to appeal just so, to (purely) rational considerations. He can be seen, that is, as requesting arguments of the kind provided by such as Gewirth, Nagel, and Darwall.

11. At this stage it would, however, be precipitate for us to proceed so; to consider, now, arguments of this kind. Though Hospers' criticism of Plato's argument (to show that moral action accords with the actor's "true interest") is cogent, Plato's is not the only attempt of this general sort and it may be that a prudential formula can be found to satisfy the sceptic's request for a reason for his acting morally. It seems preferable to first test the adequacy of the more mundane or "lower-case" responses, the responses that are, so to speak, closer to home. Only if these fail is there a real need for the moralist to defend his view in a more theoretical or speculative way.21

Of course in order to show that the moral apologist must go further afield one would have to show either that any prudential response must be flawed or show individually that each such response is flawed. My more modest aim is to point up, individually, the failure of two notable (kinds of) attempts and, thereby, to suggest the possibility that, in a like manner, any prudential response will fail.

Footnotes to Chapter 1.

1. John Hospers, "Why Be Moral?", 1961. In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970.

2. Ibid., p.745.

3. On the other hand, this view is not of necessity precluded either. The sceptic -- we must allow at this stage -- may be persuaded that his is a naive conception of self-interest and that moral action does indeed accord with enlightened self-interest. I shall, shortly, consider two attempts at such persuasion.

4. J.L.Mackie, Ethics, 1977, p.78.

5. Ibid., p.191.

6. an institution being -- loosely -- a widely accepted and supported continuing social practice; a complex of rules and appropriate behaviours, concepts, expectations, enforcements, condemnations and the like. See pp.80-82, ibid.

7. See p.82, p.100, ibid.

8. Derek Parfit, Reasons and Persons, 1984.
The Present-aim theory sees reasons as resting either on aims that the agent currently has, or on aims that under certain conditions he would have, or on aims that he has or would have and which are not "intrinsically irrational" (p.94). There are, then, three versions of this theory. On the first -- the Instrumental Theory -- any current aim provides a reason. On the second -- the Deliberative Theory -- only those aims that the agent would now have, were he thinking clearly and in full knowledge of the relevant

facts, provide reasons. On the third -- the Critical Present-aim Theory -- some aims and sets of aims simply are irrational and there may be some aims that are rationally required.

My subsequent approval of Parfit's preference for the third version over the other two is tentative only. One worry is this: how, absent ideal deliberation `a la Brandt, can an agent judge of an aim or a desire that it is rationally required? Parfit himself is tentative here. See his pp. 120-122.

9. Mackie, op.cit., p.77.

10. Parfit, op.cit., p.117.

11. Parfit observes that most of us have a non-reductionist conception of personal identity, holding that the fact of someone's identity over time cannot be reduced to more particular facts; and that most of us hold to what may be regarded as corollaries of that conception, e.g., that one's continued existence, like personal identity, is all or nothing, that personal identity must always be determinate, and that personal identity matters deeply. Parfit's intent in Part 3 of Reasons and Persons is to disabuse us of these and other, related, notions.

If Parfit is right then many present desires rest, at least partially, on false beliefs and to the extent that they do they provide only (what are called) subjective reasons, or, on the Deliberative Theory, no reasons at all.

12. Parfit, op.cit., p.194.

13. I'm quoting from Mackie here; op.cit., p.190.

14. Kurt Baier, The Moral Point of View, 1958, p.314. Quoted in Hospers, op.cit., p.743.

15. and this is not gainsaid by such as Mackie's quasi-moral approval of a certain degree of egoism -- "...we shall want egoism also as a moral principle" (Mackie, op.cit., p.173).

16. R.M. Hare, "Ethical Theory and Utilitarianism", 1976, p.119; in H.D. Lewis, ed., Contemporary British Philosophy, 4, 1976.

17. ante Parfit anyhow.

18. Parfit, op.cit., p.192.

19. See p.318 et seq. ibid.

20. Hospers, op.cit., p.746.

21. In calling the one mundane and the other theoretical I do not mean that the latter is more complex or more worthy etc.

CHAPTER 2

## A PRUDENTIAL RESPONSE

1. In the light of these more recent efforts to overcome what Sidgwick saw as an "unresolvable tension" between the moral and self-interested viewpoints -- efforts aimed at reconciling the moral to the prudential -- it is important, where the question is " Why should I do X? " where X is the act morally required and is one that is against one's self-interest to do, that we be careful how that last condition is spelled out. If X is such that doing it is tedious or painful or will incur a monetary loss or loss of time etc. and is against self-interest on that account, but circumstances are also such that one's not doing X is noticed by others and thereby issues in a greater (though indirect) loss to one, then obviously doing X is not only the moral thing to do but the prudent thing to do. An observation like this contains the seed to Gauthier's argument to constrained maximization.

Again where the agent is so constituted -- unlike Gauthier's "economic man" who enjoys a Nietzschean freedom from moral qualms -- that not doing X is more costly (e.g. because of feelings of guilt consequent on the omission) than doing X, then doing X is not only the moral thing to do but is what is in one's real interest (or, `a la Plato, one's "true interest") to do, even though, because of the loss of individual utility mentioned, one might be tempted not to do X. And again an observation such as this is suggestive of a general approach to a justification of morality that takes account of what are seen by many as pervasive and virtually inexpungable psychological needs and traits.

Neither spelling out captures the poignancy of the question. Rather what is required is that the question, "why should I do X? ", be put where X is both the moral thing to do and where doing X is overall detrimental to one's self-interest, and one knows both of these facts. Clearly here one has a moral reason for performing the act and a reason of self-interest against that performance, so one is not asking for either a moral or self-interested reason but rather asking how to decide between the two. One is saying "From the moral viewpoint I should do X, from the point of view of self-interest I shouldn't; which point of view should I adopt? Can you give me a reason, or reasons, to take one point of view in preference to the other?"

There is an important general approach to this question which purports to show that the agent either in virtually all cases faces no dilemma here or, on a different version, that he never does. A common feature of both versions is the relegation of choices of action in particular cases to a secondary position but the rationale for this, between the two versions, is interestingly different. In this chapter I consider one version, in the next chapter the other. As I remarked earlier I think both versions fail.

2. On the one -- derived from Plato and Aristotle, echoed in Hume and other sentimentalists, and finding modern expression in such as Mackie and, to an extent, in Gregory Kavka -- choices of action give way to the more fundamental "...choice of dispositions, of characters, of overall patterns of life".1 A disposition is virtuous if it is such that having it better enables one to achieve the good life than does one's having any alternative disposition. So, to use Mackie's example, courage is that virtuous disposition or attitude with respect to possible harms such that its possessor is more likely to lead a fuller life than the coward and is more likely to be successful in his

various enterprises than either the coward or the foolhardy. And a like characterization can be given of the other virtues, both non-moral and moral. For example, the moral disposition which consists in the tendency to keep one's promises, or the more inclusive disposition, integrity, are such that their exercise enables one to engage to one's benefit in cooperative endeavours.2

3. The case for the primacy of the choice of dispositions appears to rest on the notion of an individual's life being (to some extent) integrated and purposeful over time, a notion which at once accords with what we see as desirable or sensible, and which is, in broad strokes, an accurate description of how we are; viz., as governed by dispositions. Now, as Mackie points out, though they can be cultivated (and presumably lost, or become flaccid, through disuse) and so change and develop, dispositions are both fairly persistent -- they cannot be switched on or off at will -- and not too finely discriminating. In consequence "...not every choice in which they are manifested will be advantageous considered on its own".3 That is, the prudent man, who is prudent by virtue of his optimal dispositions, will be led, at least occasionally, to act imprudently. These occasions may be less frequent however than we might at first suppose, for although the agent's actions on many occasions may issue directly in a loss to his self-interest, because of his strong moral tendencies the moral course and the prudent course will, for the reasons touched on earlier, diverge less often for him than they would for a less moral person.

There is, admittedly, a paradoxical flavour to this. One might raise an eyebrow at the idea of the exemplary prudent man being prone to suffer feelings of guilt for an action that, absent these feelings, would issue in a boon to self-interest. The paradox, or

rather the air thereof, can be dispelled. There is indeed an instability in the notion of the rational egoist "choosing" and nurturing a strong moral temper with a view to advancing his selfish ends. But our exemplar should not be seen as prudential calculator. Rather he is one who in being sincerely moral has those dispositions that a theory of rational egoism would recommend. He exemplifies, in fact, the obverse of a familiar phenomenon, in that in striving for something other than his own happiness he is happy.

4. There is here, I think, a forceful argument for adopting the moral point of view, or, given the elucidation in the last paragraph, a prudential justification for an individual's having moral dispositions. I think Mackie is right when he maintains with respect to his own (fuller) argument that it does much to reduce the tension that worried Sidgwick.4 Nonetheless the argument allows, as Mackie admits, that "...moral reasons and prudential ones will not always coincide"5 and though this admission precedes the appeal to the (moral) kind of person one is and wants to be, an appeal which functions to reduce the gap, I think he would have to admit that, at least for some people some of the time -- in a "cool hour" and possibly, or even likely, at a fateful juncture in their lives -- the gap can be considerable.6 So although an answer has been delivered which advises adoption of the moral point of view in preference to the prudential one (and this for a prudential reason) it is not an answer to the question as the latter was intended.

5. This can, I think, be brought out even if we transform Hospers' questioner from (fairly) comfortable egoist to just a one as we've been discussing. He, well attuned to the moral viewpoint, is brought up short by a stark opposition. Given the conditions as spelt out earlier he knows (or is convinced) that following the counsel of prudence now will

ensure him a considerably happier life than would be his lot were he to be guided, as is his wont, by conscience, and, let us suppose, the moral transgression would be a proportionately serious one. Though the approach which stresses as primary the choice of dispositions answers an important question it is not this one. This man wants to know what to do now.

To repeat, the response offered fails to address the question asked. And it is easy to see why. One's taking up (or keeping up) the moral point of view cannot be vindicated on prudential grounds in any case where, ex hypothesi, the moral action results in an overall loss of self-interest.

6. But -- it might be asked -- why the expository exercise if the question is set up such that the purported prudential formula must fail? The point is to show that even if we allow the force of this formula (and not everyone would), if we concede that the eminent prudential choice for the individual in a society where morality holds sway is that of moral character, moral dispositions, this concession does not amount to admitting a full reduction of the moral to the prudential. And even if complete reduction is not claimed (Mackie, for instance, does not so claim) the feeling may be that since adoption of the moral point of view is championed and moral character is vindicated -- renegation from the opportunistic act being seen as an inevitable (if unfortunate) consequence of what is after all the best choice -- so also morality is vindicated altogether. But this is not so, or not so without further ado; and there are difficulties (which to me seem insurmountable and which I will discuss in a moment) attendant on that further effort.

7. Again it might be objected that in specifying the situation as I have done I have simply inserted "Sidgwick's tension" as a postulate and so begged the question. This would be a damning objection were it the case that all such cases (exemplifying that tension) were contrived, were what Nielsen dubs "Desert-Island examples",7 but it is not plausible to hold that they are or -- even less so -- that they must be. Consequently they pose a problem for the rationale underlying Mackie's first order normative theory, a theory as to which he is indifferent that it be called a rule-right-duty-disposition utilitarianism or rule-right-duty-disposition egoism.8 For in all such cases it is clear, even if we adopt the latter characterization, that the first order theory will guide us in one direction while its rationale will guide us in the other. The underlying theory, that is, is internally inconsistent.

Footnotes to Chapter 2.

1. J.L. Mackie, Ethics, 1977, p.192.

2. If it is the case that "Aristotle tells us that the well-being or eudaimonia which is the good for man is an activity in accordance with virtue" (Ibid, p.186. My underline) it may seem that my interpretation of a virtuous disposition as simply end-based is an unsubtle and possibly distorting one. To this I can only reply that my concern is with the use, in the present context, to which talk of dispositions can be put and indeed to which Mackie puts it.

3. Ibid. p.189.

4. It might be objected that the tension is reduced only for one who is exemplarily moral; that for one who is less moral the prudential and moral courses will more frequently diverge. But this is to miss the point. The latter would be better off were he such that the two courses diverge less often. The fault is with him and not with the moral/prudential theory.

It might also be pointed out that Mackie's position is open to empirical disconfirmation; for example, by the observation that moral men are not all happy (or successful) men, or worse and more to the point, the (putative) observation that some are unhappy because they are moral. The argument does, I think, fail to address many specific worries, couched as it is in the most general terms. How moral would the exemplary prudent individual be, for instance? I qualifiedly call the argument "forceful" in that it shows that for the non-specified or "average" individual (in ordinary circumstances) the moral point of view is the one the "adoption" of which is most likely to issue in his happiness (or success).

Incidentally, our exemplar should not be taken as moral in the pejorative sense where, for example, prudery would be a natural concomitant, or fanaticism for that matter. He is your ordinary genuine good fellow.

5. Ibid. p.190.

6. I'm not suggesting that Mackie would resist this. Indeed he suggests that we can reconcile ourselves "more easily" to the admission that prudence and one's (moral) ideals may sometimes clash by reflecting on the comparable relation between present desire and prudence; a point which, as I touched on earlier, Parfit fully exploits.

7. Kai Nielsen, "Why Should I Be Moral?", 1963. In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970, pp.763/764.

I have benefited from Nielsen's discussion of Hospers' answer to that query. Nielsen follows Prichard's line (to be discussed later here) on the non-rational nature of the choice between the moral and self-interested viewpoints when these are opposed (ibid., p.759). Subsequently, however, he makes too light of the problem.

8. Mackie, op.cit., p.200.

CHAPTER 3

## THE RATIONAL/PRUDENTIAL RESPONSE

1. In regarding a (true) charge of inconsistency as embarrassing to a theory we invoke a background theory of reasons and this leads naturally to a consideration of a second version of the general approach to our question, the general approach being the strategy of relegating (choices of) actions to a secondary position.

Consider again our moral exemplar. Being sincerely moral, if he were challenged to justify his behaviour, he would find it natural to give a moral justification; that is, he has not adopted moral dispositions for the sake of self-interest. Mackie's talk of the "choice" of dispositions smacks of the incongruous if taken literally. Rather we may more plausibly see the acquisition of moral dispositions as resulting from (pace Nietzsche) a fortunate combination of genetics and upbringing. Yet he has those dispositions the possession of which are, more than any other set, most conducive to self-interest.

This poses a problem for the rational egoist as that character is normally conceived. His predominant disposition is to "cooly" assess actions solely in the light of their impact on his well-being and this predilection is seen as barring him from the kind of co-operative endeavours that would benefit him. Short of undergoing a radical change of character -- brought on perhaps by hypnosis1 or, say, drug-induced -- the rational egoist, ironically enough, seems doomed to a less-than-optimal existence, and

this in virtue of his having the disposition most consonant with a belief in the paramount importance of self-interest.

We may either suppose that such a drastic character change is not technologically feasible or, supposing it is, that our rational egoist is averse to such a procedure either because, though he believes his paramount objective will be achieved thereby, he cannot countenance himself in the role of moral dupe, or, more defensibly, he is concerned that circumstances may change in an unforeseeable way such that being moral is no longer prudentially justified; and because he sees that he will, in any case, be predisposed to renege his advantage should the previously-described situation arise for him and he is not convinced that his apostasy here is prudentially justifiable.

2. On any of these suppositions a Gauthier-like solution may seem promising.2 The rational egoist's assessment and determination of actions and choices has proceeded in the light of standards provided by a particular substantive conception of practical rationality; viz., rational egoism. Even were his conception the correct one, he has been, in Gauthier's words, "...rational in only a restricted and mechanical sense".3 Struck, perhaps, by the aforementioned irony, he is spurred to reflect on the rationality of those standards themselves; on the rationality of that substantive conception. But how is he to assess that conception? Or how can he decide between conceptions if the only standards he might use are themselves provided by one or other of the different substantive theories; theories that differ precisely because they embody one rather than the other conception? Each theory purports to give ultimate standards of rational choice.

Gauthier's response is to invoke a more general or background theory of practical rationality; that is, he appeals to certain formal requirements that, it is plausible

to hold, any substantive theory of reasons must satisfy. One such is universality, that the theory applies to all agents, but, not surprisingly, this requirement provides virtually no filter at all on substantive theories. Another is a consistency requirement, that the theory provide coherent guidance in all circumstances, a demand which, I have held, Mackie's brand of egoism fails to satisfy. With respect to the theories Gauthier is concerned with however, viz., rational egoism and a moral substantive theory, there is considered to be still a stalemate, both theories being thought to pass the test. But a third requirement, that of self-support, is regarded by Gauthier as a tie-breaker.

3. It is important here to be clear on how Gauthier conceives rational egoism. His preferred term is "unconstrained individual utility maximization",4 a phrase which is intended to portray the rational egoist as one who in each separate situation will do what he sees as best for himself within the context of that particular situation; that is, as one who operates on the principle "always maximize individual utility".5 Also, what he sees as best etc. reflects the preferences he would have "were he sufficiently informed and reflective".6

Is this conception of rationality, this principle of action, adequate? Gauthier's test is to ask of a particular substantive conception whether it would be rational according to it to choose to act on it. If the answer is in the affirmative the conception is self-supporting, if not, not. And rational egoism -- on the above description -- does not pass this test, for to operate on that conception is to ensure one's exclusion from mutually beneficial arrangements with others, arrangements which have a large impact on an individual's well-being. What Gauthier calls "constrained maximization" is a superior conception in that, in advising agreement-keeping in such situations and

unconstrained maximization otherwise, it ensures for the individual who acts accordingly that he does as well where other agents are unconstrained maximizers and much better where they are not.

4. Remember how the criterion of self-support is motivated. The unconstrained maximizer is rational in only a restricted and mechanical sense; though "...a conscious agent, [he is] not fully a self-conscious agent, for he lacks the freedom to make, not only his situation, but himself in his situation, his practical object".7 By contrast, a fully rational agent is, on Gauthier's view, self-critical; that is, unlike our rational egoist (prior to revelation), he subjects his substantive principle(s) of action to critical assessment. He will be guided only by principles that he himself can critically endorse. In this way he is self-governing, autonomous.8

Plainly the fully rational agent cannot endorse unconstrained maximization and on the proviso that criteria other than those generated so far are not forthcoming and that there are but two players in the field -- rational egoism and some moral substantive view -- he must opt for the one which fulfils all available criteria, viz., the moral view; his conception of that being constrained maximization.9 And, we can presume, our rational egoist would, on reflection, undergo a suitable conversion.

5. This second attempt to justify action in accordance with moral principles is again flawed in that, in much the same way as the first, it is disconcerted by the moral sceptic's crucial question. But given the tenor of Gauthier's argument it is more perspicuous, in revealing the flaw, to borrow and adapt a distinction made by Hare where he is by way of defending utilitarianism against "intuitionist" charges.10

6. The distinction is between two kinds of moral thinking, Level-1 and Level-2, and the principles employed at the different levels. Level-1 is everyday thinking under the usual press of limited time, limited information, etc. and the principles it employs are suitably general and relatively easy to follow. Level-2 thinking is that by which the everyday principles are selected and/or judged as to their suitability, is that by which Level-2 principles of any required degree of specificity are generated in response to any specific problem or case, and is thinking that is conducted in ample leisure and full knowledge of the facts. This archangelic thinking, as Hare dubs it, will generate Level-1 principles with the highest acceptance-utility. Since these principles may vary in complexity across kinds of individuals when differences in intellectual sophistication, moral character, etc. are taken into account, let us confine ourselves to just the one individual and so the one set of everyday principles.

Hare makes two further distinctions. One is that between the (morally) rational and the (morally) right action: the rational action is that action which is most likely to be right (the right action being the one which does or would, in fact, yield the best outcome). The other is that between the good and the right action: the good action is that which issues from the optimal everyday principles and is most usually the rational action and, consequently, most usually the right action but it could be one and not the other of either of these, or neither of them.

Cases where the good action is not the right one can be of two kinds, one where it is the rational action, the other where it is not. So it might be that on occasion the good general principles deliver the wrong answer but the agent (though he might be

doubtful about the deliverance), because of the usual exigencies, is unable to determine this. Action on the principles here is the action most likely to be right and so rational.

Rather more exceptional circumstances can arise, however, where the deliverance is again wrong but where the agent, not beset by the usual limitations, recognizes this and yet because he is so wedded to his principles of action he cannot act against them. His action here, though good, is not only not right but irrational.

7. It is this last case which bears on the objection to Gauthier. Rendering Hare's moral terms in the prudential key gives us actions as prudentially "good", prudentially rational and prudentially right. Good actions are those in accord with that principle adherence to which more than adherence to any other principle yields the best overall outcome for the individual. The optimal principle is that of constrained maximization and it behoves the agent, given the condition of mutual transparency of motivation assumed by Gauthier, that he be firmly wedded to it. And indeed the fully (or ideally) rational agent, in that it is the principle he endorses, is so wedded.

Now face this agent with the sceptic's question. Put him in a situation where the moral (constrained) act is not the prudent one, where the moral and self-interested viewpoints collide. Here any loss incurred on performance of the immoral action, e.g., one's reputation as an agreement-keeper is tarnished or even destroyed, is more than offset by direct gain. Clearly the moral act is here neither the right nor the rational act.

8. How might Gauthier reply? His position, in the debate with Parfit over self-support, is interpreted by Darwall as follows:

...rationality of conduct is inherited from the rationality of character: conduct is rational only if it accords with a principle of choice by which an ideally rational agent would be guided. Since an ideally rational agent would be governed by C [constrained maximization], it follows that C is the criterion of rational action.11

So, even though he cannot hold the moral (prudentially good) act as the right one -- ex hypothesi it yields less individual utility -- he will hold that it is yet the rational one. That is, he will resist Hare's characterization of the rational action as the one most likely to be right, a characterization which in the present instance, because the agent knows the right act is the immoral one, ensures that the immoral act is the rational one.

9. But can he justifiably take this tack? As Darwall points out, and as is evident from the description of Gauthier's approach in these pages, the Kantian element is fused with (indeed subsumed under) an end-based view of rational character, the approbative term "rational" being conferred on an agent's character only if it is such that the agent's choices yield maximal individual utility overall.12 Primacy of character with respect to act and primacy of utility with respect to character is an unstable combination however. If utility is a measure of rationality of character then why isn't it a measure of rationality of act? The criticism available here is the same as that levelled against rule-utilitarianism;13 if utility is the measure of a rule then why isn't it the measure of an act?

Since an important function of Hare's distinguishing the everyday (rule) from archangelic (act) thinking is the welding of rule and act-utilitarianism into a coherent whole -- thereby defusing this criticism14 -- might Gauthier have like recourse? Might he agree that constrained maximization is a working (Level-1) principle? But to do this is to

concede the argument; it is to admit that in cases like the one under discussion the agent can, and should, do a spot of Level-2 thinking.

On the other hand, even if he resists assimilation to the Harian model and holds to the notion of the ideal rational agent as one who is critically self-reflective, subjecting even his substantive conception of rationality to critical assessment, it seems that this case calls for a re-assessment of the principle earlier endorsed. He can hardly hold, in the face of cases of this sort, that critical assessment, once done, is done for all time.

10. It might be protested that such a case will not arise. But not even under conditions of mutual transparency (unless in addition (other) agents have access to a crystal ball) is it plausible to suppose that, at least for some constrained maximizers at some time in their careers, there will never arise a situation where the opportunistic immoral act will ensure a better overall outcome. Such cases will of course be rare but this is precisely why the everyday principles, being of necessity general (as against specific), are not adequate to deal with them.

11. Another objection might be that, given the ideally rational agent is choosing a principle to guide him through his entire life, it is clear that the principle of constrained maximization (call it C) is superior to that of unconstrained maximization (call this U), and that forgoing opportunistic advantage is an inevitable consequence of opting for the better principle.

In attempting a reply to this one needs to distinguish two possible scenarios. One is where the agent chooses as Gauthier suggests and at a later stage finds himself in a situation where the moral and prudential viewpoints collide in the way indicated (let

us call such a situation X). He was indeed justified in choosing C in preference to U but now, in X, C is unjustified by the same criterion employed in the earlier choice -- the rational principle being that which by following it one maximizes individual utility. Following C here is, by that criterion, to directly defeat the purpose for which C was chosen.15

What has gone wrong? The mistake lies in thinking there were but two options in the original choice. C is superior to U but inferior to C` which advises action in accordance with C except when in X when one should act by U. For an agent with a history as described, following C here, in X, is indefensible rule-worship.

12. On the second scenario the objection is not so easily handled. Take our onetime rational egoist who, concerned that his life go for him as well as possible, is reflecting on a choice of principle. Let us assume that he has carried out the cogitations presented here and so sees his choice as between U, C, and C`, and that he is not currently in an X-situation. Can he endorse C` or would a disposition to act on it be (indirectly) self-defeating?

If the majority of agents in the vicinity were U-actors, the few C-actors might be willing to engage with him in co-operative endeavours, but warily, i.e., when they are confident that an X-situation will not arise for him. If, on the other hand, there are plenty of C-actors about, he is more likely to be excluded from the more important arrangements. It does seem that, insofar as choosing C` will likely lead to a reduction in his already slim chances of finding himself in an X-situation and to his partial exclusion in more civil society, he would be better advised to choose C as his principle of action.

But picture the kind of transparency of motivation that is required for this to be true.16 If we grant transparency it is because we allow that we can quickly divine an agent's motivational disposition from his actions, that though he might try to camouflage his abiding intent his actions will inevitably betray it. A U-actor will be quickly identified as such. Now consider the C`-actor. In all situations other than the rare X-cases his actions are those of a C-actor. How then to identify him? Perhaps by an air of calculatedness; but how to distinguish this from the normal caution exercised by many C-actors? Indeed, given the rarity of X-situations and the likelihood, should one occur for the actor, of its being easily noticed by him, he need not have any pronounced air. What seems to be needed is an awareness of others' mental worlds on a par with the introspective awareness that some maintain we have of our own; a burgeoned Huxleyean scenario. We can reject this post-Huxleyean storyline and hold that C` goes through, that our onetime rational egoist can endorse it in preference to C.17

13. The theory underlying Gauthier's substantive conception of rationality as constrained maximization fails in much the same way as does Mackie's rationale for a rule-right-duty-disposition egoism; by failing to satisfy the earlier consistency requirement. The rationale for endorsing constrained maximization is a rationale for abandoning it in X-situations. In that C` is the more rational principle -- on Gauthier's schema -- the opposition between the moral and prudential viewpoints is preserved. By this I mean not merely that the two viewpoints can on occasion give conflicting direction -- that they do is stipulative of an X-situation and I have claimed that an X-situation can arise for an agent -- but that the fact that they can cannot be accommodated by a theory which purports to (fully) justify moral constraints on prudential/rational grounds.18

Footnotes to Chapter 3.

1. See Parfit, op.cit., p.19.

2. I rely on Darwall's characterization of Gauthier's position as this is expressed in the latter's "Reason and Maximization", 1975. See Stephen L. Darwall, Impartial Reason, 1983, esp. pp.182-200.

Darwall again considers Gauthier's position, including the statement of it in the later Morals by Agreement, 1986, in Stephen L. Darwall, "Rational Agent, Rational Act", 1986, in Philosophical Topics, Volume X1V, No. 2, Fall 1986.

3. See Darwall, 1983, p.195.

4. See Darwall, 1983, p.194.

5. or, if we allow, with Parfit, that he may sometimes act on a more particular desire, as one who will never do what he sees as worse for himself; i.e., as one who is in every situation never self-denying. See Parfit, op.cit., p.6.

6. Gauthier, 1975, p.415. Quoted in Darwall, 1983, p.195.

7. Gauthier, 1975, p.431. Quoted in Darwall, 1983, p.196.

8. This is the Kantian strain in Gauthier's thought. He continues: "Although we began by agreeing, with Hume, that reason is the slave of the passions, we must agree, with Kant, that in a deeper sense reason is freedom" (ibid. p.431).

9. The principle of unconstrained maximization (U) is rejected because it lacks self-support; it is not rational according to that principle to act on it. Accordingly it is tempting to see the principle of constrained maximization (C) as being endorsed because it is self-supporting. This is partly what Gauthier has in mind and it is in harmony with the view of the fully rational agent as self-governing.

However, as we shall see, it is even more tempting to see C as endorsed because it better achieves the goal encapsulated in U, viz., maximal individual utility.

10. R.M. Hare, "Ethical Theory and Utilitarianism", 1976. In Contemporary British Philosophy, 4, H.D. Lewis, ed., 1976. See pp.122-129.

11. Darwall, 1986, p.48.

12. "We identify rationality with utility-maximization at the level of dispositions to choose. A disposition is rational if and only if an actor holding it can expect his choices to yield no less utility than the choices he would make were he to hold any alternative disposition." From Gauthier's Morals by Agreement, 1986, p.183. Quoted in Darwall, 1986, p.47.

13. by, for example, Smart in J.J.C. Smart and B. Williams, Utilitarianism: For and Against, 1973. See also Darwall, 1986, pp.50/51, where he employs this criticism in pointing up the "deep tension" in Gauthier's view.

14. See Hare, op.cit., pp.125 & 127/128.

15. This is not to say that C is directly self-defeating. Note that this direct defeat is crucially different to the way in which action on C issues in a less than optimal immediate outcome in non-rare situations.

16. Darwall protests that Gauthier's argument against U assumes "...conditions of perfect information about agents' motivations" (Darwall, 1983, p.197). I want to show here that a more forceful charge of this kind can be made against an argument against C`.

17. Suppose a possible world, the same as our's except perfect transparency holds. Here C is the better principle. Suppose now an X-situation arises. Action on C is rational according to C but irrational by the criterion by which C is judged to be better. Since we must take that criterion as fundamental we must see the ideally rational agent as having to act irrationally. A theory's dictating that an agent act irrationally on occasion would be seen by some as an untoward consequence of any theory of rationality. Parfit, however, argues that although the Self-interest Theory has this consequence this is no objection to the theory. See his pp.13-17, op.cit. We can grant this and yet hold that having this consequence is an objection to a theory, such as Gauthier's, which views ideal agency as critical self-governance.

18. Gregory S. Kavka is kin to Gauthier and Mackie in holding an instrumentalist view of morality. In his Hobbesian Moral and Political Theory, 1986, he propounds a rule-egoism both as a moral theory and as a system of rules guidance by which is most prudentially rational for the individual.

Like Mackie, he advocates a morality that is practical in the sense that it is not too demanding. As well, in his appeal to socio-biology to explain the (limited) altruistic side of our nature, he shares Mackie's mundane approach. Mackie, in fact, anticipates that appeal thus:

...moral sentiments ... will enable social groups in which they take root to flourish. Consequently the ordinary evolutionary pressures, the differential survival of groups in which such sentiments are stronger, either as inherited psychological tendencies or as socially maintained traditions, will help to explain why such sentiments become strong and widespread. Since evolution by natural selection is the standard modern replacement for divine providence...(op.cit., p.113).

Since (again like Mackie) he allows that it is not always prudentially rational to follow the (prudentially) best set of rules, he appears to concede my case.

For some tensions in Kavka's view which bear more than a little on our topic here, see Wesley Cooper's "Critical Notice", Canadian Journal of Philosophy, Volume 19, No. 3, September 1989, pp.491-508.

CHAPTER 4

# THE DESIRE TO DO WHAT IS RIGHT

1. I expect that, were he alive today, H. A. Prichard would look askance at these modern endeavours to show the individual utility of personal moral behaviour. With respect to the attempt of Plato (and that of Bishop Butler) to establish "a necessary connection between duty and interest" he gives voice (and substance) to the feeling that "somehow they are out of place, so that the real question is not so much whether they are successful, but whether they ought ever to have been made".1 I myself share this feeling2 while at the same time holding that were one of them successful our moral sceptic would be provided a response to his query which he could legitimately accept. That we are pulled in different directions by our intuitions here is evidenced further by the fact (that it is a fact I take from Prichard)3 that Kant too, "in whom perhaps we should least expect to find such a proof", tried to exhibit what, in one vein or mood, we might consider an unholy alliance of self-interest and duty.

2. Let us follow, very briefly, Prichard's reading of Plato's position. That position was held in reaction to a view put forward (according to Plato) by the Sophists (the intellectual forbears of our contemporary sceptic). They argued (and this is an unstated premise) that for something to be a duty the doing of it must benefit the doer, or, for an action to be really just it must be advantageous to the agent; that many actions which ordinarily we feel bound to perform issue in a direct loss to the agent, a loss without

compensatory indirect gain; and that consequently these actions are not just, or not duties at all. Our ordinary moral convictions must then be illusory, they inferred, and they went on to construct an error theory to explain why and how we came to hold them.

What Prichard finds striking about Plato's rebuttal is its concentration on the second listed premise; its exclusive attempt to show that moral action issues in personal benefit. Nowhere does Plato suggest that the presupposition expressed by the first is, or even may be, false, and his preoccupation with meeting the Sophists on their own ground, his failure to suggest the irrelevance of the second premise, suggests that he too shared that presupposition. But, if it is true that if an act is right it must be advantageous, then, Prichard argues,4 it is precisely its advantageousness which renders it right, and if this is so our ordinary moral convictions are undermined rather than vindicated.

3. But remarkable as it may seem that Plato held that (tacit) belief, and even though on the strength of the fact of that belief alone we seem to have ready explanation for his concentrating on the second premise, it is an alternate explanation -- by way of the attribution of another belief to Plato, one tied in perhaps with the first -- that is more germane to Prichard's thesis and to our discussion. Plato's concentration on the second premise, his concern to show that right action benefits the agent, can only be explained, Prichard holds, by a conviction on his part "...that desire for some good to oneself is the only motive of deliberate action".5 Now this conviction -- that a man is motivated solely by his own good -- entails that to have him do what is right it is useless to convince him merely that the action is right, that he morally ought to do it. Rather, to have him act morally one must convince him that moral action is to his advantage. And this, of

course, is what Plato tries to do. The attempt is misguided, however, in that the conviction on which it is based -- in its implication that there are no benevolent or disinterested actions or that "there is no such thing as moral goodness" -- simply will not stand "the test of instances".6

4. There is in Plato, then, a mistaken theory of human nature and, its corollary, the (false) doctrine that morality needs a sanction. The theory, according to Prichard, has two negative implications. One is that the thought of duty is motivationally void; the other that there is in us no desire to do what is right. One can discredit the theory either by showing that in the very thought itself of one's having a duty one has a motive, or by showing that there is in us a desire to act rightly, that we possess a "sense of duty". Prichard maintains (pace Kant) that, on examination, the first collapses into the second, for in having a motive we have a purpose and a purpose is "...that the desire of which for its own sake leads us to do the action"; and that Kant would likely have accepted this "... had he not had the fixed idea that all desire is for enjoyment".7

What we have here, summarized in a few moves (Prichard indicates the availability of a fuller vindication), is an argument to (what seems to me) a quite momentous conclusion. For not only is there a threat here to the force of a distinction which, one imagines, is crucial to Kant's argument to freedom of will or autonomy -- the distinction between motivating thought and desire -- but, to put it conservatively, the hint of a slide from deontology to either a value-based or (possibly) a welfarist consequentialism.8 It is not surprising that Falk should have attacked him here.

For the moment, however, we will stay with Prichard who holds that the resolution of the problem with respect to the motivation of good conduct lies in

maintaining, as it is most plausible to do given our experience, the existence of a desire to do what is right.

5. If we grant this empirical claim we must allow Prichard's success in disengaging self-interest and duty. But the separation itself generates a problem (which Prichard himself brings to our notice) in that a man can be said to have "two radically different desires", the objects of which are "completely incommensurable", and how can he choose between them, in those cases where they clash, given that they share no "comparable characteristic" by which he might judge one to be preferable to the other? Prichard's response is that it is "wholly inappropriate to speak of a choice" in these cases; that though an agent may come to act on the one desire rather than the other and so can be said to have decided what to do, he has not chosen the one act in preference to the other.9

This perhaps surprising analysis seems, initially at least, to lean rather heavily on a stipulative distinction between decision and choice. What can Prichard mean by it? Clearly in these cases there is a reason of self-interest to do one thing and a moral reason to do the other. What seems to be meant is that there is no way of weighing these reasons one against the other and that the decision, whichever way it goes, will not rest on a reason. That is not to say that it will be irrational; rather that it will be non-rational, that rationality can get no foothold here.10

This conclusion, discomfiting though it is, has a ring of truth to it. I daresay it is the experience of many a reflective individual that he runs out of considerations here, that he sees himself faced with making (in the frightening sense of the word) a "free" decision. If the experience is not misleading, if it is an accurate reflection or reading of

the facts of the matter, then an objective validation of (our) moral convictions is not available.11

6. The average man-in-the-street however, who takes his morality seriously, is likely to find the foregoing conclusion indigestible, for one of his convictions is that his moral convictions are well-founded. Nor of course will it be easily endured by the rationalist moral philosopher. But there is another way in which Prichard's position gibes "our ordinary moral convictions" and the sensibilities of the rationalist. What this consists in can be revealed by evaluating the adequacy of that position as a response to our sceptic's demand for an "acceptable reason" for his acting against his interest. And, as we shall see, the objection which can be levelled here encourages an alternative response, one which, should it succeed, provides too a compelling rejection of the view that in deciding between moral and self-interested behaviour we are faced with an existential choice.

7. In arguing for the assimilation of motive to desire (an argument aimed at justifying his strategy, as against Kant's, of maintaining the existence of disinterested desires) Prichard says: ...if we face the purely general question "Can we really do anything whatever unless in some respect or other we desire to do it?" we have to answer "No".12 And shortly afterwards he holds that (for an agent) "...his desire to do what is right, if strong enough, will lead him to do the action in spite of any aversion from doing it which he may feel on account of its disadvantages".13

Now if the "ought implies can" principle is coupled with the first assertion, which expresses a condition on the possibility of voluntary action -- "one can only if one

desires" -- it follows that "one ought only if one desires". Taking "one ought" as equivalent to "one has a reason to" it follows that "one has a reason to (do what is right) only if one desires (to do what is right). So if, as may be the case, the sceptic lacks the relevant desire, on Prichard's schema he lacks a reason to act morally.

And even if Prichard was sanguine on the topic of human nature -- supposing that the desire to do what is right is had by all and ever-present -- he can hardly have supposed (and he nowhere states) that this desire is always strongest, so, in those cases -- which we can plausibly assume are many -- where the desire for personal well-being is more ardent, he cannot hold, while holding to his second assertion, that the moral reason is overriding.14 So on either scenario the sceptic's rejoinder -- that he has yet to be given an acceptable reason for his acting morally -- is justified.

But even if the sceptic takes morality very seriously (scepticism often enough being symptomatic of concern) he is unlikely to find this kind of response appealing. He is not likely to be impressed by the accidental nature of the reasons he might have for performing his moral [illegible] -- that they be contingent on the happenstance of his having the appropriate desire(s).15 And in this he is at one with many of his adversaries. For although Prichard's strategy succeeds in rebutting the view that moral behaviour requires the sanction of self-interest it does so by replacing it with a view which implies that there is no real need for, and indeed no possibility of, moral action in the absence of the desire to do what is right. In short, Prichard's position fails to preserve what has been called (if I recollect rightly) the "peculiar stringency of the moral", a failure which may upset the moralist and (one kind of) moral sceptic alike.

8. Before proceeding to the view which this objection engenders -- the alternative response I mentioned earlier, to wit, Falk's analysis of moral obligation -- it would be as well to consider the use to which I have put the "ought implies can" principle. This consideration will serve to modify the criticism voiced above, and it will serve too, I hope, to point up the different aspects there are to the problem we're discussing, the problem of the vindication of moral conduct.

To begin with there is the issue of how the principle is to be interpreted. Frankena claims that it "...need not be construed as asserting a strict logical implication"; rather that it "may plausibly be understood" as saying that moral judgements presuppose/contextually imply/pragmatically imply "that the agent is able to act as proposed..", or as saying that there is no point to uttering moral judgements, or that it would be wrong to exercise them, in cases where the agent is or is thought to be unable.16

And a similar question can be raised as to the status of (iet us call it) Prichard's principle. Does "one can only if one desires" express a conceptual truth (is it analytic) or a psychological fact or a constraint on utterances of a certain kind?

Now Frankena is concerned to refute an argument to the (internalist) view that a judgement of moral obligation logically involves (affirmative) reference to the existence of motivation on the part of the subject of the judgement; or to the view that sees motivation as "internal" to obligation, by which is meant that the existence of moral obligation involves the existence of the requisite motivation (either occurrently or dispositionally) or, put another way, that one's being motivated is (at least partly) constitutive of one's being morally obligated.17 And it is perhaps important in that argument -- ought implies can, can implies motivation, therefore ought implies

motivation -- that the premises express logical implications.18 Indeed, I am inclined to side with Frankena's pushing the softer interpretations, in that to admit entailment here is to needlessly concede too much to the internalist.

The question arises then: how to determine what falls out when either or both principles (the premises of the argument) are given a soft interpretation? I want to avoid this question. I want instead to exercise the "ought implies can" principle (on the strong interpretation), not to argue a position, but to illustrate the different positions that can be taken.

Perhaps by expounding some related difficulties I can make this intent more precise. Apart from the problem of the logical/non-logical character of "ought implies can" the kind of argument we've been considering employs ambiguous terms. "Desire", "motive", "reason", "ought", and even "can" are each subject to different construals. For example (borrowing from Frankena), we can construe "ought" as indicating nonmoral or moral obligation, or indeed (and not pointed out by Frankena) normative obligation that is not moral.

Correspondingly, "reason" can be meant as "exciting reason", i.e. as referring to something with causal power; or it may be intended to refer to something that has justificatory force. Frankena dismisses (correctly, in my judgement) a number of anti-externalist arguments on the ground of their failure to make this "prima-facie" distinction.

9. With this distinction in hand let us re-examine the argument "One ought only if one can, one can only if one desires, so one ought (or one has a reason to) only if one desires". If we take the initial "ought" as expressing nonmoral obligation, or, more correctly, non-normative obligation, the antecedent of the first conditional reads "One is

obliged to, or one has to (or one has an exciting reason to)" and the conclusion reads "One is obliged to, or one has to (or one has an exciting reason to) only if one desires". That is to say the argument is about motivation and its conclusion is that there is no motivation -- that one lacks a reason (in the sense of motive) -- in the absence of desire.19 And, if we want to be sure that we are being fair to Prichard, this is how we should interpret it in criticizing his position. For Prichard was concerned not so much to provide a justificatory reason for one's performing a duty -- presumably he thought that an action's being in accordance with duty is justification enough for one's doing it -- but with showing (against Plato) the availability, independently of one's egoistic desires, of enabling reasons for that performance.

And this is where Falk initially criticizes him. The position against which Prichard set himself, it should be remembered, was that which sought to show a positive relation between duty and self-interest but where the duty was determined independently of that relation; where it rested on, for example, the fact that one promised. And this amounts to the view that "morality needs a sanction".20 But while Prichard is successful in rebutting the specifics of that position his own view, Falk argues, is not sufficiently distanced from it. For on this view, though someone admits to a duty, "...there is no convincing him that he has a motive except by considerations additional to those which already convince him that he has the duty".21 In short it is yet the view that morality needs a sanction, though "...admittedly a more respectable one" viz., one's disinterested desires. It follows too that when one has a duty that one has a motive is not a "foregone conclusion" but rather an "...open question, turning on a benevolent dispensation of providence in the shape of a singular psychological disposition."22

10. There is, however, another sense in which we can take the initial "ought" in the argument which begins with "ought implies can", a sense that is not (overtly at any rate) motivational or psychological but rather normative, and given that the moral "ought" is accepted, that the question as to the existence of a moral duty is settled, we may see this initial "ought" as superordinate to the moral one and as serving to qualify it. The conclusion then reads "one rationally ought (to do what is right) only if one desires".23 Again morality is seen as in need of a sanction but there is a different emphasis here in that what is in abeyance -- and what is contingent on one's desires -- is not a reason in the sense of motive, in the sense that one is psychologically obliged, but an authoritative or justifying reason. There is an appeal to this (rational) "ought" in my earlier assessment of Prichard's position as a response to the sceptic and it is a feature too of Falk's criticism of Prichard that he pictures him as seeing disinterested desires as providing a justification as well as a motive for one's acting as morality dictates.

But, as I suggested earlier, it may be inaccurate (and unfair) to attribute such a view to Prichard. If we wish to criticize him in this vein it must be not for what he said or for what is implied by what he said but for what he left unsaid. As a response to the sceptic his position is unsatisfactory in two ways. First, absent disinterested desires (or desires of sufficient strength) the sceptic is powerless to do what duty bids him do; this is the charge we have already dealt with -- that morality requires an additional psychological sanction. Second, though he does not see the authority of the moral "ought" as derivative of this addition but rather accepts duty as bearing its own authority, this acceptance does nothing to satisfy the demand that the (alleged) force of the moral "ought" be explicated, that it be shown that moral reasons really do have justificatory force. Prichard, while providing an answer as to how it is possible to do one's duty, has,

it seems, simply assumed that the answer as to why has already, in the very idea of duty, been provided for.

11. There is yet a third way in which we can construe the "ought" in the argument and this is to see it as a moral "ought", notwithstanding it is already accepted that the subject has a moral duty and perhaps indeed is aware that he has. The conclusion now reads, "One morally ought only if one desires ", and on the assumption that the subject lacks the relevant desire it follows that it is not the case that he has a duty here. We may then see this as a reductio ad absurdum of Prichard's position provided, of course, that the moral "ought" is used unambiguously throughout. But need it be? Might it be rather that the duty assumed at the outset is viewed as a demand from without, that the "ought" is the externalist one, while the duty in the sub-argument starting with "ought implies can" is viewed as emanating from the "voice of conscience"; that is, that the "ought" disqualified by the absence of desire is the internalist one?

It is in fact, Falk avers, just this ambiguity in the terms "ought", "duty" and "obligation" which infects the thinking of the ordinary man and the moralist on these matters, and which inspires the attempts of Prichard and his opponents to demonstrate the efficacy of the moral "ought" while simultaneously vitiating their solutions.

Footnotes to Chapter 4.

1. H.A. Prichard, "Duty and Interest", 1928. In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970, p.690.

2. as does Falk, who expresses himself precisely thus and as, I suspect, do most of us.

3. Ibid., p.690.

4. The argument is on pp.694-696, ibid. I find it convincing.

5. Ibid., p.700. The argument runs from p.698. Prichard cites evidence for Plato's having that conviction from the Republic and the Gorgias.

6. Ibid., p.702.

7. Ibid., p.703, his italics.

8. See Darwall, 1986, footnote 2, pp.53/54, for a nicely condensed account of these different sorts.
On reflection, I wonder at the justification for my suggesting a hint of a slide etc. I will leave the suggestion in on the slim chance that there is something that I saw then that I fail to see now. No argument of mine hangs on it.

9. Ibid., p.703.

10. Nielsen is of like mind. For the individual faced with adopting a personal guide to action the moral and self-interested viewpoints seem to constitute "two strands of discourse...with distinct canons of justification". As to the choice between them, "...we just choose and there can be no reasons for our choice" (Nielsen, op. cit., p.759).

11. Nor, for that matter, is prudence objectively vindicable either. I use "objective" in something like (but perhaps not identical to) the sense intended by Mackie, that is, as suggesting an authoritative standard (somehow given rather than invented) against which human activities can be measured as to their worth, their rightness or wrongness.

12. Prichard, op. cit., p.703. See G.C. Field's "A Criticism of Kant", 1921, in Sellars and Hospers, 1970, pp.704-707, for a like view.

13. Ibid., p.703. That this doesn't square with the claim of a non-rational deciding in these matters is indicative of some unease on Prichard's part. Either the objects of the opposing desires are commensurable after all or the absence of a common measure interferes not at all with our behaviour, our action being in accord with the stronger desire.

14. This does not follow from the second assertion as explicitly stated but a natural reading of this, given the context, is that strength of desire is not only a sufficient but a necessary condition on action. This reading, "one can only if the desire is strong enough", with "ought implies can" yields "one has a reason to only if the desire is strong enough".

15. Picture for a moment the sceptic as one who experiences the moral impulse but who is worried as to the justification for his acting on it. Told that the justification lies in his having the impulse, he may feel that the moralist has not met the challenge; he may hold, with e.g. Parfit and Darwall, that desires or impulses to action are criticizable.

16. William K. Frankena, "Obligation and Motivation in Recent Moral Philosophy", 1958. In Essays in Moral Philosophy, A.I. Melden, ed, 1958, p.60.

17. As we shall see later, the latter is probably the more perspicuous phrase.

18. I say "perhaps" because although Frankena can insist on a logical sense of "implies" in an argument about judgements of obligation it is not clear, on his own view, that he can when the argument is about the existence itself of obligation. That is to say, his emphasis on the distinction between (in his case, physical) fact and logical necessity (see his example p.60) suggests that logical entailments, though (possible) properties of judgements (or statements), are not properties of facts. Frankena is concerned here to rebut an argument to the view that "the state of having an obligation includes or is identical with that of being motivated in a certain way" (ibid., p.58), the view called "existence internalism", but his rebuttal is directed at a judgement internalism (again, see p.60).

19. Should the reader find this use of the principle a bit odd, even strained, I ask him to view it as heuristic, as serving to point up a distinction which is easily lost. And even if the derivation be deemed artificial (and non-rigorous) the conclusion itself is a fair representation of Prichard's position.

See Mackie, op.cit., p.73-77, on how "...different uses...of "ought" introduce different kinds of reason".

In Falk (pp.506/507. See footnote 21) "ought" drifts towards "must", but here the "must" is not merely motivational but obligatory also in the sense of being "rationally necessary".

20. Prichard, op.cit., p.701.

21. W.D. Falk, ""Ought" and Motivation", 1947/48. In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., first edition, 1952, p.497.

22. Ibid., p.499.

23. This might be seen as a dictum derived from a Self-Interest theory of rationality via a Desire-Fulfilment theory of self-interest. But one's desires could be (predominantly) other-directed. Should classification be based on whose desires they are that are to be fulfilled or on the objects of those desires?

CHAPTER 5

## THE VOICE OF CONSCIENCE -- AN INTERNALIST RESPONSE

1. Prichard's complaint against the views of his opponents, you will recall, was that they fail to vindicate our ordinary moral convictions. So he censures Plato and Butler on their assumption that we do our duties for none but interested reasons. In contrast, for Prichard, though justification and the enabling motivation are separate, they can be brought together through the existence of disinterested desire. His position, then, provides a mixed conditional response to the sceptic -- you are justified in doing what is right and you can do it provided that you have the desire.

But against this, Falk insists, our "ordinary moral thinking" involves the conviction that "the very fact of duty entails all the motive required".1 One is reminded here of Hospers' categorical retort with respect to a right act -- "doesn't the rightness provide a sufficient reason for performing the act?" -- a reply which (if we take "reason" as having justificatory and motive force) Falk would say we have dismissed too precipitately. And it is true that Hospers taps into our intuitions here.

But it is also true that, as Falk admits, we do use the moral "ought" in an external sense, as signifying a demand made on us from without and without express regard for our motivational capacities, and it is, again as Falk points out, a "natural corollary" of the view employing this use that morality needs a sanction. Indeed the burden of much of Falk's paper, which he so ably discharges, is to show that in our moral thinking the "...external and internal uses of "ought" remain undifferentiated, and are imperceptibly

juxtaposed and confused" (p.509); that is to say, that our convictions are in tension. This being so a simple appeal to these will not help us. Rather what is required, since "we cannot obey two masters", is, as Falk says, a "...decision about what fact would most nearly correspond to our intentions in the use of moral language.."(p.510).

One of my concerns, in this review of Falk's response, will be to elicit what it is we must give up were the decision to favour (as Falk would have it) an internalist reading and to show that we cannot easily do this.

2. A tandem concern is this: the moralist cannot rest with the mere settlement of this intra-mural dispute. If, against the externalist view of the moral "ought", he champions the conviction that "...if anyone has a duty, that he has it is inseparable from the existence of some real check on his freedom to act otherwise.." (p.499), he must not only persuade us that his is the most revealing articulation of our moral intuitions but must also show (as Prichard omitted to do) that it has force for the sceptic who, presumably, does not as yet share these intuitions or, if he does, is worried about how well founded they are. Convictions are one thing, how to sustain them, yet again as Falk admits (p.499), another.

I shall attend to two crucial and related arguments, one of which can be regarded as preliminary, the other as final, whereby Falk endeavours to show that the moral "ought" on the (correct) internalist reading "...expresses a rationally necessary or objective determination of a person's will" (p.507), a conclusion which, were it to stand, should banish the sceptic's disbelief, or dispel his worries, while also discrediting the notion (discussed in the previous chapter) that when deciding between duty and self-interest we are faced with an existential or non-rational choice.

3. It is perhaps the case that Falk views desire in much the same way as Kant, seeing it as, like our natural appetites, something we are lumbered with and against which we must struggle on occasion if we are to exercise our autonomy. There is another reason however, closer to the current task, for his disavowal of Prichard's identification of (moral) motive and desire and that is that so identified, and because of the contingency of our desires, we can never be assured of our ability to do as duty bids us do.

But some caution is needed here. Though not incorrect, this way of introducing Falk's position may be misleading in that it seems to cast duty as a demand from without and so encourages the externalist reading he opposes. His position is perhaps best introduced by my remarking on the fact that from the start, and without any ado, he identifies motive and reason. Now this may seem to leave him open to the charge which Frankena brought against Falk's confreres -- that they failed to make the distinction between exciting and justifying reasons. But this is not the case here, however, as for Falk (and for Darwall after him) justification and motivation fuse at the final (ideal) stage of the deliberative process.

4. We get an inkling of this (impending) fusion on an examination of what I have called Falk's preliminary argument, an argument intended to move us away from the linking of moral motive with desire. Here he accuses Prichard of failing to note the distinction between having a motive occurrently and having it dispositionally (or "potentially and reflectively") and he goes on to claim that though there is a connection between having a motive occurrently and desiring (though none so close as Prichard

makes out), there is "none whatever" between desiring and having a motive dispositionally and it is the latter case which is most germane to the issue on hand; whence he moves us to accepting Kant's view of moral motive as independent of desire or inclination and as (possibly and in fact often) antagonistic to it, and to the view of moral action as self-induced behaviour.

The argument depends crucially on Falk's characterization of a reason or motive and it is here that I wish to take issue with him. I will suggest that his account, which is designed to evince Kant's view of moral action as willed or autonomous behaviour, may have the implication that precisely the opposite is the case. The characterization goes as follows:

In the sense relevant to this discussion, a reason or motive is a moving or impelling thought, the thought of that for the sake, or in view of which, some act is done; and I myself see no intelligible alternative to saying that it "moves" or "impels" in the sense that it functions as a cause of actions, in the conventional sense of cause as an antecedent implying a consequent by a rule of invariable connection. I should therefore, describe a motive as a causa rationis, a mental antecedent which, when attended to by a person, and in otherwise comparable conditions, will invariably be followed by an orientation of his organism towards the action thought of, in a way which, except for the intervention of distractions, counter-motives and physical impediments, will terminate in the action itself. Such a thought may be said to constitute simply a reason or motive, if, when attended to by itself alone, it thus causally implies action; and a sufficient reason or motive when, in addition, it persists in doing so even in view of opposing motives in the situation. Now, that someone has a motive may be used in either an occurrent or a

dispositional sense;.....The motive may be his either actually or potentially and reflectively so (p.495).

It would be idle then, on this conception of motive, to point out to someone in a state of practical doubt that "...he had a "sense of duty" in the occurrent sense", for in that sense "...for someone to have a motive strong enough for doing some act is for him to be subject to the operation of a law implying action; and a law cannot be said to depend for its operation on anyone's knowledge that it is operating" (p.496).

Falk continues:

In fact, it is in the dispositional sense only that it makes sense for people to claim that their ability to do some act depends both on their having sufficient cause for doing it, and their realizing that they have it and in what it consists. For a sufficient reason which exists dispositionally only will not actually function as a cause as long as the thought that constitutes it is not being dwelt upon; and if anyone wanted to make it function as a cause he would first have to realize that he had it and what it was. He would then, in the knowledge of the reason which dispositionally he has for doing the act, have a means for inducing himself to do it (pp. 496/7).

He concludes that

...Prichard should have acknowledged that if a man induced himself to act under the influence of such a motive he would not be acting simply from desire, or inclination, as the motive thought would here not function as a cause unless the agent had enabled it to do so (p.497).

5. A really clear reading of the previous two paragraphs is difficult to make and I shall neither attempt to describe the "correct" one nor to elicit the possible plausible alternatives. Rather I want to put what I see as a generic problem. My way of putting it may suggest one reading rather than some other but it can be adapted, I believe, to any of them.

Given the definition of motive as cause and its characterization in terms of the operation of a law, how does a man who has dispositionally a sufficient motive transform it into an occurrent one? It is Falk's claim itself -- that a dispositional motive will not function as a cause unless the agent enables it to do so -- which forces this question, for, on that claim, how can we understand the agent's inducing himself to do the act for which he has a sufficient reason dispositionally, other than as seeing him effecting this transformation? That business done, the now occurrent motive functions as sufficient cause and the agent performs the action.

But that business -- his effecting the transformation, his "inducing himself" -- is itself an action and itself requires a motive, a separate one, and whence this latter motive? Either he has this occurrently or he hasn't. If the former (and if strong enough) then he transforms the dispositional motive into an occurrent one, but invariably. If the latter then in order to acquire it as effective motive he must have it dispositionally and again there is required yet another (sufficient) motive. For him to act on the original dispositional motive there must be some related occurrent sufficient motive, but, again, this functions invariably as cause and (if he has it) issues in the agent's performing the act. It seems -- on this picture -- that the agent is never in a position to decide; agency, itself, seems to fall away.

A retort to all of this might be that what we have here is not a picture of agency dropping out but the very expression of rational agency, i.e., domination by thought or reason rather than desire or instinct. But still the feeling is that it is the thoughts themselves which generate action through (as against by) the agent; the picture is that of the "agent" as passive medium.

Or, it might be said: haven't you simply taken out agency by your treatment of such phrases as "inducing himself", "trying", and "dwelling upon"? But how else to fit the notions intended by these phrases into the Falkian schema?

6. There is another issue related to Falk's emphasis on dispositional motives, one which can be raised independently of the objection I have just made. This is the issue of what we might give up if we adopt the internalist reading of moral obligation; if, in place of, say, the view of duty as some kind of external demand we see it as some kind of internal demand, that in speaking of duty we are talking about "a certain kind of psychological fact".2 The issue is that of how well grounded is the fear of internalism's subjectivist implications.3

The problem is set by our experience that not only are the behaviours of different individuals in like circumstances quite varied, but that individuals vary greatly as to their moral sensibilities. If this variation is a reflection of (or identical to) a variation in motives (occurrent or dispositional), then on Falk's analysis, at face value anyway, what is a duty for one individual may not be for another in an otherwise like situation, and this conflicts with the commonsense view of a duty as, ceteris paribus, universally applicable. So in the very process of securing, in line with our beliefs, the necessary

connection between duty and motive (and so ability to perform), Falk may offend against another, and arguably a firmer, moral conviction.

It is true that, in dismissing Prichard's argument to a non-rational decision in those cases where duty opposes interest, Falk maintains that Prichard has been forced to that cc clusion by his failure to admit that "...our very thinking that we ought to do some act already entails that, by comparison, we have a stronger reason in the circumstances for doing it than any other" (p.500), and that this failure is a result of his not being prepared to hold that "... the sense of duty, even dispositionally, must always be everyone's strongest motive" (p.501). And we can take it that Falk affirms what Prichard failed to affirm. So -- it might be suggested -- if one can hold, on an internalist analysis, this position with respect to everyone's sense of duty, doesn't the analysis avoid the aforementioned offence?

But can one hold this position in the right way? One could, for example, point out that "our very thinking that we ought...etc.", if it is true, is true also of "oughts" other than that of duty. But let us leave that aside for the moment and concentrate on the second assertion. When one says "the sense of duty, even dispositionally, must always be everyone's strongest motive ", one is not making an empirical assertion. In the first place one would say "is" rather than "must be"; in the second place, where duty is regarded as distinct from, and assessed independently of, considerations of self-interest (and likely to oppose them), it is hard to see how anyone could be in a position to make such a claim. Rather what one seems to be doing is making a statement of necessity. The question then is how are we to construe "duty" so that the statement is true. One way -- and the only way I can see -- is to regard someone as having a duty only when, having stacked up the various reasons (motives) he has, the sum of moral reasons

outweighs for him that of the competition. That is to say, though one may have moral reasons for performing an action, one has a duty to perform it only when these reasons are weightier. On this construal of duty the statement is true because tautologous, but on this construal too there seem to be strong subjectivist implications, implications that are very much at odds with our moral convictions.

7. The same disturbing feature can be elicited from Falk's own comparison of his and Prichard's positions. Remember that on Prichard's analysis, though one's having a duty is not a function of one's psychological condition, one's ability to do one's duty is; that there is a "subjectivism" with respect to the ability to act morally. Falk's way of putting that position is this: that what moves us to action is not "...a motive implicit in the very fact of duty itself; but a motive constituted by the thought that we had the duty, and existing apart from and additionally to the fact of duty" (p.497). His own position, in contrast, is that an agent "...in the very thought that he ought to do the act, [has] dispositionally a motive for doing it .." (p.497).

Though Falk is not altogether clear here, there seems to be, between the two positions, a reversal of the order of dependence. On Prichard's view the true thought of duty depends on there being the fact of duty, and the existence of a motive depends on a desire to act on the thought. On Falk's view it seems that the fact of duty is constituted by the motive; that is, if no motive (impelling thought), then no duty. Now this may seem an overly strong interpretation and it may be -- though I am not at all sure -- that it is one which Falk would reject. But what he must reject, or, rather, what he must refute, is the charge that implicit in his fusion of the "how" and the "why" of duty (as against their separateness in Prichard) is a subjectivism with respect to the fact of duty itself. For, for

all that has been shown thus far, there is this implication and we have a view of duty which does violence to our ordinary moral convictions, indeed moreso, I feel, than does the view of Prichard which it seeks to replace.

8. In urging a use of "ought" or "duty" in a "purely formal motivation sense", i.e., that sense where "...when a person asks "need I really, or have I, if only dispositionally, a sufficiently strong reason for doing this act?", he might as well have said "should I or ought I really to do it?"..." (p.503), a use which ensures the "purist" view of a necessary connection of "ought" with motivation, Falk appeals to Kant's position as follows:

It is worth remembering that Kant who insisted most on the purist view of the connection insisted also on a use of "ought" in a purely formal motivation sense. Modern deontologists overlook that Kant would have rejected offhand their view of duty as an objective requirement of the situation, as he did reject the view of duty as consisting of the demands of a deity, or of some inward compulsion of a special quality. For he objected to anything being called a moral imperative with regard to which a person could still ask whether the act required of him was one which, in a motivation sense, he really ought to do, i.e. in Kant's language, whether the act was one which rationally a person would be necessarily and unconditionally determined to will; and he insisted on "ought" or "duty" being used for nothing but the very fact itself that rationally a person would thus be made to will an act. Kant's very definition of "ought" therefore makes it a tautology that anyone who has a duty has, on this account alone, a reason, though not necessarily an impulse or desire, sufficient for doing the act; and Kant thus dispenses with a "sense of duty" in the shape of any singular and contingent psychological disposition (p.504).

Now the last sentence alone appears to deny the subjectivism of which I spoke, but really, by itself, it represents no advance in Falk's position. A "sense of duty" (i.e., Prichard's "desire to do what is right") is indeed dispensed with but only (as Falk allows) by Kant's very definition of "ought", a move which, I have argued, may merely replace one subjectivism with another (less palatable) one. The burden is squarely on Falk to show us that his view does not have this consequence, and an introduction of the term "will" (as noun or verb) will not itself effect this task. For then Falk must show that, for example, Frankena is wrong when he states that what one is "...impelled to do even after reason has done its best is still dependent on the vagaries of one's particular conative disposition..".4

And indeed that he is prepared to attempt this is signalled by his invoking the notion of rational necessity, a notion which is central to what I have called his final argument. Perhaps this notion can provide that further stricture required to ensure the universality of the moral "ought" (in a specific kind of situation).

9. Falk explicates the notion thus: "...any "ought"...expresses a rationally necessary or objective determination of a person's will.. [where what is necessary is].. what no trying will alter, or what is invariable, and "rationally necessary" or "objective" what would not be altered or varied by reason, or by mental operations". Then with respect to Kant's distinction between a conditional and an unconditional or absolute "ought" he says:

The crucial point is simply that not every motivation "ought" does as yet express a completely conclusive reason for doing an act. If an act must dispositionally be willed only on account of some other end which is desired, whether it also must or can be so willed in every respect would still be an open question; for it may still apply that this end

itself need not or could not, on consideration, be willed, either on its own account, or on account of the implications of pursuing it in a given situation. A conclusive reason would be one with regard to which no further question could be asked: which was thought rationally unavoidable on account of the act itself (or on account of some other end thought rationally unavoidable in itself), and rationally, in the given circumstances, unavoidably stronger than all opposing motives. An "ought" embodying such a reason would be "absolute" in the sense that, in relation to given circumstances, it was formally complete; and it is when such an "ought" is identified with a moral "ought" or duty that the connection of duty with sufficient motivation becomes logically necessary (pp. 507/8).

Now Falk is, up to a point, correct here. There is certainly a sense in which something is rationally necessary; for example, given that one wills A and that B is (and is known to be) the only means to A, then one not merely would will B but must will it. And if there were a "conclusive reason", or a "formally complete "ought"", it would be as Falk describes it. And if such an "ought" were identified with a moral "ought", then that one morally ought would entail not only that one can but that one rationally must.

But is there such a thing as a "conclusive reason" in Falk's sense? How does he (or how might he) persuade, for example, the sceptic, that he should side with Kant here and against Hume; how might he show that there is such a thing as a rationally unavoidable end?

Again, supposing there to be such an end, such a "conclusive reason", and given that we are using "ought" in the purely formal motivation sense, that is, as content-less, how is it shown (or how can it be shown), in cases where duty and self-interest conflict, that what one ought rationally to do (what one is rationally compelled to) is one's duty?

That is to say, how might one argue to the identification of the formally complete "ought" with the moral one?

And finally -- and this last worry may or may not be addressed if answers to the previous two are forthcoming -- how to ensure that what is a formally complete moral "ought" for one is so for another in comparable circumstances? Whereas it is indubitable that there is but one rationally necessary willing available to all agents in the kind of means-end reasoning touched on earlier, how can it be shown that there is a like universality with respect to conclusive reasons; that there is something to the rational, or to being a rational agent, such that when (or if) the Kantian equation of the rational with the moral holds for one it holds for all?

Let me dub these three worries the "three indeterminacies" with respect to Falk's appeal to rational necessity. The label is justified in that, though Falk has given us the conditions on a rational vindication of the moral "ought", he has not at all shown that these conditions can be satisfied.

10. I hope to have shown that Falk's account of a reason as a "causa rationis" has untoward implications with respect to agency; that his internalist analysis of the moral "ought" threatens a subjectivist account of duty which offends our ordinary moral convictions while failing to sway the moral sceptic; and that his appeal to the rational does not go nearly far enough either to dissolve that subjectivism or to show indeed that, where duty and self-interest collide, we are not faced with a non-rational or existential choice.

Footnotes to Chapter 5.

1. W.D. Falk, " "Ought" and Motivation", 1947-8. In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., first edition, 1952, p.499.
Unadorned references in the text and in the footnotes in this chapter are to this work.

2. p.505. This fact is later described as ".. an ideally inescapable inner entanglement, a dictate of conscience" (p.509).

3. "subjectivist implications" is Falk's own phrase. See p.509. I shall try to justify Frankena's charge that internalism runs the risk "...of having to trim obligation to the size of individual motives" (Frankena, op.cit., p.80).

4. Frankena, op.cit., p.77. My italics.

CHAPTER 6

# AN APPEAL TO FORMAL FEATURES OF REASON AND RATIONAL AGENCY

1. Gewirth1 cites this contention of A.J. Ayer's

> Given that a man has certain moral principles, we argue that he must, in order to be consistent, react morally to certain things in a certain way. What we do not and cannot argue about is the validity of these moral principles.2

and claims that proponents of "the moral point of view", though they hold a contrary position with respect to the validity of moral principles, have thus far, and for a variety of reasons, failed to vindicate their convictions here.

It is his intent to repair that deficiency; to provide a rational justification of a moral principle -- specifically of the egalitarian-universalist principle embodied in "the moral point of view" as this is expressed by such as Baier. This principle will provide "...a body of rules and criteria for individual action and social policy which accord to all men certain basic equal rights...[rights which] consist primarily in freedom from coercion by other persons and in certain essential aspects of well-being" (p.116).

Success in this will not just supply a vindication of this particular principle but, being a rational vindication, will allow us to construe "...the difference between what is morally right and what is morally wrong as objective and universal and hence as knowable by moral judgements on which all persons who use rational methods must agree" (p.113). So it will allow us to reject Ayer's contention and, incidentally, will allow

us to discount any related view, e.g., that of Mackie, who claims that we can, logically, simply opt out of moral discourse.3

2. Gewirth's thesis is that the egalitarian-universalist moral principle can be derived from an appeal to purely formal features of agency and rationality; that is, that the principle can be logically deduced from analysis of the concept of reason itself and the concept of rational action/rational agency.

Reason (or rationality), for Gewirth, is characterized by its "...avoidance of, and opposition to, arbitrariness" (p.120); it (specifically as embodied in deductive and inductive logic) opposes the arbitrary "...insofar as it takes account of what is cognitively necessary or coercive either logically or empirically" (p.120).

One example of logical necessity is (what he calls) the principle of universalizability. This states that "...if a predicate belongs to one subject for a certain sufficient reason, then it logically must belong to every other subject which fulfils that sufficient reason" (p.127), and so it generates a "requirement of consistency" (p.137).

Though this principle plays a central part in the derivation of egalitarian-universalism, of itself it does not disjustify particularist moralities of the sort championed by such as Aristotle, Nietzsche, Hitler (p.116). This is so because it allows "...two important kinds of variability with respect to content" (p.128), first as to the actions it allows and second as to the criteria of relevant similarity. That is to say, an agent can inject whatever content suits him -- can fill out "predicate" and "certain sufficient reason" however he wants to -- and not offend against the principle.4

However, though ineffectual of itself, when coupled with necessary features of agency there emerges an argument to an egalitarian-universalism, this view (or principle) being, in fact, implicit in that combination.

3. The argument, put in point form is, briefly, this:5

1. An action (in the relevant sense) is that which is voluntary and purposive -- the agent knows what he is doing, chooses and controls his behaviour and acts for a purpose, for something he sees as good. That is to say, voluntariness and purposiveness are the "categorial features" of action or agency.

2. It is necessarily true of every agent that qua agent he participates voluntarily and purposively in transactions in which he is involved. (A transaction is an action in which an agent acts on at least one other person, the recipient.)

3. The agent, in acting for a purpose, regards his action as justified by that purpose.

4. So, (from 3.), he regards himself as having a right to perform the action and he makes a corresponding right-claim. (Gewirth importantly qualifies "justification" in 3. and "right" in 4. I will discuss the qualification later).

5. (From 2. and 4.) Every agent necessarily claims, at least implicitly, that he has a right to participate voluntarily and purposively (which are equivalent, respectively, to the rights of freedom and of well-being) in transactions in which he is involved.

6. The only non-arbitrary (rationally justified) description under which one claims the right(s) in 5. is that of oneself qua prospective agent who wants to realize some purpose of his. That is to say it is only this -- the necessity of his possessing the

categorial features of agency, what he must have to be an agent -- which can non-arbitrarily ground his right-claim.

7. These features are held by all other (prospective) agents.

8. (From 5. 6. 7. and the principle of universalizability) He must hold that all other agents have the same right (to freedom and well-being).

9. (from 8. and the logic of rights-talk) He must hold that he (or any other agent) has a corresponding duty; that is, in that agents' rights entail a correlative obligation, he must hold that he (and others) must allow them basic freedom and ought not to harm them.

What falls out then is a precept which can be addressed to every agent: "Apply to your recipient the same categorial features of action that you apply to yourself"; a precept which Gewirth calls the Principle of Categorial Consistency (p.135). And this is an egalitarian-universalist moral principle "...because it requires of every agent that he be impartial as between himself and his recipients when the latter's freedom and welfare are at stake" (p.136).

In short, in that an agent rationally must allow that others have as much right to the goods of freedom and well-being as he does, and given that he "necessarily claims"[6] his right to these goods, he has, and must admit that he has, a moral obligation with respect to others.

4. Gewirth sees this argument as constituting a derivation of a moral principle from "...rational considerations about the features which necessarily pertain to all action" (p.137).[7] By substituting "rational necessities" for the "contingent contents" allowed by the principle of universalizability, he sees it too as providing a refutation of all

particularist (inegalitarian) moralities (p.138). And, finally, he sees it as demonstrating, contra Hume, that rationality (reason) is not morally neutral (p.139).

5. Should it go through, the argument seems, at a stroke, to resolve the three indeterminacies infecting Falk's position. For what is argued to is a "conclusive reason" similar in kind to that touted by Falk in that it is rationally unavoidable; that is, it is conclusive in that, falling (logically) out of claims that an agent necessarily makes, it expresses a rational necessity. It is also argued that this conclusive reason is to be identified with a moral "ought"; i.e., that it is a moral obligation which is rationally required. And, finally, it is argued that not only does it hold true of one (or another) agent that he has this moral obligation but, from the fact that it is qua rational agent that he incurs it, that every agent is likewise morally obligated; that is, it is argued that the Kantian equation of the rational with the moral, in holding for one, holds for all.

6. The argument does not, however, go through. In explicating 3. and 4. Gewirth says:

> It must be emphasized, however, that the justification and rights here in question...need not be moral ones; they vary according to whatever criteria the agent invokes explicitly or implicitly in the purposes for which he acts (p.131).

Now what can he mean by a non-moral "right"? It seems that -- like a non-moral justification -- it must be simply a reason. On this construal 4. reads: "he regards himself as having a reason to perform the action"; and if he is to claim something it would be that he has such a reason. Now if this filters down to 8. -- the appropriate intermediate

changes being made -- what we have there is "All other agents have the same reason..etc ". But, whereas one's having a right may be inextricably bound to others having a duty, there is no such obvious relation between persons' reasons.8 Nothing like 9. will fall out. That 9. falls out for Gewirth is a consequence of his substituting "right" for "reason", qualifying this (odd) use of "right" and thereafter omitting or discarding the qualification.9

7. We can, perhaps, get a notion of how things have gone wrong by looking at Gewirth's preliminary sketch of the argument (on p.130). Here he claims that every agent, on the ground that he is a prospective agent, "necessarily claims as rights for himself" the categorial features of agency, those features that he must have to be an agent; that agents "logically must" make this claim. But one may claim those features as a necessary (logical) condition of one's action and yet not claim, in the sense needed for the transition from 8. to 9., a right to these features, a right to acting. There is not the entailment here that Gewirth thinks there is.

There are a few other things that we can say about agents/agency that might be thought (mistakenly) to yield the necessity in question. For example, it is true of someone that on every occasion of his acting he has in fact the freedom etc. But his mere having it doesn't (not directly anyway) have a bearing on whether he ought to have it in the sense that he has a right to it which must be recognized (even by him). And if he wants to (verbally) assert his status as agent he logically must claim possession of (as against a right to) the categorial features. What he may have -- and what prima facie every person does have on the ground that he is a prospective agent -- is a reason to ensure his continued possession of these features.

Now, if an agent claims a right on the ground of his being a prospective agent then, in order to be consistent, he must, as Gewirth contends, accord the same right to relevantly similar others. And an agent may easily slip into claiming a right where he has (or thinks that he has) a reason, in that by claiming a right he secures (or tries to secure) his ability to act on his reason. But need this claim be granted? A person may well grant that another has a reason but nonetheless may think himself under no obligation to support him nor indeed to refrain from frustrating his end. Indeed he may have reason to do just that -- frustrate that end -- even where the end is to secure the conditions of (continued) agency.

To say this is not simply to deny what Gewirth affirms. Rather it is to say again that, unlike the way someone's rights have a bearing on another's duties, one person's reasons may have no bearing upon another's, and to say that Gewirth has not shown that they have.

Footnotes to Chapter 6.

1. Alan Gewirth, "Moral Rationality", 1972. In Freedom and Morality, John Bricke, ed., 1976. Unadorned page references in the text and footnotes in this chapter are to this work.

2. p.114. From A.J. Ayer's Language, Truth and Logic, 1948, pp.111/112.

3. Mackie, op.cit., p.100.

4. See Mackie's like analysis of the "first stage of universalization" in Mackie, op.cit., chapter 4.

5. The argument runs from p.130 to p.136. I shall not discriminate exact quotation from paraphrase in this summary of it.

6. Indeed the PCC should be written "Apply to your recipient the same categorial features of action that you do and must apply to yourself".

7. The argument would show, then, that for any agent it is rational for him that he be guided by the PCC. So it would have force for our moral sceptic, who, we have allowed, is open to being convinced; and it would obviate the existential choice, at least in those basic cases covered by the principle.

8. That there is such a relation (though not an obvious one) Nagel tried to argue in The Possibility of Altruism but he has since pulled back from the view expressed there.

9. See e.g., pp.137/139/142.

# PART II
# THE SIXTH ATTEMPT

# REASON AS IMPARTIAL

CHAPTER 7

INTRODUCTION

1. If I have been right about the responses we have considered thus far, then we have reason to believe that our moral sceptic remains unswayed, and we can see Darwall's thesis in Impartial Reason1 as yet another -- and needed -- attempt to convince him. This thesis is, I think, appropriately the last such attempt to be dealt with here in that it has important relations to all the positions hitherto expressed.

So, although Darwall is concerned to refute (against e.g. Mackie) the claim for the force of present desires as reasons, and is concerned too to deny (against e.g. Prichard) that the ability to act morally is contingent on the agent's desires, he sees this alleged connection between desires and reasons as motivated by what he calls "the good sense of Humean internalism" (p.80).

Like Falk, then, he embraces an internalist view of reasons and he is kin to Falk too in the emphatically Kantian strain of his thought. Indeed we can see him as attempting to provide a resolution of the indeterminacies in Falk's position, though a resolution which, as we shall see, Falk might not endorse.2

Again, though he rejects the positions of Baier and Gauthier, he applauds (and incorporates) their "Hobbesian reflections" and their respective appeals to a background theory of reasons and of rational agency.

Putting it roughly, he regards Baier's background theory as "questionable" (p.19) because it either starts from a bias towards the moral (p.186) or it suffers Kant's

objection to any "materialist" theory of the nature of morality in that the rationality of action on moral reasons is made conditional on a desire -- the desire that oneself flourish or the desire that people generally flourish -- that rational agents may lack.3

As to Gauthier's contribution, the self-critical or self-reflective aspect of ideal rational agency is endorsed but the requirement of self-support derived from this is too weak, Darwall claims, to bear the weight put on it; that is, it does not generate the (moral) principle of constrained maximization.4 Nonetheless, self-criticalness is an important component in Darwall's derivation of a stronger requirement, that of universal impartial self-support.

For Darwall then, as for Baier and Gauthier, Hobbesian reflections need to be supplemented by an appropriate background theory of practical reasons -- by "...additional, more formal aspects of practical rationality" (p.19) -- and in this addition his argument resembles that of Gewirth's in that (in Darwall's words) "... [it proceeds] 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" (p.202/203).5

2. Not surprisingly the outcome of this eclecticism is a complex argument. We can get a rough or introductory idea of the strategy underlying the argument by summing up what I have just said; by seeing Darwall's thesis -- that some basic moral considerations are overriding reasons..etc. -- as proceeding from an amalgam of several theoretical approaches:

1. An internalism with respect to reasons and rational principles. This involves the explication (and subsequent use) of the motivational aspect of reasons and of

practical reasoning. We can expect such a theory to have something to say on how we can act on or for reasons.

2. Formal considerations of rationality or the consideration of formal aspects of rationality; what Darwall calls "the normativity of reasons to act" (p.19). This involves the generation and explication of a background theory of reasons. We should expect such a theory to have something satisfying to say on justification and the related notion of the authority of reasons and of rational principles or norms; to have something to say on why we should act one way and not another.

3. "Hobbesian reflections": consideration of the results for oneself of one's acting in a certain way where what one does affects others and vice versa, and consideration of the results of all agents acting in one way as against another.

We have already touched on each of these approaches but (except when we dealt with Falk) in relative isolation, one from the other. In Darwall a combination of the internalist thesis with the normativity thesis issues in a particular expression of the ideal of rational agency (of the ideal rational agent), one which, in turn, leads to a condition on any principle's being a rational principle -- that it have universal impartial self-support. Then, on the supposition that many different principles can satisfy this condition, Hobbesian reflections, it is argued, -- via the employment of means-end or relative rationality -- narrow the range to moral principles of some kind.

This sketch is very general and incomplete. Internalism enters in at different points, the normativity thesis -- accentuated in the later critical stages of the argument -- is pervasive, and Hobbesian reflections too play a part additional to the one indicated. Moreover there is the further complication that, though it figures large in the text of

Impartial Reason, the criterion of universal impartial self-support is (by his own admission) not essential to Darwall's argument.6

Perhaps a better sense of that argument can be obtained by considering some of the ground-clearing which must be done in order that Darwall's aim -- to show the rationality of moral conduct -- be capable of being realized. This consideration will serve also to give some indication of my focus.

3. We can start by asking what it would mean -- what the import would be -- were it the case that Gewirth's argument goes through, were it not objectionable in the way I've described. Were that the case it would follow that for any agent, were he to contravene the Principle of Categorial Consistency, he would be open to the charge of being inconsistent and, thereby, irrational. What is the force of this charge; e.g., must it have practical effect in that if the agent admits its correctness he will, or will be inclined to, mend his ways?

Gewirth, in reply to the objection that the consideration of self-contradiction is not usually practically efficacious in this way, draws the distinction between motivating and justifying reasons and claims that the "morally practical question" in the relations between rational agents is not who can influence the other but whose position is rationally justified and that "...the most basic way of proving that a position is not rationally justified is to show that it involves a contradiction".7

Now Gewirth is surely right here but there is still the worry that, while claiming his right to freedom and basic welfare, were an agent to get away with avoiding his corresponding obligations to others then he could simply ignore the "morally practical question". We can say to him that he is being irrational; that his position is rationally

unjustifiable. But now, again, why should he be bothered by that? We are not asking whether or not he will be bothered given the kind of person he might be. The question is, rather, should he or ought he be bothered? Now this may be thought to be a senseless question.8 After all the "ought" is the rational "ought" and the question then appears to reduce to "ought he to do what he ought to do?", the "ought" having the same sense in both instances. But the question doesn't so reduce. Rather it asks if the rational "ought" is operative here. Does it have an application; is it apposite? That is to say the question challenges the scope -- and hence the authority -- of practical reason.9 One has the unsettling suspicion that rather more to the point is the unlikelihood of the agent's getting away with it; that more germane is the fact that others, on equal footing with the agent, would simply not allow this asymmetry. And with this comes the corresponding suspicion that for such an agent, who has as well a considerable power advantage, the charge of irrationality loses its sting roughly in proportion to the degree of that advantage.

Gewirth gets close to addressing this worry when he considers the objection that, for example, it may be rational for a Machiavellian ruler to make inconsistent statements in order to advance the end of maximizing his power; the objection in effect being that the charge of irrationality (`a la Gewirth) is empty here, it being the case, ex hypothesi, that the tactic is (instrumentally) justified. Gewirth's reply is that the end of self-aggrandizement has not itself been subjected to scrutiny "...by criteria which are independent of his own [the ruler's] predilections".10

But our worry is precisely this -- are there such criteria? Is there more to rationality than means-end or relative rationality? Gewirth merely assumes that there is.

What we need, then, is to have it shown that there is such a thing as substantive rationality -- that desires, intentions, purposes, preferences, ends etc., are, contra Hume, criticizable; that reasons can be given for or against them. Then, even though someone may be unmoved by the charge that he (or his predilection) is irrational, the charge would yet have force in that it correctly points up an important deficiency.

4. Now one of Darwall's arguments to the existence of a substantive rationality is that formal theories of decision, in their uncontroversial assumption that sets of intransitive preferences are irrational, require in fact that preferences be criticizable by reasons. I shall not deal with the details of this argument11 but I shall query the scope of the conclusion and suggest that a more limited scope than that intended by Darwall is both consistent with decision theory and sits better with Darwall's internalism. Darwall is of course not content with this conclusion alone, with showing merely that some ends may stand and others fall on rational scrutiny. But not only does he want to exhibit the eminence of moral reasons, he wants to show too that reasons for action do not depend on, and need not refer to, present desires (or preferences) at all.12 He has, then, two motives.

Given that there is a substantive rationality, one could champion any of a number of positions. For example, one could hold that only self-interest figures and so only self-regarding desires are rational. Or that only one's desires figure, where one might have, for example, desires with respect to the well-being of others, or for states of affairs, as well as desires regarding one's own well-being.13 Or one could hold that moral reasons figure and that in a Hobbes situation they dominate.14

Now one motive is to show that the third position is the correct one. Although Darwall allows that self-interest generates reasons he wants to show that (in the world we have) when these reasons conflict with moral reasons they are defeated.

The other motive is to show that moral reasons figure independently of desires. That is, he wants to show, against e.g. Prichard, that there exist, independently of desires, reasons which are not only justifying but are enabling also; that an agent's ability to act on a moral reason is not contingent on his having the appropriate desire.15 And, as with Falk, the impetus here is Kantian and it too is two-fold.

First there is the concern that any materialist theory of reasons -- where "materialist", for Kant, meant any "practical principles which presuppose an object (material) of the faculty of desire as the determining ground of the will" -- fails to guarantee the rational necessity of moral action because, again in Kant's words, "we cannot know, a priori, of the idea of any object, whatever the nature of this idea, whether it will be associated with pleasure or displeasure or will be merely indifferent" (p.174). That is to say, there is no (putative) object of desire such that it (rationally) must move one (everyone). And if we fail to demonstrate rational necessity we also fail to show moral principles as being valid for all agents.

The second (and related) concern is to exhibit moral agency as an expression of, or a realization of, self-governance. Moral agency is then incompatible with a view of reasons as being based on desires, in that, in being guided by the desires he just happens to have, an individual is guided heteronomously.

These two concerns can be put as one: the concern to show that it is possible for any agent to act morally regardless of his current desires; that he can (against his

desires, if need be) enable himself to act -- as he sees that he ought, as he rationally must -- by an exercise of his free will.

5. This explication of Darwall's thesis, perhaps not surprisingly in that they are both pushing the Kantian line, seems to present it as an iteration of Falk's position. There is the same Kantian equation of the moral with the rational with the autonomic. They are, however, not identical positions.16 They diverge by virtue of Darwall's additional Hobbesian reflections, the addition forcing a different interpretation of the equation. This may be the cost -- if cost there be -- of arguing to that equation other than by way of what Darwall calls "... [Kant's] difficult metaphysical doctrine of noumenal freedom" (p.200).17 But we can perhaps best capture the tenor of Darwall's argument -- and perhaps sense the possibility of tension therein -- from two of his remarks on it, the first prefatory, the second consisting in his final words on the subject.18

Having acknowledged Rawls' influence on the work he says of Impartial Reason that it is "...an attempt to provide Rawls's theory of justice with a theory of practical rationality to anchor its Kantian interpretation" (p.9). Now Rawls, though he admits his theory is not precisely Kant's view,19 has had occasion to defend it against charges of heteronomy levelled (fairly or unfairly) at it.20 And these charges stem from an uneasiness with respect to (and are levelled at) the apparently Hobbesian reflections of the parties in the original position. Rawls replies by emphasizing that the parties, though "mutually disinterested", are not thereby egoistic as "...they are moved in the first instance by their highest-order interests in developing and exercising their moral powers.." and these aims are "...entirely fitting and proper".21 Nor, he argues, does his theory allow in heteronomy in another way, i.e., via the constraints on the parties behind

the veil, because these constraints are expressive of the moral powers of the persons whom the parties represent.22

Now, prima facie, a formally similar line of defence seems open to Darwall as well but, unlike Rawls, he cannot appeal to our deep-seated moral convictions23 as it is (some of) these very convictions he wants to vindicate. That is, in that Darwall is not preaching to the converted, to those whose disputation has to do with which of the moral philosophies (of those available to us) is best, his approach is necessarily "hard-headed" and this hard-headedness involves -- as we shall see -- Hobbesian reflections additional to those employed by Rawls.

Much later, having urged the view that Rawls' theory of justice has a Kantian interpretation by urging that principles chosen from the original position and with a view to self-governance "...express the Kantian ideal of autonomy" (p.248), Darwall identifies his view with the view of Kant in the following manner:

> Kant's idea that moral principles would be the object of will for all in a realm of ends ties them to the more abstract perspective provided by our thick veil. In discussing his notion of such a realm he wrote: "By "realm" I understand the systematic union of different rational beings through common laws. Because laws determine ends with regard to their universal validity, if we abstract from the personal difference of rational beings and thus from all content of their private ends, we can think of...a whole of rational beings as ends in themselves" ...This abstract perspective is the intersubjective standpoint of a rational agent considered as such.
>
> I have argued that the fullest expression of practical reason is to act on principles we could choose for all from this point of view (p.249).

6. I have, I hope by now, given a good idea of the purpose in Darwall's ambitious project and some idea of the strategy he adopts. The question is whether the strategy secures the purpose. I think it does not. I shall confine myself to the attempt to show four things which I give now in what I see as an increasing order of importance and in an order which reflects too the increasing degree of confidence with which they are put forward.

1. that the thesis is not one of which Kant would have approved; or -- more cautiously -- that it diverges from Kantian orthodoxy. In that my knowledge of Kant (and of Kantian orthodoxy) is quite severely limited and almost entirely derivative this position is tentative. Still, I don't wish to undercut it altogether and I shall suggest that it is in virtue of an additional Hobbesian reflection that the thesis would be rejected by Falk, for example.

2. that the thesis is unstable. Even if we grant his argument to Rawls' original position, the ends of the parties in that position, in view of which they choose the principles by which they will guide their conduct when the veil is lifted, are determined by Darwall in a way which leans too heavily on Rawls' conception of "a rational conception of a person's good" (p.105), his defence of that conception being inadequate. That is to say, the force or authority of the rational "ought", which is deemed to derive from an ideal of rational reflection, is vitiated by its being derived, ultimately, from a view of a "rational conception of a person's good" with which one can disagree and over which there is, in fact, much disagreement.

3. that there is a considerable tension in Darwall's argumentation. On the one hand he embraces an internalism with respect to reasons in the manner of Falk, and

emphasizes (what was implicit in Falk) the necessary connection between motivation and all reasons, not just moral ones. This view, normative internalism, he argues, gives an altogether more satisfactory account of reasons than does its alternative, normative externalism, in that it is not subject to the objection of ontological queerness (nonnatural properties) nor the consequent epistemological and motivational problems. Strategically he needs the view to move to the intersubjective standpoint.

On the other hand he holds that the motivation involved need not depend on any desire (or preference) the agent may happen to have; that is, he embraces motivational externalism. And he needs to do this in order to preserve the features of autonomy and rational necessity. But not only is motivational externalism a difficult doctrine to sustain, Darwall's characterization of normative internalism strongly suggests the existence of what I call a base preference set, that is, a set of one or more uncriticizable preferences -- the engine, as it were, of all motivation. That is to say, Darwall's formulation of normative internalism sits altogether better with a motivational internalism than it does with the motivational externalism he holds to.

This criticism is related to the criticism in 2. and both it and 2. generate a suspicion that (contra Darwall) there is not, after all, a substantive rationality in the strong sense of its being the case that all preferences are "...criticizable in terms of reasons" (p.77).

4. that the argument to the intersubjective standpoint (the impartial standpoint of the rational agent as such) does not work; that it is either fallacious or it rests on a dubious version of judgement internalism.

Footnotes to Chapter 7.

1. Stephen L. Darwall, Impartial Reason, Cornell University Press, 1983. Hereinafter, unadorned page references within and without parentheses, in text and footnotes, are to this work.

2. Though inspired by Falk, Darwall feels "..[he] would almost certainly reject the theory I offer here" (p.9). It is, I think, arguable that Kant would reject it too and for the same reason. I will take this up later.

3. See p.174 and p.191.

4. as I've tried to show in Chapter 3.

5. As we shall see, Darwall's "formal conception" is richer than that of Gewirth's. Though it owes much to Nagel, it differs from the latter's conception too. See his chapter on "Nagelian Objectivity". See also his commentary on Nagel's later work, The View from Nowhere, 1986, in "How Nowhere Can You Get (and Do Ethics)?", in Ethics, Volume 98, 1987. I do not have the space to treat of that difference.

6. B.C. Postow in his "Darwall and the Impartial Standpoint", in Philosophical Studies, Volume 94, 1986, rejects premises leading to that criterion. Incidentally, Postow's attempt to hone Impartial Reason to its essentials illustrates well the complexity of the argument there.

In "Darwall's Kantian Argument", Canadian Journal of Philosophy, Volume 18, Number 1, March 1988, George Terzis directly attacks the self-support condition. Darwall's claim that his theory need not appeal to self-support is in his "Reply to Terzis" (same journal, same issue), 1988.

My criticism of the argument depends neither on the inclusion nor the exclusion of the self-support condition.

7. Gewirth, op.cit., p.145.

8. See Darwall pp.215/216.

9. Obviously in the manner of Hume. There are echoes of Nielsen here too. See his remarks regarding the subjectively evaluative tone in our use of the rational "ought" in Nielsen, op.cit.

10. Gewirth, op.cit., p.148. Note the similarity to Darwall's view: "The rational person is not constituted by whatever ends or preferences he happens to have at any given moment. Rationality consists, at least partly, in our capacity to make our ends and

preferences the object of our rational consideration and to revise them in accordance with reasons we find compelling" (pp.101/102). I shall have more to say on this.

11. See his Chapter 6.

Also see Thomas E. Hill, Jr, "Darwall on Practical Reason" Ethics, Volume 96, 1986, pp.607-609, for a possible coherentist (or relativist) response.

12. nor indeed on or to any desires the agent may ever in fact have, were he to fail to realize his rational potential.

13. This we will call -- following Darwall -- a self-centred theory of reasons, as against the earlier "self-regarding" one.

What I have given is, of course, a loose formulation. On one construal it allows a Humean whim -- to scratch or not to scratch one's finger -- as providing a reason. On another it may allow only some desires as providing reasons e.g., those that survive a full Brandtian reflection.

It might be claimed that on the former construal the view reduces to the thesis that there are no nonnrelative reasons; that rationality has to do only with the choice of means and takes the ends (what one desires) as simply given. This claim is -- prima facie at any rate -- incorrect in that the view has something to say about preferences e.g., that it is rational to act on the preferences one has and presumably -- though there might be some difficulty here -- irrational to act (let us say) otherwise. The other thesis, by contrast, has nothing at all to say about preferences.

14. A Hobbes situation will be described later. Our normal circumstances is such a situation.

Darwall has an important proviso on the overridingness of moral reasons; or, put another way, a proviso on when something is a moral reason.

15. though it is, he allows, dependent on his capacity to be moved. See pp.41/42.

16. When Darwall suggests (p.9) that Falk would not accept his view he does not elaborate either then or later.

17. Later in this essay I propose that there are two arguments running in Impartial Reason, one explicit, the other not. When the former, which is designed to circumvent Kant's difficult doctrine, runs into difficulty the latter, which (I suggest) leans somewhat on that doctrine, makes an appearance. That there are two arguments is perhaps more evident when one compares the text of Impartial Reason with Darwall's elaboration of his theory in the later essays.

I bring this up now to forestall the objection that, since the additional Hobbesian reflections occur after both arguments (explicit and implicit) have run their course, it cannot be that Darwall's interpretation differs from that of Falk because it seeks to circumvent Kant's doctrine. In reply I would say that Darwall could assume some of the tenets of that doctrine e.g., that to be rational is to be autonomous, while yet rejecting the purist conception of Kant's position, based on that doctrine; namely, that his position is starkly inconsequentialist.

This last, I admit, is speculative, relying on nods and winks rather than Kant scholarship. Still -- but only insofar as I remark on the difference between Darwall and Falk -- I am gambling on it.

18. in Impartial Reason. The later essays, to which I referred in the previous footnote, are "Rational Agent, Rational Act", Philosophical Topics, Volume X1V, No. 2, Fall 1986, pp.33-57; "How Nowhere Can You Get (and Do Ethics)?", Ethics, Volume 98, 1987, pp.137-157; and "Reply to Terzis", Canadian Journal of Philosophy, Volume 18, Number 1, March 1988, pp.115-124.

19. John Rawls, "Kantian Constructivism in Moral Theory", The Journal of Philosophy, Volume LXXV11, No. 9, September 1980, p.517.

20. See p.527 and pp.531/532 ibid. See also Darwall, 1983, pp.247/248.

21. Rawls, ibid., p.531.

22. Ibid., p.532.

23. See Rawls, pp.518/519, ibid. I wish to be not simplistic on a complex position but I take it that the appeal is there though not so stark or unsupported as my words might suggest.

CHAPTER 8

# THE MOTIVATIONAL ASPECT OF REASONS

1. The motivational aspect of reasons is evident in Darwall's "general internalist characterization" of reasons to act, a characterization which claims that "...a fact is a reason for a person to do something if he would be motivated to do it, other things equal, were he to consider that fact rationally" (p.20). It is evident too in the later more specific formulation which states that "p is a reason for S to do A if, and only if, p is a fact about A awareness of which by S, under conditions of rational consideration, would lead S to prefer his doing A to his not doing A, other things equal" (p.81, his italics). This later formulation is perhaps preferable in that it serves to emphasize as well the normative aspect of practical reasons.

Rational consideration -- that which generates the normativity of reasons -- is elucidated in two stages, the first yielding "the initial account", the second an extended account; the latter revising, or qualifying in an important way, the first.

2. The initial account has been likened to the theories of rational choice put forward by Brandt and Rawls1 and it consists largely in a sophisticated version of the recommendation to look before you leap.

There is the obvious requirement that one should have "reflective awareness" of a fact, of its connection to the act in question (p.86). With this condition satisfied a fact is a presumptive reason if it motivates. It is merely presumptive in that it is defeasible; it

may on further scrutiny lose its motivating power or it may be seen to conflict with other, weightier, reasons.

A fact that is a reason need not be universally so; that is, a fact may be peculiarly a reason for one (or more than one) person. But for any fact to be a reason (peculiar or not) it must satisfy "a universally valid standard", viz., that on reflection it motivates the particular agent (p.88). This universality claim figures large in the extended account, which, Darwall argues, reveals reasons as grounded in "...principles valid for all rational agents" (p.88).

A fact that is (presumptively) a reason moves one to a preference. One can, however, be moved to conflicting preferences. Rational decision here requires of the subject that he adopt a standpoint from which he can get both preferences in view, that he assess them "in a unified way" (p.91), a process which reveals an all-things-considered preference.

There is, though, a further wrinkle. We can be the slaves of our preferences (or desires). To resolve this paradox of our being compelled to do what we want, Darwall has recourse to the interesting distinction between first-order and second-order desires, the latter -- "essentially reflective" (p.91) -- taking as their objects first-order desires. One has, or can have, preferences about one's preferences (something which is of a piece with -- and probably finds its purest expression in -- the notion of a preferred life). Pondering the influence (or lack of it) of second-order desires on those of the first order, Darwall notes that in passionate moments the second-order desires may be simply aligned to those of the first and so be "not necessarily rational" (p.93). There is generated, then, the requirement of dispassionate consideration, consideration which corrects for the undue influence of the immediate; that is, it corrects for perspectival

errors in that it allows facts that are temporally or psychologically differently situated their true or proper weight.

Add to this the requirement that one's consideration be informed -- that it be free of misconceptions about its object, that all relevant factors are within its ken etc.; in short, that one's preference survive Brandtian cognitive psychotherapy -- and we have the notion of an informed preference. Finally, the same procedure, carried out now with the object of ordering one's preferences, yields one's informed all-things-considered preferences (p.100).

3. We have the resources now, Darwall feels, to draw some conclusions about rational preference and rational agency. On the initial account what provides a reason for a preference is not the simple fact of preference itself but the fact (or facts) that on due consideration motivate that preference. In this respect the account accords with the assumption which, Darwall claims, underlies formal theories of decision, viz., that preferences are criticizable in terms of reasons. But criticizability of preferences requires a view of the rational agent as one who "...does not merely find herself with preferences" (p.99), who "...is not constituted by whatever ends or preferences he happens to have at any given moment" (p.101). That is to say, it requires a view of the rational agent as one who accepts or rejects the preferences she has and/or adopts new ones on reasons which (on due consideration) she finds compelling. And, for Darwall, this view of the rational agent in turn "...vindicates Rawls's conception of the rational person as having a fundamental interest in rationally choosing his or her own ends" (p.102)2, a conception which (in its turn) "...plays an important role in Rawls's

arguments, both for his theory of justice as fairness and for the rejection of utilitarianism" (p.102).

It is important, then, that this view of the rational agent -- as one whose preferences have the support of reasons in the manner indicated -- be very well grounded; or, more precisely, that what grounds this view of the rational agent -- namely, the criticizability of preferences in terms of reasons -- be itself very well grounded. And, as has been mentioned, Darwall sees that thesis (that preferences are criticizable) as supported by formal decision theory and by the initial account. But there is, he feels, additional support which can derived by appeal to (what he calls) the unity of agency. Let us turn to that appeal first.

4. As we have seen on the initial account, for an agent to come to a rational decision when faced with conflicting preferences he must occupy a standpoint from which he can get these preferences in view and from which he can assess the reasons for and against them. His adopting such a standpoint and his (practical) decision from that standpoint is an instance of his unity of agency. This unity can be viewed "...both synchronically and diachronically" (p.102); as yielding informed all-things-considered preferences both at a time and over time. So, were an individual unable to display or exercise unified agency at a time he would be "...less a single agent" but rather "...two (or more) agents competing against each other" (p.103). The inference appears to be that inasmuch as the individual is a single agent he can criticize his preferences regarding contemporaneous acts or state-of-affairs by weighing the reasons for and against them. The same point is seen (by Darwall) to hold for unified agency over time in that this requires an informed all-things-considered preference for one's life.3

The argument is meant to parallel Kant's transcendental deduction from the fact of knowledge (the unity of experience) to the existence of the conditions necessary for such knowledge. Unity of agency, whether synchronic or diachronic, requires that preferences be criticizable in terms of reasons. If preferences were not criticizable in this way, if they were "completely foreign" to one another, they would pull the individual in different directions and he would have "no possible basis" for deciding between them (p.111). Preferences, therefore, are criticizable in terms of reasons.

It must be admitted that Darwall is, to an extent, right here. On the total uncriticizability scenario, where the individual possesses (or is possessed by) a chaotic bundle of preferences, agency is paralysed. We are not so paralysed; we have unity of agency in that (and to the degree that) our preferences are not so jumbled. One can, for example, act on one's preference for a healthy life by rejecting, or restraining, one's preference for (say) bacon and eggs. And most of our preferences do seem to be comparable in this way. But Darwall's thesis is stronger than this. Though nowhere stated in so many words, it is the thesis that all preferences are criticizable in terms of reasons. And to derive this by way of a transcendental deduction he must make the factual claim that we have (or can have) unity of agency of a very complete sort, that we are never (or need never be) pulled in different directions.4

5. This is a very strong claim and it goes unsupported. As it stands it simply denies Prichard's view of there being cases of non-rational decision-making. Concomitantly it brushes over Sidgwick's worry of there being an unresolvable tension between moral reason and the rationality of self-interest.5 Indeed, according to Nagel, there is a tension within morality itself, between agent-relative and agent-neutral moral reasons and it

ignores this issue too.6 Again (if I can step lightly into a difficult arena), the thesis that we have unified agency in this very complete way seems to run foul of Freudian psychology, a psychology which propounds the counter-thesis that we are subject to opposing urges, to Thanatos and Eros; that we are at once destructive and creative. All of these positions run counter to the position of Darwall either by opposing the factual premise or by opposing the conclusion of the deduction (I find it hard to say which).

But even if we allow, against these positions, that there is unified agency the transcendental deduction does not yield the conclusion that all preferences are criticizable in terms of reasons. For it is consistent with there being unified agency that there be a fundamental set of one or more uncriticizable preferences, a set which is the ultimate arbiter of other preferences, which in fact generates overriding reasons whereby all other preferences can, if necessary, be assessed. If there is more than one preference in the set then, for there to be unity of agency, these preferences must at least not be in conflict and may even -- in a limited sense -- support one another. But the requirement of non-conflict and the possibility of mutual support does not generate the picture Darwall wants.

It may seem that if one has a preference for A and a preference for B one can, apart from the fact of one's having a preference for B, either point out that there is no reason against one's having that preference or cite a reason for one's having it, viz., that it supports (or is supported by) one's preference for A. But one might too strongly construe this "giving of reasons". In the first place for either case the benchmark for criticism is the fact of one's preference for A. That is to say, the fact that one has the preference for A either fails to generate a reason against, or generates a reason for, one's preference for B. At this level preferences generate reasons and not, as Darwall

would have it, vice versa. In the second place it is misleading to talk in terms of a reason for a preference for B. That preference is, ex hypothesi, fundamental and so not derived. As such it simply stands and is not adopted for a reason. On this picture there is unity of agency because it happens that one's fundamental preferences do not collide.

The appeal to the unity of agency fails, on two counts. The appeal is itself controversial and, even were we to allow it, the desired (strong) conclusion does not fall out.

6. The latter point applies too to one of the other planks for the claim that all preferences are criticizable in terms of reasons, namely, that formal decision theory presupposes the correctness of that claim. We can concede, as before, that most of our preferences can be compared and connected (and are subject to the transitivity requirement) but we need not allow that we can give reasons for all of them.7 Decision theory works just as well on the assumption that there is a basement level of preferences, either one fundamental preference or, if more than one, a base preference set where the constituents are consistent. Should preferences be inconsistent at this level then decision theory is ineffectual and will not be called on.

7. Let us return now to the initial account of rational consideration. Recall that on that account the reasons for our actual preferences are "...the facts that motivate them, and not the simple fact of preference itself.." (p.98). Our actual preferences -- be they preferences for acts or for states-of-affairs -- have the support of reasons were it that we would yet be motivated to them on full rational reflection as per the initial account. So, for example, there is a reason for one's preference to act in a certain way (e.g. one's

preference to A) if there is a fact (e.g. that to A is to R) that, on adequate reflection, would motivate one to act in that way (e.g. would move one to A). And, of course, if there is no such motivating fact there is no reason for one's preference. In short, on the initial account preferences are criticizable in the light of reasons for or against them.

But again there is the question as to whether or not all (and not merely most) preferences are so criticizable. That all are is a position Darwall must hold if he is to see criticizability by reasons as vindicating Rawls' conception of the rational agent as having a highest-order interest in revising his final ends and, indeed, if he is to hold to Kant's thesis that rational agency is self-governance. Now the manner in which the initial account is seen to vindicate Rawls' conception is that it encourages a picture of the rational person as one who, rather than being constituted simply by the preferences he has, can revise his actual preferences by reasons he finds compelling. This is a plausible picture but it is one which allows (as does the account of rational consideration which encourages it) our seeing the rational person as one who is yet driven by latent (or dispositional) preferences. That is, we can construe "actual" as "occurrent" and see the agent as being able to revise his occurrent preferences by reasons generated by latent ones, by those more fundamental preferences which are revealed through appropriate reflection. It is important for Darwall's argument that the plausibility of his picture of the rational agent as one who can revise his preferences does not rest on our seeing the picture this way, for on this interpretation we are led, eventually if not immediately, to accept the existence in the individual of a base set of uncriticizable preferences and, thereby, to deny the possibility of his revising his final ends in the deep sense intended by Rawls and, concomitantly, to deny the possibility of personal autonomy as envisaged by Kant.

8. But it is this interpretation which -- ironically, perhaps, if not surprisingly -- naturally emerges from Darwall's formulation of Humean internalism. For it is natural to ask how it is that a fact can motivate (and so be a reason) unless it taps a preference. On the formal example, rational consideration of the fact that to A is to R moves one to A. Which is to say that one's being moved to A (on rational consideration) depends on the fact that to A is to R. It is hard (if not impossible) to see how that fact generates motivation unless there is already a preference for R-ing. Indeed that rational consideration reveals this latent or dispositional preference sits well with Darwall's explication of Hume's internalist analysis of facts having the property of being reasons: that they have this property is "...rightly judged by....noting (on reflection) a condition of one's own will" (p.56). That one's will be in that condition appears mysterious unless there is postulated something suspiciously like a latent preference.

Darwall does allow (as an internalist must) that something's being a reason depends on the agent's "motivational capacity" (p.42) but he cannot8 allow this to consist, even partly, in latent preference, for to allow that would be to disallow the possibility of autonomy, in that we must now see the individual as being slave to his desires, albeit latent ones.

Now Darwall does try to argue -- via the example of Roberta (pp.39-42) -- that a consideration's being a reason does not depend on its being linked to a prior desire (presumably whether occurrent or dispositional) of the agent. The example is unconvincing, however, and I shall not deal with it.9 It does, however, in its attempt to disengage reasons from prior motivation, point to a possible strategy and that is to claim that the process of rational consideration can itself generate motivation, that the very

process can independently yield preferences and reasons. And Darwall does seem to make this claim (p.39). But by virtue of what can it do this? How is it that this process -- coming to know (realizing thro' rational consideration) that e.g. P is a fact about A -- can move one to A if it is to do so other than by way of a preference for P? It is hard to see how an answer can be given here.10

In fact Darwall's understanding of internalism seems, itself, to preclude this kind of answer. Holding, against Nagel, that a fact that will be a reason for an agent need not be a reason for him now, he says that "...rational consideration of a fact may move [the agent] at some point but not at another" (p.108). The same rational consideration? It seems so; "...it may simply be the case that a consideration that would motivate a person on reflection at some time would not at some other time, even if the person were to give it the same careful imaginative consideration informed by no further experience" (p.108; his italics). It cannot be then that the process of reflection itself generates the motivation.

The initial account, then, is consistent with the claim that not all preferences are criticizable in terms of reasons; that is, it is consistent with (and indeed encourages) a view of the rational agent as one who can indeed revise (most of) her preferences in the light of reasons but who does so in virtue of one or more fundamental and uncriticizable preferences. But this picture of rational agency does not vindicate (as the more radical picture might) Rawls' conception of the rational person.

9. If my arguments are sound then all three planks of Darwall's argument to that conception -- the appeals to decision theory, to unity of agency, and to the initial account -- have been removed. Or, if one sees (as I do) the initial account as in fact including all

three appeals, then what I hope to have done is to have cast doubt on the distinctly Rawlsian flavour of the latter part of that account. The existence of such doubt is important in that Darwall appeals in a final stage of his argument to the "good independent grounds" we have for the initial account, which means (for him) that we have grounds "...to embrace a conception of the rational person as having a "highest order interest," as Rawls puts it, in revising her ends by reflection on reasons she finds compelling.." (p.234). I have already touched on the work this appeal is meant to do; as to the impact of a (legitimate) doubt on its intended function, I shall leave this matter for a later part of this paper.

10. I hope too to have indicated a tension in Darwall's position, to have justified my claim 3. in the last chapter. Darwall's Kantian view -- that moral agency is expressive of self-governance -- is a deeply attractive one. Freedom, it seems, is a logical prerequisite of moral agency. To assess an agent as moral/immoral in his performance of an act is, it seems, to imply that he chose to do it, that he had the capacity to refrain from doing it.11 Though Darwall's attack on the view that reasons rest in some way on the agent's desires can be seen (on the assumption that most desires are selfish ones) as an attack on rational egoism (broadly conceived) it is better seen, I think, as an attack on a view which disallows autonomy and so disallows moral agency.

My interpretation is this:

1. Moral assessment of character makes sense only on the assumption that the agent chooses to act as he does. It is a logical requirement of an act's being moral/immoral (as against its being e.g. fortunate/unfortunate) that it be freely chosen.

2. Rational agency is plausibly understood as full self-governance, as governance by principles one chooses (on adequate reflection) to be guided by. To be a rational agent is to be free of non-self-imposed guides to action, including those provided by the desires one happens to have.

3. Our notion of moral agency requires this picture of the rational agent. If an individual cannot be free in the manner depicted in 2 then (by 1) it is not possible for him to be moral.

4. An identification of the rational with the moral does not fall out from 1 and 2. The question remains whether the free agent would choose principles of action which disbar moral action or ones which allow it or ones which require it, where such action opposes self-interest or where it is contrary to current inclinations etc.12

5. Darwall's intent is to show that the rational agent would choose principles that require (under certain conditions) moral action; that moral agency issues from (rational) self-governance.

Now an externalist view of reasons may also be considered, though in a somewhat different way, as hostile to self-governance. Falk suggests that the mature individual becomes resistant to his interlocutor's "subjective recommendation use" of the moral "ought".13 This resistance would, I think, be extended by an internalist (as it is by Falk) to cover any use of the moral "ought" which depicts the moral demand as a demand from without. According to Frankena, R.M. Blake (an externalist) saw the doctrine that the moral will is autonomous as "incompatible with the existence of any categorical imperative" and so "...should be repudiated by externalists..".14 Blake's position is an inversion of the internalist one. Either way the view is (or could be) that

autonomy and externalism are mutually exclusive, that what I have suggested is a precondition of moral agency, viz., autonomy, sits better with internalism.

My intent has been to show that, au contraire, internalism as formulated thus far (per the initial account) suggests a base preference set and so encourages a view of the individual as heteronomous, as incapable of moral or rational agency in the deep sense intended. Admittedly this reading can be resisted by postulating a motivational externalism (something Darwall hints at), but not only is the latter position hard to defend, Darwall's claim -- that the same rational consideration might move one at one time and not at another -- seems to preclude it.15

11. The objections raised thus far apply to the initial account of rational consideration. The extended account is meant to modify the initial account while yet preserving it as "...at least a partial account of practical reason" (p.234). We shall have to consider the extended account and see the way in which it is linked to the initial account to form Darwall's full thesis before we can see the bearing these objections have on the full thesis.

Footnotes to Chapter 8.

1. By Hill, op.cit., p.611. See footnote 11, Chapter 7, here.

2. This "fundamental interest" is but one of the two highest-order interests ascribed to the moral agent by Rawls in his "model-conception" of a moral person. It is important to see that Darwall's vindication of this part of Rawls' conception lies in the initial account.

The other highest-order interest of the moral person is to realize and exercise his "sense of justice". See Rawls, 1980, p.520/p.525. Darwall's only explicit reference to this aspect of moral personality is in a footnote (p.231) where he says of the choosers behind the "thicker" veil that "...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", this supposition being "...comparable to Rawls's assumption that his principles are chosen for beings with a sense of justice". But Rawls' "assumption" is not merely an assumption. Rather it is an appeal to a notion either implicitly affirmed in or acceptable to "...the public culture of a democratic society" (Rawls, 1980, p.518).

Nor will it do for Darwall merely to suppose. On the contrary I take his argument to the original position, if it be successful, as showing that ideally rational beings would be guided by the principles chosen; that it provides a (rational) vindication of the Rawlsian moral person's "sense of justice". Again it is noteworthy that this one of the two moral powers, unlike the other, is vindicated (if it be vindicated) via the extended account of rational consideration.

3. This idea -- the idea of an informed all-things-considered preference for one's life -- is, Darwall feels, well captured by Rawls' notion of a rational life-plan. But whereas Rawls identifies this plan with what he calls "the person's good" or "a rational conception of the person's good" Darwall feels that what Rawls intends is better expressed by the phrase "a rational conception of the good life for that person" in that this phrase, unlike the ordinary meaning of "a person's own good", naturally allows altruistic behaviour (for example) if this is the result of the agent's "...informed, free choice" (p.105).

4. that is, pulled in different directions in the deep sense intended by Darwall. Obviously conflicting preferences pull us in different directions. His claim is that we have unity of agency in that any such conflict is capable of resolution by a weighing of reasons.

Presumably an individual can have unity of agency and not exercise it; that is, he might not do this weighing. And it is conceivable that one might find one's conflicting preferences equally supported by reasons; nonetheless they stand "compared and connected".

5. I take this from Mackie, op.cit., p.190.

6. See, for example, pp.175-185 in Thomas Nagel, The View From Nowhere, 1986.

In his review of this work Darwall does indicate a rationale, within his character-based approach, for agent-relative obligations; the existence of any such obligation depending on "...whether it would be required by principles by which a morally autonomous person would govern herself" (p.153, Darwall, 1987). And not only is it likely that some such obligations would be required in this way, but because the approach is character-based these agent-relative moral reasons are not undermined, as they are on an end-based view, by considerations of overall value. See pp.153-156, ibid.

This rationale is provided however by Darwall's extended account of rational consideration. The initial account is meant to stand, to some extent, independently of that account. Specifically it is meant to provide "good independent grounds" for Rawls's conception of the rational person (p.234), the very conception here at issue. The initial

account, in vindicating this conception, is seen to provide the end --the highest order interest of the rational person -- in view of which the choice from behind the veil is made. Darwall points out that since the principles employed in the choice are means-end principles (not substantive principles but principles of relative rationality) the overall account avoids circularity (p.234, footnote). For the same reason -- to avoid circularity in the overall account -- it is necessary that the end stand independently of the substantive rational principles chosen to realize it.

I shall have a little more to say about the linkage of the initial and extended accounts. It is enough, for now, to have pointed out that the first must be independent of the second in the way described and that, consequently, Darwall's claim that there is unity of agency ignores the tension between agent-relative and agent-neutral moral reasons.

7. A psychology of preference is bound to be a complicated affair and I do not pretend to be able to provide one. I want merely to suggest that there is an alternative schema to that provided by Darwall.

8. as he seems to suggest he can, by assimilating a "weak" notion of desire to the agent's motivational capacity. See p.42.

9. For discussion and criticism of this example see Hill, op.cit., pp.606/607; also Alfred R. Mele, "Motivational Internalism: The Powers and Limits of Practical Reasoning", Philosophia (Israel), 19 (4), December 1989, pp.427/428.

10. Mele says of motivational externalism -- the view embodied in the claim I've attributed to Darwall -- that it "...smacks of creation ex nihilo" (p.429, op.cit.).

Hitherto, in my text, the terms "internalism" and "externalism" refer to positions about the nature of reasons; to what we might call conceptual theses. (I shall continue this practice, the bare terms referring to the opposing conceptual theses). These positions Mele dubs "normative internalism/externalism". Darwall, for example, is a normative internalist. Theses like these are to be distinguished from empirical theses about the source of the efficacy of practical reasoning; that is, theses about the "causal history" of the motivation which emerges from a piece of practical reasoning. These positions Mele calls "motivational internalism/externalism". Mele, in that he holds that antecedent motivation is always part of the causal history, is a motivational internalist. I have benefited from Mele's pointing up the distinction and from his criticism of motivational externalism.

In my view, Kantians like Falk and Darwall, and like Rawls with respect to his model-conception of the moral person, must hold to motivational externalism. If rational agency is self-governance then practical reasoning must rise above any preference or desire the agent just happens to have. Mele makes a case against motivational externalism.

He also argues that his position -- motivational internalism (MI) -- has no direct implications as to what normative thesis one must hold, specifically that it does not entail a normative internalism (NI); that to get from MI to NI one at least needs a (debatable) proposition R (See his pp.431-434). I will not discuss the correctness of his claim here. I cite the argument because it is illustrative of how the water can get muddied. His R is very like, and on one construal is identical to, Darwall's normative internalism. His NI is

a version of what Darwall calls the Desire-Based-Reasons thesis. It is in fact Version III if we take this to include latent desires (see p.38, Darwall). The argument then is, in Darwallian terms, that motivational and normative internalism together may yield a Desire-Based-Reasons thesis.

More apropos to my current argument against Darwall is another claim of Mele's, viz., that motivational internalism does not lead us to Hume's characterization of reason as being the slave of the passions. His being right here would mean that Kantians are not forced to motivational externalism. I can only say that his argument, by way of the example of Theresa, a young student who herself "decides" on a particular conception of what it is for her to live an important life, is unconvincing (see his pp.429-431).

11. Nagel, in his essay "Moral Luck", points up the paradox, the apparent absurdity, of our employment, in the shadow of a scientific world-view (a view of the world as a succession of determined events), of moral epithets to individuals. See his Mortal Questions, 1979, pp.24-38. Nonetheless he regards as "intuitively unacceptable" Kant's claim that what people are like through no fault or doing of their own is not a proper object of moral assessment; this because we do habitually and naturally assess people in this way (p.33). That we do Nagel attributes to our extending to others the internal view we have of ourselves (p.32), that is, he treats the phenomenon (moral luck) along the lines of the subjective/objective split.

Alternatively the paradox can, I think, be seen as issuing from our ambivalence with respect to morality and moral character. I can only hint at an explanation here. It would hinge on our seeing morality, on the one hand, as (`a la Mackie) having a function, and moral character, on the other, as (`a la Kant) being estimable independently of the consequences that issue from it. It is not clear how this explanation might sit with Nagel's treatment, or if it could.

12. I take these alternatives from Nagel, 1986, p.200.

13. Falk, op.cit., p.508.

14. Frankena, op.cit., p.64. Frankena is "...not persuaded that a recognition of autonomy necessarily leads to internalism" (p.64).

15. It might be worthwhile to point out in addition that, so far as I can see, absent the appeals in the initial account to decision theory and the unity of agency, there is nothing in the formulation of internalism itself which prevents an agent's having intransitive preferences or which guarantees that the reflective process, properly carried out, will always yield an unambivalent conclusion.

CHAPTER 9

# THE NORMATIVE ASPECT OF REASONS

1. Though the initial account can yield reasons other than self-regarding ones, e.g., one may have an informed individual preference that a person (or persons) other than oneself flourish, it "...appears to make the standpoint of rational consideration the agent's own" (p.112). That is, Darwall sees it as offering "...no challenge ...to a fundamentally self-centred view of practical reason" (p.112). The extended account, by contrast, is intended to move us away from this "personal standpoint"; is intended to move us to accepting the "intersubjective standpoint" of "a rational agent as such" as the point of view "...from which all reasons, including those grounded in what motivates an agent from his own point of view, are ultimately assessed" (p.113).

2. The first move away from the agent's "own particular point of view" (p.117) as the perspective from which facts may be reckoned (as generating reasons) is via the distinction made by Nagel (in The Possibility of Altruism) between subjective and objective considerations. Following Nagel a consideration is subjective "...if its expression contains a "free agent variable", a phrase whose reference is given by the agent for whom it is supposed to be [a] reason" (p.118). A consideration that makes "no essential reference" to the agent himself as such is, by contrast, objective (p.130).1

As Darwall notes (p.118), Nagel held to the "thesis of objectivity"; that is, he held that only objective considerations could provide reasons. But, as Darwall also points out

(pp.119/130), this thesis alone will not yield the conclusion that Nagel was after, viz., that something's being a reason for a person bears on what others ought to do. For that conclusion the thesis of objectivity needs to be supplemented by the "thesis of universality", the thesis that for something to be a reason for someone it must be a reason for anyone "...in relevantly similar circumstances" (p.117).

Though he rejects this strong "Nagelian Objectivity" (rejecting both the universality thesis and the thesis of objectivity), Darwall is intent on showing not only that objective considerations do indeed provide reasons but that some of them (viz., moral considerations) commonly provide overriding ones.2

Now, that objective considerations do in fact weigh with humans we can allow. But should they? To that question -- "...why should a person even take into account facts that must be considered from an impersonal standpoint?" -- Darwall's reply is that, on an internalist theory of reasons and in light of the initial account, "...there is no more reason to take into account facts that can be assessed only from a personal standpoint than there is to take into account facts that must be considered impersonally" (p.135).3

What the initial account does leave open is the question of the relative weights of subjective and objective considerations as reasons. On that account "...subjective reasons may generally be weightier reasons than objective ones", or, alternatively put, a person "...may systematically rank personal preferences over impersonal ones" (p.136).4

The extended account is meant to block any such morally untoward possibility. As I have indicated, the burden of that account is to move us away from the initial account and its emphasis on the personal standpoint of the agent to an account of rational consideration which takes as central the "intersubjective standpoint" of the

rational agent as such. It will discharge that burden, Darwall intends, through its appeal to the normativity of reasons.

3. The motivational aspect of reasons and practical reasoning, on Darwall's internalist analysis, has been well canvassed. This aspect is one element of his background theory of reasons and, as we have seen, it allows that a reason for one person need not be a reason for another, nor indeed for the same person at a different time. Reasons may be idiosyncratic as to persons and times. Yet, in every case where it is said of a person at a time that he ought (that he has a reason) to do something, what is said is intended to have the same kind (if not the same degree) of force. That is to say our discourse suggests that we understand practical reasoning to have, as well, a normative aspect; that we see reasons (idiosyncratic or not) as authoritatively action-guiding. As Darwall says, without this normative aspect "...there would be nothing to distinguish reasons for someone to do something from reasons why he did or will do it.." (p.201).

That rational consideration and the reasons to which it gives rise are normative is suggested too, Darwall holds, by the way our familiar conception of rationality accords with the notion of a normative system; with the notion of a system of norms (or standards, or principles) which are seen to apply to its (the system's) subjects in that they can guide and appraise their and others' relevant actions -- the system's objects -- by them (pp.204-208).

I think Darwall is right here. We do, in our discourse on reasons etc., presuppose a normative system. But are we thereby slipping, as Falk claims we do in our thinking on moral obligation,5 into an externalism with respect to rationality? As

Darwall emphasizes, a very attractive feature of internalism is that, unlike externalism, it leaves no mystery about the agent's having the ability to do what he has reason to do. Motivation is built in. An externalist might retort however that his own view, though perhaps weak on motivation, is, unlike internalism, strong on justification; that it is a strength of his view that normativity is built in. He might ask -- how does an internalist theory of reasons capture their normative aspect?

4. It is important to see that, for Darwall, reasons are derivative of rational consideration. On his internalist view reasons are not objective entities or properties (of facts) that we discover "out there" when we deliberate.6 Rather they are discovered via, or (one might say) they are generated by, a required deliberative process. And to say that the process is required is to say -- as Darwall holds -- that all rational agents are subject to the norm or norms which define it; i.e., that these one or more procedural principles are universal.

But Darwall goes further than this. Procedural principles (the required process) give rise to substantive principles of practical reason and these substantive principles too are considered to be universal. An example of a substantive principle -- one that is considered and rejected by Darwall -- is the principle of maximizing individual utility. Another substantive principle might enjoin conduct that accords with moral requirements, and Darwall is of course concerned to show that a principle or principles of that sort do issue from the process of rational consideration.

On the other hand, as he has emphasized, the reasons generated by rational consideration need not be universal, need not be reasons for all rational agents. Facts which move me need not, though we perform the same deliberative process, move you.

This leads him to say that on an internalist account "...a fact's being a reason is not grounded in universal reason-specifying principles.." (p.210).

And, it seems, he must insist on this last if he is to preserve his account as an internalist one, an account which avoids the nonnaturalist metaphysical commitments of externalism and which, in contrast to externalism, gives "a deeply satisfying" account of rational motivation.7 But, as he also notes, motivation by itself is not enough -- "...the degree of motivation that a consideration provides is not identical with its weight as a reason" (p.81). To hold that it is is to fall back into a desire-based theory. Motivation of itself, then, does not supply rational grounds for action; does not provide the required justificatory or prescriptive force. Darwall's account must indicate the locus of that force while yet not teetering to an externalism which asserts "...that there are simply some facts that have the intrinsic property of being action-guiding.." (p.81). His solution is to locate the normativity of (action on) reasons in an ideal of rational consideration. And in Impartial Reason this latter notion is introduced and given specific content by way of the concept of a rational normative system, a system to which, it seems, we appeal -- which we seem to presuppose -- in our discourse on reasons and rationality.

5. Darwall is, I think, attempting to walk a very thin line here. Consider the nature of substantive principles. Darwall cannot rest -- and does not -- in the position that all substantive principles could be non-reason-specifying. A substantive principle requiring moral conduct while yet not specifying reasons for all agents (in like situations) would simply be malapropos. Substantive principles, it turns out, are of two kinds, those which specify relevant facts (a principle of right conduct, for example, is reason-specifying) and those that merely direct an agent to do what he is moved to do on consideration as per

the initial account (pp.236/7). With respect to our concern, viz., the authenticity of moral requirements, Darwall's theory and an externalism may yet offer competing rationales, but with respect to the conduct that is morally required the two views may well be extensionally equivalent. Though this in itself is not surprising -- externalists and internalists have after all been reared to the same moral requirements -- the contrast between the two views is now less marked, for on Darwall's internalism variation in motivational capacity is allowed full rein only with respect to such as personal projects etc., and not with respect to morally required conduct. Now, given our sense of what it is for a certain kind of conduct to be morally required, this is as it should be. What is not clear -- and what needs to be argued for -- is how this universality of moral requirement can be accommodated on a view which ties reasons to an agent's motivation.

Darwall might say that such a view can so accommodate -- and this is the argument (at least implicit in his thought) -- because each individual qua rational agent has the same capacity here. But even if we allow this, it is evident that across fully-fleshed individuals there is considerable variation in motivational economy. It is at least a respectable empirical proposition that some individuals find it easier to accede to a particular moral demand, or to moral demands generally, than do others. Now externalism will be unequivocal with respect to the existence of any such demand; if it holds for one it holds for all, irrespective of an individual's motivational capacity. It may acknowledge variation in capacity -- though perhaps not without difficulty -- through a corresponding variation in assignations of responsibility and blame. The moral demand will however in each case be the same, the moral charge perhaps being diluted appropriately. And it seems reasonable that actual variation in capacity be somehow recognized.

But if it is to be recognized on Darwall's position what is to stop that position from drifting to one which is subject to Frankena's charge against internalism: that it runs the risk "...of having to trim [moral] obligation to the size of individual motives"?8 We might allow that simply qua rational agents we are alike in motivational capacity. But if for fully-fleshed individuals we have a sliding scale of moral charge because of actual variation in motivational susceptibility, why not have -- for fully-fleshed individuals -- a sliding scale of moral demand for the very same reason?

Now Darwall is vehement in his opposition to this. He has made it abundantly clear that on his view an individual's de facto preferences, and the actual strengths of those preferences, do not provide the measure of reasons for him to act. Reasons have a normative as well as a motivational aspect. But will it do for Darwall -- in order to elicit this normative aspect, and in such a way as to debar a possible drift to a sliding scale morality -- to merely refer us to the presupposition underlying our talk of reasons and rationality? For that presupposition may be -- and likely is -- an externalist one, a presupposition of a standard or a set of standards which stands over against, and makes demands on, each individual agent. Indeed, on the face of it, it appears to be precisely the rational analogue to that "...externalist and non-psychological conception of [the moral] "ought" " with which in our ordinary moral thinking, Falk says, the internalist conception is blended, and this "perhaps" for fear of the latter's "...empiricist and subjectivist implications..".9

Even if Darwall resists my way of introducing this worry, the worry itself remains. It could, I suppose, be introduced in a different manner. And the worry is: how to capture the normativity of reasons on a purely internalist approach? Now, as I've pointed out, Darwall's intent is to do so via the notion of ideal rational consideration, and that notion is

to be filled out -- is to be given content by -- our considering the kind of reflection an ideally rational agent would engage in. Unfortunately, in Impartial Reason, the ideally rational agent is presented as one who is (self-critically) an internally self-identified subject of the rational normative system; that is, he is defined in terms of a notion -- a rational normative system -- which is, there is reason to believe, an externalist one. For the rational normative system (RNS) provides norms (or principles), both procedural and substantive; that is, provides standards, for its subjects, of deliberative process and action. And it is, after all, of the nature of standards that they stand over against that to which they apply.

6. This criticism -- that Darwall's argument is infected with externalism, that it derives plausibility by appeal to a view which he rejects -- stems from what is, I think, a legitimate reading of passages in Impartial Reason. But perhaps this reading or, more precisely, perhaps the passages which encourage this reading, merely obscure rather than vitiate the argument. Let us take that tack. Let us see the RNS (however it might be filled out) as standing over against fully fleshed individuals, its subjects; as independent of their actual desires and predilections. This is, I think, in line with Darwall's view and it is implied in his remark as to the "minimal normative content" of rational norms: that they are "...principles or standards that are normative for all agents with rational capacity" (p.208). What he does not admit -- and cannot, if he is to preserve his account as an internalist one -- is that the standards are independent too of the ideally rational agent. In fact it is precisely the opposite thesis that forms the core of Darwall's internalism; the thesis that any putative rational norm must, to pass muster, to

be a rational norm, be acceptable on reflection to an ideally rational agent. The ideally rational agent himself chooses the norms he is to be guided by.

If this conception could be made to work then, it would seem, the normative element characteristic of an externalist account would be captured on an internalist one. And we would have, it seems, a happy blend of externalism and internalism. Rational principles, both of process and substance, stand over against fully-fleshed individuals and so have an externalist aspect. But at a deeper level these principles are internal. They are internal to the ideally rational agent or to the rational agent as such; that is, the principles are such that each individual, simply qua rational agent, would be moved to adopt them.10

How then to conceive the ideally rational agent? He has two salient features. He is self-critical in that he is "...inclined to assess critically even the norms he ordinarily takes as ultimate standards for rational choice" (p.219). But how can he critically assess these given that they are purportedly ultimate standards? By asking "...whether it would be rational according to a given principle to choose to act on it" (p.219).11 Secondly, he (or she) is an internally self-identified subject of the rational normative system (for short, an ISIS of the RNS). An ISIS is one for whom the norms of the RNS "...have motivational and attitudinal weight" (p.213); she "...accepts the idea that norms...ought rationally to be acted on by all agents, including herself, simply because they, and she, are subject to them" (p.224). He is one who acts on facts (putative reasons) not only because the specific facts move him but because he judges these facts to be reasons -- "...his judgement of them as reasons is the controlling factor" (p.214). And he is such that "...appraisals of himself that he makes from the intersubjective standpoint of a subject of rational norms figure prominently in his own self-esteem" (p.214).

Does this characterization reveal the theory as being infected with externalism at the deeper, supposedly internalist, level? Perhaps not, if we view the ISIS this way: as one whose dominant concern is to deliberate and act justifiably (rationally), however it might be to do these things. Moreover it is he who, as self-critical, decides how it is to do these things. As Darwall elsewhere puts it, "...no fact exists about what are correct principles of rational conduct that is independent of what would be acceptable on reflection to fully self-critical and self-governing rational agents".12

Initially, then, the ISIS has but the bare idea of an RNS, the purely formal notion of ideal rational consideration and the principles of rational conduct that issue from it. And it is he who is to, non-arbitrarily, give the form content. (Concomitantly it is, as it were, through the eyes of the ISIS that Darwall will make form fertile; that he will have a formal theory of rationality yield some substantive conclusions). His choice of substantive principles will be non-arbitrary, Darwall intends to show, because it will be under the constraint that it be made from the "...impartial standpoint of an arbitrary rational agent.."(p.226).

7. A basic premise in the argument to that constraint is that whatever substantive principles there are (or that might be chosen) they are principles that are normative for all rational agents (p.208). This "relatively uncontroversial" claim13 as to the universality of rational principles has, however, been strongly objected to by two of Darwall's reviewers.

B.C. Postow argues that the claim is inadequately defended14 and he puts forward two rival claimants to Darwall's ISIS for the title of ideally rational agent, neither of whom claim universality for the (different) principles by which they choose to guide

themselves.15 Terzis, too, argues that Darwall has failed to show "...that there is a single set or system of norms to which all agents must appeal".16 To this last Darwall replies that if reasons have a normative character and are therefore "...tied to the "ought" of rationality" then "...the relevant norms, it would seem, are rational norms, norms that apply to all agents with rational competence, much as legal and moral norms apply to agents with their respective competences".17

My sympathies are with Darwall here. For suppose a number of rational normative systems. Then the same action by the same agent can be criticized (or evaluated) differently on the different systems. On one system he may have a reason for his action, on another not. This robs "reason to act" of its force. It might be objected that, for any agent who identifies with a particular RNS, it is the norms of that RNS that he applies universally in his criticism of his and others' actions and so for each agent the force is maintained. But an agent cannot (non-arbitrarily) apply norms universally, champion their force, and yet recognize that there are, or there may well be, a number of systems and no way to decide between them. Rational criticism is unfounded if there is a relativism at this level.

Nor are things any better if one allows, following Postow, that an agent may embrace as fundamental a principle that is, and that he sees as, a particular principle; a principle that differs from that which may, justifiably (the agent feels), be adopted by other agents.18

In either case, whether ultimate principles are held to be merely particular or, if held by agents to be universal, it is yet maintained that there are (or may be) a number of RNS's, we must forbear from much of our usual criticism and our talk of reasons etc. loses much and maybe all of its force.

The trouble is that the appeal from the normativity of reasons presupposed in our ordinary discourse to a single set of universal rational principles is the same appeal I characterized earlier as being an externalist one and it is now being made at a deeper level. Made at this level it suggests that the RNS not only stands over against the fully-fleshed individual but stands over too against the ideally rational agent. This last, of course, runs directly counter to internalism in that the picture is that of an independently correct RNS determining what it is to be an ideally rational agent, rather than vice versa as Darwall would have it.

In short, the argument runs in the wrong direction. Rather than arguing from the (assumed) normativity of reasons to a single RNS -- which, I have tried to point out, is an externalist move -- what Darwall needs to do is to show that on an internalist analysis there can be but one RNS. In showing this he would show that internalism indeed captures the normativity of reasons.

Such an argument is available but, though present in Darwall's thought, it is not properly exploited in Impartial Reason. Moreover, as we shall see when we get to it, it moves Darwall nearer to Kant's "...difficult metaphysical doctrine of noumenal freedom " than perhaps he would like.19

8. For the moment, to follow the explicit argument of Impartial Reason, let us assume arguendo the universality of rational principles; let us assume that there can be but one RNS. Let us allow too that an ideally rational agent is a self-critical ISIS.

Since it is of the nature of a rational principle or norm that it function as a guide for subjects, we may say that something is a rational principle only if it could be a guide

to rational agents. From this and from the ideal rationality of the ISIS, Darwall derives the guidance principle:

A principle is a rational principle only if an ISIS of the RNS could regard it as such (p.221).

How then does the ISIS regard rational principles? From the definition of an ISIS we know that she is motivated to act on a rational principle because she sees it as a rational principle; because she sees it as normative for her (p.222). But as well as this causative link between norms and motivation there is a justificatory link. Her action on a rational norm is justified (the ISIS judges) by its being a rational norm; by its being something "...on which agents ought rationally to act" (p.222). That is to say, the justification for her so acting (she sees) is not conditional on her having the desire to act rationally but rests merely on the status of the principles on which she acts (p.224). Indeed "...[her] desire to act on rational norms itself has a justification: their being rational norms" (p.225).20

Now (the argument continues) suppose an ISIS judges P to be a rational principle. It is this judgement -- that P is a rational principle -- that "...motivates the ISIS's preference that he act on P" (p.225). But the judgement that a principle is a rational principle is the judgement that the principle is normative for all agents; it is the judgement that it is a principle on which all agents ought rationally to act. It is, then, an "impersonal" judgement (p.210); a judgement made from the "impersonal standpoint" (p.225).

This means that insofar as an ISIS is motivated by his judgement that P is a principle on which any agent ought rationally to act, he is motivated from an impersonal standpoint. Any motivation he has, as an ISIS, for preferring that he act on P will be

impersonal, therefore. He will be motivated to prefer that he, qua rational agent, act on P. Consequently, he will have the very same motivation to prefer that any agent act on P, indeed, that all do so. .....

It follows, therefore, that a necessary condition of an ISIS's regarding P to be a rational principle is that he impersonally prefer that all agents act on P. Moreover, since his preference that all agents act on P is grounded in P's being a rational principle, it will not be conditional on anything not relevant to the latter judgement. 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. (p.225/226).

If we exercise the guidance principle on this consequence we get this requirement on a principle's being a rational principle: that it be a principle that the ISIS prefers from the impartial standpoint of an arbitrary rational agent that all agents act on it.21

But, in that the ISIS will choose (prefer) simply qua rational agent and impartially, he will in effect choose from the original position, albeit one with a thicker veil than in Rawls' characterization of that position. We may say then that a principle can be a rational one only if it is such that it would be the object of choice from behind this (thicker) veil.22

This conclusion, were it to stand, would be a momentous one. It would provide rational vindication for Rawls' original position and it would, of course, provide an adequate reply to our moral sceptic. However it doesn't stand and to see why -- why the

argument doesn't go through -- we need first to hearken back to the crucial question of standpoints, of points of view.

9. Recall that, for Darwall, the first move away from the agent's own particular point of view was via Nagel's distinction between subjective and objective reasons. Now Nagel called "impersonal" the standpoint from which (as he saw it in The Possibility of Altruism) all reasons are generated and, in that he held to both the thesis of objectivity and the thesis of universality, this appellation is somewhat appropriate. But, even so, to treat the problem of points of view in terms of a sharp dichotomy -- though such a rendering may serve to set the problem in high relief -- is to aid and abet our seeing as one the several distinct moments in the passage from the one extreme to the other.

And, indeed, in the later Mortal Questions, Nagel advises us to see the original dichotomy as instead polar opposites on a continuum which at the impersonal end is asymptotic. He says of this continuum:

At one end is the point of view of a particular individual, having a specific constitution, situation, and relation to the rest of the world. From here the direction of movement towards greater objectivity involves, first, abstraction from the individual's specific spatial, temporal, and personal position in the world, then from the features that distinguish him from other humans, then gradually from the forms of perception and action characteristic of humans, .....There is probably no end-point to this process ....23

Now what is important for us to note is that the several abstractive steps in the movement from personal to impersonal differ from one another in kind as well as in quantity or degree. But that they differ in kind is still not sharply brought out in Nagel's treatment.24

The distinction I have in mind is that between cases where the abstractive step consists in the deletion from the consideration of a reference to the considerer and cases where what is discarded is either some idiosyncratic or some partially identifying characteristic of the considerer; i.e., cases where what is abstracted from is a feature that distinguishes him from all or some other agents.

The movement from a subjective consideration to an objective one is then an instance of an abstractive step of the first kind while not being an instance of an abstractive step of the second kind. An instance of the second kind is the case where the considerer abstracts from, for example, his penchant for taking risks.

Now for Nagel a step of either kind is subsumed under the movement from the personal to the impersonal standpoint. And Darwall follows him in this. In marking the contrast between a personal and an impersonal standpoint, he writes:

[In a personal preference] 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. A preference is no less a person's preference for being impersonal (p.133).

Now what we have here is clearly a case of an abstractive step of the first and not of the second kind and what we need is a nomenclature which distinguishes the two sorts of movement. Since the case here is an instance of

preference/consideration/standpoint that we usually call "impartial" I propose that we call it so to mark the relevant difference to cases where steps of the second kind are taken, steps which move the agent towards what, I think, is more properly called the "impersonal" standpoint since these steps abstract more and more from the features which mark him out from other agents.25

10. Let us examine Darwall's argument in the light of this distinction. He writes:

...the judgement that a principle is a rational one is itself appropriately made from an impersonal, rather than the agent's own personal, standpoint. It properly concerns the agent himself only as one agent among others, all of whom are subject to rational norms (p.225).

There are two (related) points here with which we might agree, and for the following reasons. Were the judgement to be grounded on a feature of the agent (e.g., a capacity for sympathy) that is not of necessity shared by all rational agents, then the putative rational principle does not, of necessity, apply to all rational agents. It is not, then, a (fundamental) rational principle and the judgement that it is is an incorrect one. It follows that for the agent (or the ISIS) to judge of P that it is a rational principle he must make the judgement qua rational agent. In short, it follows from the universality of rational principles that the judgement of a principle as a rational one is made qua rational agent. Secondly, in judging P to be a rational principle one judges it to be a principle that applies indifferently to all; one judges it to be impartially applicable.

We can concede, then, that the ISIS's judgement is made from the impersonal standpoint and that the content of the judgement applies, and is seen by him to apply, impartially to all rational agents. Darwall continues:

This means that insofar as an ISIS is motivated by his judgement that P is a principle on which any agent ought rationally to act, he is motivated from an impersonal standpoint. Any motivation he has, as an ISIS, for preferring that he act on P will be impersonal, therefore (p.225).

Now we can allow that insofar as the ISIS is motivated by his judgement the motivation will be impersonal, if what is meant by this is that he will be motivated qua rational agent. That is, we can allow that he, qua rational agent, will be motivated to prefer that he act on P. Darwall's reading is however the quite different one: that the ISIS "...will be motivated to prefer that he, qua rational agent, act on P" (p.225). And from this it follows that "...he will have the very same motivation to prefer that any agent act on P, indeed, that all do so" (p.226).

The difference is, of course, crucial. Whereas we can allow, I claim, abstractive steps of the second kind, steps which abstract from differences among preferrers, Darwall is pushing an abstractive step of the first kind, a step which deletes reference to the preferrer in (the statement of) the preference.26

In defense of the first reading one can say that the ISIS's judgement (that P is a principle on which any agent ought rationally to act) is the judgement that acting on P is rationally justified, and that it is plausible to see that judgement (made qua rational agent) as motivating the ISIS (qua rational agent) from the personal standpoint, where "personal", here, opposes "impartial". After all, that the ISIS sees himself as rational "...figure[s] prominently in his own self-esteem" (p.214). So, supposing P to be the principle of individual utility (as Darwall does), one can say that the ISIS will prefer qua rational agent that he act on P and allow that he will see not only his preference but any

other agent's like preference as justified, and yet deny that the ISIS will (or must) prefer that all agents act on P.

Darwall considers this objection and part of his response is:

But if what justifies and motivates an ISIS's preference that he act on P is (his judgement) that P is a principle on which all agents ought rationally to act, how can that judgement, taken by itself, motivate a preference simply for his acting on P? It could do so only if the judgement were essentially a judgement about his own conduct, made from his own standpoint. Since it is not, and since the judgement itself motivates the ISIS's preference that he act on P, it motivates equally his preferring that all agents do so (p.227).

11. This is a difficult passage, open, I think, to a number of construals. Obviously it asserts what the objector has denied; that the preference motivated by an impersonal judgement (a judgement made qua rational agent about all rational agents) is, or must be, an impartial one. And one might see this assertion as stemming from a failure to distinguish the impartial from the impersonal standpoint; that is, one might see the argument to that assertion as resting on an ambiguity in (the use of) the term "impersonal", as involving a covert shift from "impersonal" to "impartial" as I've defined these terms.

Or one might suspect that Darwall's picture of the ideally rational agent is that of someone who is motivated by his judgement to a preference that he act on P (act rationally) simply because, or for the reason that, (as the agent sees it) he is an instance of rational agency, and falling out from this (for Darwall), via the principle of sufficient reason, he is motivated to a preference that all agents act on P (act rationally). That is,

the picture of the ISIS might be that of someone whose concern to act rationally is concern that that rational person, simpliciter, act rationally. And if this is Darwall's picture of ideal agency, one can then say that impartial preference is built into it and that the question is begged.27

Or, lastly, one might construe the passage as containing, not a fallacious argument, but an appeal to an inchoate special version of judgement internalism. I will deal a little later on with this reading.

12. The second part of his response to the objection is to say that what the objector may have in mind "...is a sort of rational egoist.." who is "...an egoist first and a rationalist second" (pp.227/228); an egoist who, to be consistent, accepts the rationality of others acting as he does. But this reply is unconvincing. One could retort that what Darwall has in mind is a sort of rational moralist who is a moralist first and a rationalist second.

The retort is not superficial. There is, prima facie, nothing that forbids a picture of the ISIS as one who is egoistic because egoism is rational, any more than there is to see him as one who is moral because moral behaviour is rationally recommended. Contra Darwall's seeming presumption, this (egoistic) ISIS of rational norms can be "...concerned first and fundamentally to act rationally" (p.228).

These are of course very different pictures. This ISIS will not prefer that all agents share his rational preferences; indeed the rational norms may dictate that at times he try to ensure that they do not. His self-identification as subject of the RNS will not, then, issue in that kind of respect for rational norms that would be characteristic of an ISIS of the RNS, were the norms of the RNS in fact moral ones. But it may be

peculiar to moral appraisal (and, possibly to a lesser extent, to appraisals of etiquette) that in thinking that people ought one also wants them to. To insist that the ideal agent's respect for rational norms must be of this moral variety (let us call it that) is to legislate impartiality into the choice of rational principles.

This is Hill's worry. He fears that "...the ISIS has a built-in Kantian disposition [to impartiality]" and that this Kantian bias infects all subsequent argumentation.28 Consequently, he has strong reservations concerning the guidance principle.

And Darwall's text does indeed lend itself to such a reading as Hill's. The guidance principle, for example, though it has a plausible ring to it -- it can be seen as an accessibility condition of the kind put forward by Nagel,29 or as an expression of the (internalist) dictum that ought implies can -- is immediately likened to "...Kant's idea that the concept of moral duty "contains that of a good will" " (p.221). Indeed we are encouraged to see the ISIS as the rational analogue to Kant's good will (p.221). This early stage of the argument is hardly an appropriate place to draw that analogy.

But apart from this there seems to be good reason to see impartial preference as merely stipulated via the definition of the ISIS. The ISIS is defined as one who is motivated to act on rational norms because they are rational norms. In turn, by the guidance principle, a principle is a rational norm only if the ISIS can accept it, only if it can motivate the ISIS. Now clearly Darwall means to add something by introducing this principle but it is difficult to ascertain what it might be. From the definition of an ISIS we know that a principle's being a rational principle is sufficient for its being acceptable to an ISIS and from the guidance principle we know that being acceptable to an ISIS is a necessary condition of a principle's being a rational principle. If, as it appears on the face of it, the guidance principle (being of the form "p only if q") is merely the obverse of

the definition of an ISIS (being of the form "if p then q") then the guidance principle will not contain the seed of impartial preference if the definition of the ISIS doesn't.

13. We have been following what I have dubbed "the explicit argument" in Impartial Reason. I hope to have shown that, on either of two plausible interpretations of it, the argument is fallacious; that it either rests on an equivocation in the term "impersonal" or (as Hill suspects) it begs the question. Though we might allow that through it Darwall has succeeded in imposing some layers of the veil -- those that make preference or choice impersonal -- we can deny that, on its basis, he has succeeded in imposing that critical layer which guarantees impartiality. Moreover, the argument's limited success can be attributed to its appeal to a prohibited view, viz., externalism.

14. To talk of an explicit argument is, of course, to suggest that there is something more going on, something beneath the surface, as it were. And I think this is the case. Recall that it is Darwall's aim that he vindicate Kant's view -- that principles of practical reason are such that they "...would be the object of choice of rational agents from a perspective impartial between them.." -- but that he do so without recourse to Kant's "...difficult metaphysical doctrine of noumenal freedom" (p.200). Now that doctrine, I assume, was meant to demonstrate the connection between practical reason and autonomy; was concerned with showing that practical laws are such that (ideally) self-governing agents would choose to be guided by them, that such self-guidance is the (only) expression of full autonomy, and that the practical laws are moral ones. I assume, that is, that the burden of this doctrine was to vindicate moral principles by establishing a relation between reason and freedom.30 And my aim, thus far, has been to show that

the wish to circumvent this "difficult doctrine" has resulted in an argument which is not available to an internalist and which, on two construals of it, doesn't work anyway.

Nonetheless, what informs the argument is that very doctrine which Darwall has sought to avoid. Darwall admits as much in his Reply to Terzis. He writes:

What lies behind [Impartial Reason] is a Kantian "internalism of Autonomy": an internalism that derives from thinking that rational conduct is autonomous conduct, and that whatever principles constrain rational agents as such must be fit to be critically endorsed by them in a way that enables guidance by such principles to be self-government. What [Impartial Reason] takes to be fundamental is .... a picture of the fully rational agent as one who is autonomous because he is guided by his own critical endorsement of principles as normative for rational agents.31

Darwall's salient argument in Impartial Reason is, then, informed by a doctrine which he has consciously sought to avoid but which wells up, as it were, into the conscious plan. The ideally rational agent as subject of (and to) a rational system, for example, is a transmogrification of the ideal rational agent as autonomous and, I hope to have shown, an unhappy one. The result of all this is that to read Impartial Reason with a stress on the explicit argument is to guide one's boat by the prevailing wind while ignoring the prevailing current. Having discovered that, let us now drift with the current; let us follow the more genuine internalist argument.

15. It is worth remarking that in his response to Terzis Darwall fails to note that in openly avowing an internalism of autonomy he is embracing Kant's view and not defending it. In Impartial Reason he speaks of Kant's argument from the universal normativity of rational principles to the categorical imperative as "... never been entirely

clear to students of his thought" (p.200). Presumably the same holds for the argument to rationality as autonomy. If this is so then Darwall's avowal may have little weight for non-Kantians. This is a disadvantage, indeed that very one that in Impartial Reason he sought to avoid.32 On the other hand the approach seems to have the clear advantage of furnishing a quick and internalist route to the impersonalist element in the original position.33 Let us see how it might do this.

Practical principles are the issue for each agent of his autonomous choice. Choice is not autonomous if it is dictated by predilections an agent (fully-fleshed) just happens to have, marks, that is, of his personal history or his extra-rational nature. These, then, are abstracted from prior to making the choice. But it is such predilections which mark him out from other agents. In choosing in abstraction from them he is choosing qua rational agent, and qua rational agent he differs not at all (except numerically) from any other agent qua rational agent. His choice is impersonal, therefore.

Furthermore, in that there is qualitatively but one chooser, each chooses what any other would choose and practical principles are indeed universal. Autonomous choice -- choice qua rational agent -- fits hand-in-glove with the universality of rational principles, with there being but one rational normative system. So we have here an argument in the right direction; from one rational system to the normativity of reasons rather than vice versa (the latter, I have claimed, being an externalist move). In short, an internalism of autonomy captures the normativity of reasons.

Now to equate rationality with autonomy is to transfer the "ought" implicit in the first term to the second term. In that one ought to be rational one ought to be autonomous.34 Again, in that autonomy is self-guidance by principles endorsed by the

agent from the standpoint of the rational agent as such, we can say that an agent (any agent) ought (rationally) to adopt that standpoint in assessing putative practical principles. We can say, that is, that reflection from this standpoint is ideal rational reflection, reflection which any rational agent ought to engage in.

16. What I have given in the last three paragraphs is not so much a paraphrase of some argument of Darwall's in Impartial Reason as it is an articulation of the internalist element35 in the externalist/internalist argumentation there, and then only a rough one perhaps, and only inasmuch as I find it persuasive. Let us link it now to what Darwall says. He holds that

> ...because the requisite normative ideal of reflective guidance is universal, connected to a [procedural] principle that is normative for all rational agents as such, IR [the theory of reasons in Impartial Reason] holds that candidates become normative in this way by virtue of being practically endorsable from a standpoint that is impartial between rational agents. Therefore, for IR, the question of what substantive rational norms there are, and thus what reasons there are for agents to act, depends ultimately on what principles can be practically endorsed from this standpoint.36

Now there is a sense in which Darwall is right here. The impersonal standpoint just is that standpoint that is impartial between rational agents as choosers. To this extent Darwall has argued to the original position; to the propriety of the parties behind the veil being equally situated. And the principles chosen from that standpoint are, too, impartially applicable; they are universal principles, principles that any agent ought rationally to follow. But what has not been shown is impartiality in the choice; that in choosing principles an agent takes no account of numerical difference (the only

difference between himself and any other qua rational agent). That is, it has not been shown that the choice is made by each agent from a standpoint of equal concern for all; from the impartial standpoint.

But this is what Darwall holds -- that the choice is impartial in just that way. He argues that since the agent, to be fully autonomous, must endorse principles not from his own point of view but from a standpoint that is common to all rational agents, and since what he endorses is a universal principle -- "...any rational agent's guiding his conduct in this way" --, then he must have "...impartial motivation to prefer that any person reflectively guide conduct in this way".37

As it is important to see what mistake, or what appeal, Darwall is making here, perusal of another passage (from another essay) making the same point, will prove useful.

Since ultimate principles apply to all if, and only if, they apply to any, the relevant standpoint [of choice or endorsement] will be one that is impartial between rational persons as such. And the relevant question will be: What principles are choiceworthy to guide the conduct of all when a choice is made from a standpoint that is impartial between them?38

We have already agreed with the first part of this if "impartial between" is taken as "situated equally"; if what is being described is the impersonal standpoint. And we can agree with the second if what is meant by "principles.. choiceworthy to guide the conduct of all" is "principles such that all ought guide their conduct by them", or "principles such that each and every agent (qua rational agent) would choose that he be guided by them". But only if it is given that sense, only if it is that statement, does it follow from what has gone before. But by "principles..choiceworthy to guide the conduct

of all" Darwall means "principles such that each agent would choose that all agents follow them" or, put another way, "principles such that, if all are going to follow them, each agent would choose". That is, he wants it to follow that rational principles are such that each agent qua rational agent would prefer that all agents follow them. And this doesn't follow.

Again we may see the argument here, like the argument in Impartial Reason, as either resting on an equivocation in (the use of) the term "impersonal" or as resting on a picture of the ideal agent (this time the agent as autonomous) that builds in impartial preference and so begs the question. As with the argument in Impartial Reason there is, however, a third construal we might make. Let us turn to that.

17. In an early stage of Impartial Reason a distinction is made between judgement internalism and existence internalism. The former is a view about the relation of motivation to (genuine) judgements while the latter is a view about the relation of motivation to (the existence of) reasons.39 Though nowhere expressly stated, Darwall is, I think, an advocate of both forms. On his analysis, to judge something to be a reason is (in part at least) to recognize in oneself the appropriate attitude; that is, preference is internal to the judgement. He writes:

...in the process of rational consideration itself, we judge whether something is a reason by how we are motivated, by our attitude. The presence of the appropriate attitude appears to be itself part of what we judge to be the case when, within rational deliberation, we judge something to be a reason, or reason enough, for us to act (p.128).

And the same is true of the ideal agent's judgement of a principle as a rational principle:

...critical endorsement [of principles] must itself involve motivation; otherwise, guidance by critically endorsed principles will depend on further motivation, external to the rational self.40

Now an agent can judge something as being a reason where part of what he judges is his preference that he act on that something. And, in that a reason need not be universal, there need not arise any question as to his preference with regard to the conduct of others.

The case with respect to principles is, however, more complex. Rational principles are universal, so the ideal agent, in judging a principle to be a rational principle, is judging for all and not just for himself and, presumably, he knows this. Since part of what he judges in judging something to be a rational principle is his preference, can we say that, since the principle holds for all, what he is motivated to is a preference that all act on it? This is what (on the third construal) Darwall wants to say.

Now the moral person, when faced with alternative courses of action, might well ask himself Kant's question: Is this the way I'd want everyone to behave? This is arguably the paradigmatic moral question. What Darwall wants to say, on the grounds that a rational principle is binding on all agents, is that, in attempting to ascertain whether or not a principle is a rational one, the rational person must (or the ideally rational agent will) ask himself precisely the same question.

Though not clearly false, this claim is not plausible. Certainly it does not express the "rationality" underlying a rational egoism but this, in itself, is not surprising. Let us take a more neutral example. While watching a game of chess (say) I might think that a player at such and such a stage ought to X but my judgement here does not imply that I want him to X, nor indeed that I want everyone in that situation to X. Rather it seems to

imply only that were i that player I would (want to) X.41 And this even though -- supposing X to be the correct move; that not to X ensures defeat by a skilled and careful opponent -- the motivation to X (to act in accord with the principle "If in situation P, then X") is available from the standpoint of one (rational) chess-player among others.

One can offer many such examples where, though one might well want someone to act as he rationally ought, one's judgement that he ought does not involve nor imply one's wanting him to.

So, on this (most favourable) construal of the argument, Darwall appeals to a form of internalism that is at least controversial. And since it plays a pivotal role in his attempt to demonstrate the rationality of moral conduct, the burden is his to provide a more compelling case for it. As it stands he has incorporated, groundlessly, too much Kant into his internalism of autonomy.

Without an appeal to this "impartial internalism", Darwall has taken us (I hope to have shown) merely to the impersonal standpoint, the standpoint of the rational agent as such, and not to the "intersubjective" standpoint, the impartial standpoint of the rational agent as such. He has failed, that is, to take us to the original position.

Footnotes to Chapter 9.

1. And, of course, a consideration that is not subjective is objective (p.118). Note again that consideration from one's own particular point of view does not automatically

preclude altruism, e.g., that a person other than oneself flourish is a subjective consideration and may be a (subjective) reason.

2. By "commonly" I mean that moral considerations are overriding (it is to be shown) for individuals in historically familiar conditions; that is, for individuals who, as we do and human beings are very likely to continue to do, face a Hobbes situation. Darwall (pp.176/177) follows James Wallace's characterization of a Hobbes situation as one where

1. Everyone is apt to benefit if all or most people in the situation conform to B, but this benefit is not realized unless most conform.

2. Conforming to B generally involves some sacrifice, so that, other things being equal, it is apt to be maximally advantageous for an individual not to conform when most conform.

3. As a rule, the sacrifice involved for an individual in conforming to B is small in comparison with the benefits to him of the conformity of most people to B.

4. It is apt to be maximally disadvantageous to an individual to conform to B when not enough others conform to realize the benefit.

For Darwall, though, there is an important rider on moral considerations being overriding in such a situation and that is that what it is rational to do (and presumably what it is moral, or at least not immoral, to do) depends on "...what others will in fact do" (p.244). Rational practical principles will reflect what he calls "...the fixed conditions of the situation" (p.243). This is the additional Hobbesian reflection of which I spoke in the introductory chapter. The (as I see it) embarrassing complication to which it gives rise will be discussed later.

3. More correct to say, I think, that at this stage of the argument it is an open question. I shall have more to say on Darwall's use of "personal" and "impersonal" for the contrasting standpoints.

4. To show that some objective considerations have a much greater weight than the initial account suggests occasions a lengthy excursion into what Darwall calls "intersubjective value". We shall not follow him in this as his argument does not require his showing the existence, for humans, of such value. Indeed Darwall himself later invokes Kant's objection to any "materialist" theory -- that it fails to guarantee universal validity -- and for this reason he jettisons intersubjective value as a basis for the rational justification of action that accords with moral requirements. See his p.174.

He is loth however to turn his back completely on Humean sympathy and the teleological ethic which this (human) capacity generates. Intersubjective value, he says, turns out to be a complicated fact about the world (p.147). Later he suggests that the "...teleological and deontological aspects of our conception of ethics, having their different sources in different aspects of us, is most adequate to the complicated truth" (p.163).

Nonetheless Kant's objection stands and "...the theory of practical reason is not properly conceived as fundamentally teleological" (p.199). Rather he sees his own argument as being based on "...a theory of practical reason that is fundamentally deontological" (p.199).

The right prior to the good? Well, not quite. What is fundamental is neither right nor good but an ideal of rational character, the agent as self-critically autonomous.

Darwall is much more clear on this in the later essays. See Darwall 1986, esp., pp.41/42, 49, 52; and Darwall, 1987, pp.153-157.

5. See Falk, op.cit., esp., p.509.

6. There is a sense nonetheless in which they are discovered, if only derivatively. Reasons are those considerations (or facts) which move the agent because action on them accords with rational principles. In his review of Nagel (1986) Darwall characterizes rational principles as those principles an autonomous agent can critically ratify, where critical ratification "...is, as Nagel says, a "motivational discovery". It is a discovery of what can intelligibly be objectively willed" (Darwall, 1987, p.154).

Ideal rational agency as autonomy, as self-governance by critically chosen principles, is what underlies (by his own admission) Darwall's thinking in Impartial Reason, but this "Kantian internalism of Autonomy", as he calls it (Darwall, 1988, p.119), does not receive the stress here that it does in the later essays. I am following now the argument as explicitly laid out here, an argument which, I hope to show, is a somewhat different argument and one which, I shall suggest, an internalist may not advance.

The underlying argument -- the internalist one, what we may call the argument from autonomy -- will be considered later.

7. See his Chapter 5. As well as its difficulties with respect to an explanation of how we are moved by reasons and its inviting "metaphysical mystery", the externalist doctrine poses an epistemological mystery. On this view there are not only nonnatural properties but "...a special sort of intuition or insight to discern them" (p.55).

8. Frankena, op.cit., p.80.

9. Falk, op.cit., p.509.

10. There appears to be a silent assumption in Darwall of the equivalence of ideally rational agent, rational agent as such, and individual qua rational agent (at least when the latter two are making impartial judgements). At any rate he appears to move easily between the alternative terms. The appearance, if not misleading, may have to do with the Kantian assumption that each agent has it in him to be an ideally rational agent.

There is, for Darwall a qualitative equivalence between each agent qua rational agent, but this is worked towards, by a process of abstraction, as we shall see on considering his move to the impersonal standpoint. Nonetheless the same Kantian assumption appears to underlie it, in that what is seen to be left after abstraction is the full pure or ideal agent as against, possibly, a paltry highest common factor among rational agents as such.

I do not quarrel with that assumption but I do make use of the embarrassment to Darwall's theory posed by counter-claims made by Terzis and Postow which, in effect, deny that rationality is homogeneous across rational agents or indeed across ideally rational agents.

11. This is Gauthier's self-support requirement. Though too weak of itself, it contributes (Darwall argues) to a more discriminating requirement, viz., the criterion of universal impartial self-support. I will say more on this later.

12. Darwall, 1986, p.52. Compare Rawls' analogous claim that there are no moral facts prior to construction. In Rawls, 1980; see esp. p.564.

13. Darwall, 1988, p.118.

14. because it depends on the false claim that to say that one rationally ought to act as one has reason to act is to say something that is more than merely trivial. Postow challenges the view that principles are "more basic" than reasons. Postow, op.cit.(footnote 6, Chapter 7), p.136.

15. Postow, op.cit. See esp. pp.135-139.

16. George Terzis, op.cit.(footnote 6, Chapter 7), p.104.

17. Darwall, 1988, p.116.

18. Two agents may, according to Postow, see both the other's choice of principle and her own choice as "maximally well-grounded" (Postow, op.cit., p.138). But maximally well-grounded in virtue of what? Postow acknowledges that the charge of "brute givenness or brute chosenness" can be brought against either agent's adopted standard and he suggests that this charge might be answered by showing the standard to be "...part of a normative theory that is in wide reflective equilibrium with relevant background theories as well as with considered rational judgements or intuitions" (ibid., p.139).

Admittedly, if one allows that this could be done for different (and mutually exclusive) standards then one allows for a relativism with respect to rationality, but prima facie there is nothing about the notion of wide reflective equilibrium that encourages belief in the existence of several equilibria rather than one.

Darwall holds that on Postow's view "...one is left with relativism at the most ultimate level, viz., regarding rationality". See Postow, op. cit., p.137. As I've suggested, this is the appropriate response to Terzis also.

19. See p.200 for Darwall's characterization of Kant's doctrine.

Darwall feels he avoids this doctrine by an argument which "...proceeds simply by taking seriously the idea that principles of practical reason are principles that are normative for us simply because we are rational" (p.200). My point has been that this "taking seriously" involves a covert externalism.

20. Falk says of our ordinary moral thinking:

The external and internal uses of "ought" remain undifferentiated, and are imperceptibly juxtaposed and confused. There may be an unnoticed switch in the use of "ought" from the one to the other,...an alternation between a nebulous externalist and an internalist interpretation of one and the same thing (Falk, op.cit., p.509).

I see precisely the same juxtaposition and confusion -- though this time with respect to the rational "ought" -- in this segment of Darwall's argument. The guidance principle provides the internalist strain -- the vindication of a principle hanging on the ISIS's being capable of being moved by it. Here the direction is from motivation to

justification. The succeeding passages contain the externalist element where the direction is reversed.

21. Darwall does not employ the guidance principle at this point. Rather the conclusion of the argument quoted here is first combined with that other aspect of ideal rationality, self-criticalness, to yield this requirement on an ISIS's regarding a principle as a rational one: that it "...be rational according to it for her to choose, from an impartial perspective, that all agents act on it" (p.229). This is the requirement of universal impartial self-support and, from the guidance principle, a principle's being a rational principle is conditional on its fulfilling this requirement.

Though much emphasized in Impartial Reason it is not altogether clear how this requirement figures in the argument there. Darwall does use it to reject the principle of informed individual utility maximization, this principle, not surprisingly, failing to satisfy UIS. On the other hand, a principle advising (moral) constraint in the pursuit of informed individual preference in a Hobbes situation is seen to support itself in the required way. See p.233. This leads Darwall (who identifies the initial account with the former principle) to say that "...reasons grounded in our initial account of rational consideration cannot be the only reasons to act" (p.233).

Nonetheless there appears to be some redundancy in the argumentation that includes and follows the introduction of this requirement. In the face of objections to UIS raised by Terzis, Darwall indicates that his theory of reasons "...can be presented without reference to any sort of self-support condition.." (Darwall, 1988, p.124). I think that with respect to the explicit argument in Impartial Reason he is right about this. As we shall see when we examine the way in which the initial and extended accounts are linked, the choice of rational principles (made by the agent qua rational agent and impartially for all to act on) is an instrumental one, the initial account serving to provide the goal or end in view of which the choice is made. So, in addition to the constraint that it be made from behind the veil, the choice is subject to the requirements of means-end (or relative) rationality. This double set of constraints serves, I think, to adjudicate between putative rational principles in much the same way as the criterion of UIS is intended to do.

Darwall is attracted to the notion of self-support because in expressing the self-critical activity of the ideal agent it is tied to the germinative notion in his thought, a Kantian internalism of autonomy (See Darwall, 1988, p.124). My point is that self-support is redundant in Darwall's explicit argument because that argument, in its attempt to avoid Kant's difficult (or obscure) metaphysics, ceases to be a fully internalist one.

22. In that the ISIS qua rational agent is (qualitatively) identical to any agent qua rational agent (See footnote 10, here) the reference to the ISIS can be deleted to yield this condensed and more general formulation of the requirement: that a principle be such that one (anyone), qua rational agent and impartially, would prefer that all agents act on it; that is, that it be such that it would be the object of (anyone's) choice from behind the thicker veil.

23. In the essay "Subjective and Objective", Nagel, 1979, p.206.

24. Nor, though it is present, is this difference accentuated in Rawls' description of the veil of ignorance. I am going on that description in Section 24, pp.136-142, A Theory of Justice, 1971.

That the distinction is not emphasized is possibly due, in Nagel's case, to the catholic scope of the inquiry, in Rawls', to the specifically moral nature of the enterprise, but it may also spring from the fact that, though conceptually the two kinds of abstraction are clearly different, particular abstractive steps may incorporate both elements or may be such that it is not clear whether they are instances of the one kind or the other. I am not saying that either author was unaware of the distinction nor, indeed, that Darwall was (or is); quite the contrary. Rather I am saying that the distinction is not marked in some key passages in Nagel's and Rawls' (influential) works and that, on one (legitimate) construal of his argument to the original position, Darwall is neglectful of it.

The imposition by the veil of the several layers of ignorance is equivalent to a series of abstractive steps. It is, in fact, by virtue of this equivalence that the extended account -- which advances the view of ideal rational consideration as a process of abstraction -- can be seen as providing a rational foundation for the original position. Darwall and Rawls are evidently knitting to the same pattern.

25. There are two senses of "personal", then; one where it opposes "impartial", the other where it opposes "impersonal". This allows for four broad categories of preference or standpoint. Speaking in terms of preference we can have:

a personal/impartial preference, e.g., the preference that all abstain from red meat.

an impersonal/personal preference, e.g., the preference that (under certain conditions) one survive. This is (arguably) a preference that everyone has.

an impersonal/impartial preference, e.g., the preference that each has (supposing an ideally moral population) that each (indiscriminately) prosper.

a personal/personal preference, e.g., the preference that one complete a marathon.

26. viz. the step to impartiality, the remainder of the passage (on p.226) being devoted to demonstrating the impersonality of the preference.

Successful employment of both kinds of step would yield preference from the "intersubjective" standpoint; would yield the proposition that he (the ISIS), qua rational agent, will be motivated to prefer that he, qua rational agent, act on P.

27. Or, one might think, the question is not begged but the picture of the ideal agent involves or invokes a metaphysical assertion. I get occasional hints of what I can only term a quasi-religious adherence on the part of the ISIS to rationality for its own sake; hints of a kind of Moorean realism of the value of rationality as such. Perhaps perversely? Darwall does reject Moorean intrinsic value (p.146); it is the paradigmatic case study of the problems of externalism. Nonetheless, if I have been right, Darwall's approach via the ISIS is more than a little tainted with externalism.

In Rational Agent, Rational Act (pp.52/53) Darwall, avoiding the idea of the rational agent as subject (ISIS), puts forward a character-based approach to rationality that requires but a "thin" theory of the good. My criticism of the initial account is relevant here.

28. Hill, op.cit., p.618. The objection prompted by the second construal of the passage quoted is quite similar to that raised by Hill. He treats it at greater length. See his pp.617-619. The second part of Darwall's response supports our taking this construal.

29. He writes: I do not believe that the truth about how we should live could extend radically beyond any capacity we might have to discover it... Nagel, 1986, p.139.

30. See Impartial Reason, pp.200, 246.

31. Darwall, 1988, pp.119/120.

32. Darwall might say that I have got things wrong here; that his stress, in Impartial Reason, on the self-critical aspect of ideal rationality is a stress on autonomy. But in Impartial Reason the sole employment of self-criticalness is to generate (in tandem with internal self-identification) the criterion of universal impartial self-support and, as Darwall rightly claims, his theory of reasons "...can be presented without reference to any sort of self-support condition.." (Darwall, 1986, p.124).

33. Darwall sees it as yielding the original position tout `a fait. I will come to that.

34. Prima facie, "that one ought to be rational" is analytic. It may be -- I am not sure -- that that is how Darwall sees it. Speaking of the ISIS's desire to act on rational norms he says that, like Butler's "principle of reflection", it "...implicitly claims its own authority" (p.225). See also his response to Parfit's attack on the self-support condition (Darwall, 1986, p.124).

35. the element more clearly delineated in the later essays.

36. Darwall, 1988, p.120.

37. Ibid., p.120.

38. Darwall, 1986, p.52.

39. See p.54.

40. Darwall, 1988, p.120. But, in that something's being a rational principle is dependent on the ideal agent's judging it so, we can say too that preference is internal to (the existence of) rational principles, and, derivatively, is internal to (the existence of) reasons. It is at this stage -- at the level of ideal rational consideration -- that, on Darwall's internalist analysis, motivation and justification fuse.

41. Gibbard's analysis starts along these lines. Alan Gibbard, "A Noncognitivist Analysis of Rationality in Action", Social Theory and Practice, Volume 9, 1983. See esp. pp.207 et seq.

CHAPTER 10

THE FULL THESIS

1. We have yet to consider the way in which, to form Darwall's full thesis, the extended and initial accounts of rational reflection are connected. But, before we do this, we should look at a feature of the choice (from the original position) of rational principles which marks out Darwall's schema from that of Rawls' and which, I believe, marks his conception of rational conduct out from that of Kant.

2. Like Rawls' contracting parties, who choose principles of justice for persons with a sense of justice, Darwall's arbitrary rational agent (agent qua rational agent) chooses principles for beings who are "...rational in the sense that they are willing, and able, to constrain their pursuit of informed preference by principles that would be chosen" (p.231, footnote 7). And on the assumption that moral principles of some kind would be chosen1 it seems that it would be rational for an agent to constrain his conduct accordingly even in transactions with persons who, he knows, are (irrationally) egoistic. As Darwall puts it, his theory seems to have "...the untoward result that it is rational to sacrifice oneself to others who have no intention of cooperating in joint action for mutual benefit.."(p.243). But, he claims, his theory does not have this implication because the principles chosen -- rational principles -- will take such circumstances, such "fixed conditions of the situation", into account.2

While no doubt reassuring to our onetime rational egoist (now pondering conversion), one of whose fears is precisely that a commitment to rational (moral)

principles would, on occasion, cast him in the role of moral dupe,3 this qualification excites one to wonder whether the latter's fearful characterization of moral agency is nearer to Kant's own conception than is Darwall's.

For how like Kavka's rule-egoism Darwall's rational/moral conception now is! Kavka, who "...assumes that a kind of egoism underlies practical rationality", puts forward what he calls the Copper Rule (in contradistinction to Kant's Golden one), a rule which calls on the agent to constrain his selfish pursuits where others are constraining theirs, but not otherwise.4 And this is how Darwall's view is rendered by a Hobbesian reflection additional to that employed by Rawls, a reflection which imparts that hard-headedness of approach demanded, Darwall might feel, in debate with the moral sceptic. But, needed as it might be in such a debate, it removes from moral agency the heroism characteristic of Kant's conception of it.

Related to this diminution in moral conception, and perhaps more important, is the (arguable) proposition that the conditional nature of moral requirements threatens the autonomy of the rational agent. For on this account of rational guidance the agent's co[illegible]t is dictated, at least in part, by the conduct of others. Now an agent's adapting his behaviour to the "fixed conditions" of the situation makes good sense, one might suppose, on almost all theories of rational action, but, exceptionally, it does not on Kant's. Because, for Kant, the requirement and the possibility of moral/rational conduct stands independently of the agent's contingent circumstances, whether external or internal.

So, not only does this extra reflection force a different interpretation of the Kantian equation of the moral with the rational with the autonomic,5 it suggests too that

the latter part of the equation does not hold. For, in that his conduct and the rationality of his conduct rests on what he knows others will in fact do, the agent is heteronomous.

Darwall might say that the agent's adapting his conduct to that of others is justified by its being the necessary means to his legitimate goal of preserving the conditions (possession of primary goods) for his continuing rational agency. But it is not clear that this response will do. For, it seems, in the very attempt to preserve his rational agency he loses it.

As noted earlier, Darwall suspects that Falk would reject his theory. The grounds for that rejection, I suggest, lie here, in Darwall's additional Hobbesian reflection.

3. Let us turn now to the full thesis. I have already remarked that Darwall sees the extended account of rational consideration as qualifying the initial account; as endorsing some, but not all, of the considerations that move an agent from the personal standpoint. At the same time he holds that there are "...good independent grounds for supposing our initial account to be at least a partial account of practical reason" (p.234). He must hold this because, while the extended account is intended to provide the position from which the choice of principles is made (viz., the original position with a thicker veil than that in Rawls' description of it), the initial account is meant to provide the goal in view of which the choice is made. And that goal is constituted by the "highest order interest" of the rational person in exercising his capacity to revise his final ends etc.6 As Darwall remarks elsewhere, the choice of principles must be guided by a "thin" theory of the good, by the interests intrinsic to rational personality.7

I have argued, in Chapter 8, that Darwall has not vindicated Rawls' conception, that he has not shown that an individual does have this capacity; that indeed his own internalism suggests that an individual has a set of one or more uncriticizable preferences. I have nothing to add here to those arguments. There remains, though, an interesting question: if those arguments are sound, could Darwall yet hold to his position? That is, could he allow that indeed there is a basic uncriticizable preference and yet hold that it is a preference that carries its justification with it; it being the preference to preserve one's interests qua rational agent, or the preference to exercise one's rational agency?

This is a moot point. If what is being made is a factual claim it would seem to depend on a particular view of the human individual and his faculties, e.g., Kant's metaphysical view. For if merely an empirical hypothesis there is not only the problem of showing evidence to support it but it is, ex hypothesi, subject to disconfirmation. And that doesn't seem good enough. So if a factual claim it is one of a peculiar non-empirical sort, i.e., a metaphysical claim.

If, on the other hand, it is a normative claim, if it is the claim that the (potentially) rational agent ought to have this preference as fundamental, there are other complications. For example, what can an individual do (on a normative internalism) if this isn't his fundamental preference? Or, if a normative claim, how to defend it? For even if Darwall had argued successfully to the original position, the agent choosing principles qua rational agent and impartially for all to act on, the end in view of which he (instrumentally) chooses rational principles is uncriticizable by reasons deriving from the principles expressly adopted to promote it. Argument as to the justifiability of that end is confined, then, to an initial-account arena. This being the case, Rawls' highest order

interest has one obvious competitor -- that provided by a self-interest theory of rationality, viz., the interest that one's life go for one as well as possible.8

It might be argued that if it is the individual qua rational agent who adjudicates these rival claims then he will opt for the interests intrinsic to rational personality. But the agent qua rational agent is choosing for the fully-fleshed individual, for one whom he knows has preferences, and though he doesn't know what these preferences are, he knows (I assume) that some of them are for other than that of securing his interests qua rational agent. So, absent an appeal to a metaphysical view, it is not at all clear that the pre-eminent goal in view of which the choice is made ought to be that of preserving his highest order interest `a la Rawls.

4. In the last few paragraphs I have dealt hastily with a tangle of problems. But my aim has been not so much to set them out in detail (let alone suggest solutions) as to point out that they are there and that the initial account does not resolve them. One consequence of this is that even if the choice is made from the original position it may issue in principles of the good rather than principles of the right.

There is another, more serious, consequence. Central to Darwall's theory of reasons is the position that "...the normative character of reasons is inherited from a normative ideal of something internal to the rational agent: an ideal of rational reflection".9 On this view, ideal reflection, reflection from the original position, issues in practical principles and, derivatively, in reasons. On the other hand, as Darwall points out, the idea behind the hypothetical imperative is "...that the force of reasons is, as it were, transferred back and forth along the line connecting an end and its necessary means.." (pp.47/48). This means, amongst other things, that the justification of the

means is derived from that of the end it serves. And this brings an instability into the full thesis. For the normativity of reasons -- their authority, as conferred by ideal rational reflection -- is diminished if there is uncertainty about the fundamental end in view of which rational principles (and, derivatively, reasons) are chosen; and there is, I have tried to show, such uncertainty.

5. With this we have come full circle; we have come back to Sidgwick's worry of an unresolvable tension between the dictates of morality and those of self-interest. In light of this so far unresolved tension, the dialogue between the moral apologist and the moral sceptic must continue.

Footnotes to Chapter 10.

1. See pp.241/242.

2. See pp.243/244. See also, Darwall, 1988, p.123.

3. See my section 1, Chapter 3.

4. See Cooper, op.cit., pp.492/493. The words quoted are his. Of the Copper Rule he says "...[it] is exemplified in Kavka's treatment of Hobbes's laws of nature as having a two-part logical structure, "Do X, provided others are doing so as well," in which the main clause requires behaviour of a traditionally moral kind, but the qualifying clause releases the agent from that requirement if others are not satisfying it, leaving him free to promote his own interests without moral restriction".

5. See my Section 5, Chapter 7.

6. See pp.102 and 234. See Rawls, 1980, p.525.
7. Darwall, 1986, p.53.
8. See Parfit, op.cit., pp.3/4.
9. Darwall, 1988, p.118.

# BIBLIOGRAPHY


Cooper, Wesley. 1989. "Critical Notice", Canadian Journal of Philosophy, Volume 19, No. 3, September 1989, pp.491-508.

Darwall, Stephen L. 1983. Impartial Reason. Cornell University Press.

------- 1986. "Rational Agent, Rational Act", Philosophical Topics, Volume X1V, No. 2, Fall 1986, pp. 33-57.

------- 1987. "How Nowhere Can You Get (and Do Ethics)?, Ethics, Volume 98, (1987-88), pp. 137-157.

------- 1988. "Reply to Terzis", Canadian Journal of Philosophy, Volume 18, Number 1, March 1988, pp. 115- 124.

Falk, W. D. 1947/48. ""Ought" and Motivation". In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., first edition, 1952. New York: Appleton-Century-Crofts.

Field, G.C. 1921. "A Criticism of Kant". In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970. New York: Appleton-Century-Crofts.

Frankena, William K. 1958. "Obligation and Motivation in Recent Moral Philosophy". In Essays in Moral Philosophy, A.I. Melden, ed., 1958. Seattle: University of Washington Press.

Gewirth, Alan. 1972. "Moral Rationality". In Freedom and Morality (The Lindley Lectures), John Bricke, ed., 1976, University of Kansas.

Gibbard, Allan. 1983. "A Noncognitivist Analysis of Rationality in Action", Social Theory and Practice, Volume 9, 1983, pp.199-221.

Hare, R. M. 1976. "Ethical Theory and Utilitarianism". In Contemporary British Philosophy, 4, H.D. Lewis ed., 1976. London: Allen and Unwin.

Hill, Jr, Thomas E. 1986. "Darwall on Practical Reason", Ethics, 96, 1986.

Hospers, John. 1961. "Why Be Moral?". In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970. New York: Appleton-Century-Crofts.

Kavka, Gregory S. 1986. Hobbesian Moral and Political Theory. Princeton University Press.

Mackie, J.L. 1977. Ethics. Harmondsworth: Penguin Books.

Mele, Alfred R. 1989. "Motivational Internalism: The Powers and Limits of Practical Reasoning", Philosophia (Israel), 19(4), December 1989, pp.417-436.

Nagel, Thomas. 1979. Mortal Questions. New York: Cambridge University Press.

------- 1986. The View From Nowhere. New York: Oxford University Press.

Nielsen, Kai. 1963. "Why Should I Be Moral?". In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970. New York: Appleton-Century- Crofts.

Parfit, Derek. 1984. Reasons and Persons. Oxford University Press.

Postow, B.C. 1986. "Darwall and the Impartial Standpoint", Philosophical Studies, Volume 49, 1986.

Prichard, H.A. 1928. "Duty and Interest". In Readings in Ethical Theory, Wilfrid Sellars and John Hospers, eds., second edition, 1970. New York: Appleton-Century-Crofts.

Rawls, John, 1971. A Theory of Justice. Cambridge, Mass.: Harvard University Press.

------- 1980. "Kantian Constructivism in Moral Theory", The Journal of Philosophy, Volume LXXV11, No.9, September 1980.

Smart, J.J.C. 1973. Utilitarianism: For and Against. J.J.C. Smart and B. Williams, Cambridge University Press.

Terzis, George. 1988. "Darwall's Kantian Argument", Canadian Journal of Philosophy, Volume 18, Number 1, March 1988, pp.99-114.