out how under-powered the supposed rational choice under ignorance is.

Rawls' theory tries, in effect, to link politics with morality, and morality (or at least the relevant parts of it) to a self-interested choice under uncertainty. He indeed links politics with a Kantian conception of morality, but the supposed choice under uncertainty seems in fact to have the morality already packed into it, and as an exercise in decision theory, or anything like it, compares unfavourably even with Pascal's celebratedly bad bet.

8 Internal and external reasons

Sentences of the forms '*A* has a reason to ϕ' or 'There is a reason for *A* to ϕ' (where 'ϕ' stands in for some verb of action) seem on the face of it to have two different sorts of interpretation. On the first, the truth of the sentence implies, very roughly, that *A* has some motive which will be served or furthered by his ϕ-ing, and if this turns out not to be so the sentence is false: there is a condition relating to the agent's aims, and if this is not satisfied it is not true to say, on this interpretation, that he has a reason to ϕ. On the second interpretation, there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate motive. I shall call the first the 'internal', the second the 'external', interpretation. (Given two such interpretations, and the two forms of sentence quoted, it is reasonable to suppose that the first sentence more naturally collects the internal interpretation, and the second the external, but it would be wrong to suggest that either form of words admits only one of the interpretations.)

I shall also for convenience refer sometimes to 'internal reasons' and 'external reasons', as I do in the title, but this is to be taken only as a convenience. It is a matter for investigation whether there are two sorts of reasons for action, as opposed to two sorts of statements about people's reasons for action. Indeed, as we shall eventually see, even the interpretation in one of the cases is problematical.

I shall consider first the internal interpretation, and how far it can be taken. I shall then consider, more sceptically, what might be involved in an external interpretation. I shall end with some very brief remarks connecting all this with the issue of public goods and free-riders.

The simplest model for the internal interpretation would be this: *A* has a reason to ϕ iff *A* has some desire the satisfaction of which will be served by his ϕ-ing. Alternatively, we might say . . . some desire, the satisfaction of which *A* believes will be served by his ϕ-ing; this

difference will concern us later. Such a model is sometimes ascribed to Hume, but since in fact Hume's own views are more complex than this, we might call it *the sub-Humean model*. The sub-Humean model is certainly too simple. My aim will be, by addition and revision, to work it up into something more adequate. In the course of trying to do this, I shall assemble four propositions which seem to me to be true of internal reason statements.

Basically, and by definition, any model for the internal interpretation must display a relativity of the reason statement to the agent's *subjective motivational set*, which I shall call the agent's S. The contents of S we shall come to, but we can say:

(i) An internal reason statement is falsified by the absence of some appropriate element from S.

The simplest sub-Humean model claims that any element in S gives rise to an internal reason. But there are grounds for denying this, not because of regrettable, imprudent, or deviant elements in S – they raise different sorts of issues – but because of elements in S based on false belief.

The agent believes that this stuff is gin, when it is in fact petrol. He wants a gin and tonic. Has he reason, or a reason, to mix this stuff with tonic and drink it? There are two ways here (as suggested already by the two alternatives for formulating the sub-Humean model). On the one hand, it is just very odd to say that he has a reason to drink this stuff, and natural to say that he has no reason to drink it, although he thinks that he has. On the other hand, if he does drink it, we not only have an explanation of his doing so (a reason why he did it), but we have such an explanation which is of the reason-for-action form. This explanatory dimension is very important, and we shall come back to it more than once. If there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action (it does not follow that they must figure in all correct explanations of their action). The difference between false and true beliefs on the agent's part cannot alter the *form* of the explanation which will be appropriate to his action. This consideration might move us to ignore the intuition which we noticed before, and lead us just to legislate that in the case of the agent who wants gin, he has a reason to drink this stuff which is petrol.

I do not think, however, that we should do this. It looks in the wrong direction, by implying in effect that the internal reason conception is only concerned with explanation, and not at all with the agent's

rationality, and this may help to motivate a search for other sorts of reason which are connected with his rationality. But the internal reasons conception is concerned with the agent's rationality. What we can correctly ascribe to him in a third-personal internal reason statement is also what he can ascribe to himself as a result of deliberation, as we shall see. So I think that we should rather say:

(ii) A member of S, D, will not give A a reason for ϕ-ing if either the existence of D is dependent on false belief, or A's belief in the relevance of ϕ-ing to the satisfaction of D is false.

(This double formulation can be illustrated from the gin/petrol case: D can be taken in the first way as the desire to drink what is in this bottle, and in the second way as the desire to drink gin.) It will, all the same, be true that if he does ϕ in these circumstances, there was not only a reason why he ϕ-ed, but also that that displays him as, relative to his false belief, acting rationally.

We can note the epistemic consequence:

(iii) (a) A may falsely believe an internal reason statement about himself, and (we can add)

(b) A may not know some true internal reason statement about himself.

(b) comes from two different sources. One is that A may be ignorant of some fact such that if he did know it he would, in virtue of some element in S, be disposed to ϕ: we can say that he has a reason to ϕ, though he does not know it. For it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate; otherwise one merely says that A would have a reason to ϕ if he knew the fact. I shall not pursue the question of the conditions for saying the one thing or the other, but it must be closely connected with the question of when the ignorance forms part of the explanation of what A actually does.

The second source of (iii) is that A may be ignorant of some element in S. But we should notice that an unknown element in S, D, will provide a reason for A to ϕ only if ϕ-ing is rationally related to D; that is to say, roughly, a project to ϕ could be the answer to a deliberative question formed in part by D. If D is unknown to A because it is in the unconscious, it may well not satisfy this condition, although of course it may provide the reason why he ϕ's, that is, may explain or help to explain his ϕ-ing. In such cases, the ϕ-ing may be related to D only symbolically.

I have already said that

(iv) internal reason statements can be discovered in deliberative reasoning.

It is worth remarking the point, already implicit, that an internal reason statement does not apply only to that action which is the uniquely preferred result of the deliberation. '*A* has reason to ϕ' does not mean 'the action which *A* has overall, all-in, reason to do is ϕ-ing'. He can have reason to do a lot of things which he has other and stronger reasons not to do.

The sub-Humean model supposes that ϕ-ing has to be related to some element in *S* as causal means to end (unless, perhaps, it is straightforwardly the carrying out of a desire which is itself that element in *S*). But this is only one case: indeed, the mere discovery that some course of action is the causal means to an end is not in itself a piece of practical reasoning.[1] A clear example of practical reasoning is that leading to the conclusion that one has reason to ϕ because ϕ-ing would be the most convenient, economical, pleasant etc. way of satisfying some element in *S*, and this of course is controlled by other elements in *S*, if not necessarily in a very clear or determinate way. But there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in *S* can be combined, e.g. by time-ordering; where there is some irresoluble conflict among the elements of *S*, considering which one attaches most weight to (which, importantly, does not imply that there is some one commodity of which they provide varying amounts); or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment.

As a result of such processes an agent can come to see that he has reason to do something which he did not see he had reason to do at all. In this way, the deliberative process can add new actions for which there are internal reasons, just as it can also add new internal reasons for given actions. The deliberative process can also subtract elements from *S*. Reflection may lead the agent to see that some belief is false, and hence to realise that he has in fact no reason to do something he thought he had reason to do. More subtly, he may think he has reason to promote some development because he has not exercised his

[1] A point made by Aurel Kolnai: see his 'Deliberation is of Ends', in *Ethics, Value and Reality* (London and Indianapolis, 1978). See also David Wiggins, 'Deliberation and Practical Reason', *PAS*, LXXVI (1975–6); reprinted in part in *Practical Reasoning*, ed. J. Raz (Oxford, 1978).

imagination enough about what it would be like if it came about. In his unaided deliberative reason, or encouraged by the persuasions of others, he may come to have some more concrete sense of what would be involved, and lose his desire for it, just as, positively, the imagination can create new possibilities and new desires. (These are important possibilities for politics as well as for individual action.)

We should not, then, think of *S* as statically given. The processes of deliberation can have all sorts of effect on *S*, and this is a fact which a theory of internal reasons should be very happy to accommodate. So also it should be more liberal than some theorists have been about the possible elements in *S*. I have discussed *S* primarily in terms of desires, and this term can be used, formally, for all elements in *S*. But this terminology may make one forget that *S* can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent. Above all, there is of course no supposition that the desires or projects of an agent have to be egoistic; he will, one hopes, have non-egoistic projects of various kinds, and these equally can provide internal reasons for action.

There is a further question, however, about the contents of *S*: whether it should be taken, consistently with the general idea of internal reasons, as containing *needs*. It is certainly quite natural to say that *A* has a reason to pursue *X*, just on the ground that he needs *X*, but will this naturally follow in a theory of internal reasons? There is a special problem about this only if it is possible for the agent to be unmotivated to pursue what he needs. I shall not try to discuss here the nature of needs, but I take it that insofar as there are determinately recognisable needs, there can be an agent who lacks any interest in getting what he indeed needs. I take it, further, that that lack of interest can remain after deliberation, and, also that it would be wrong to say that such a lack of interest must always rest on false belief. (Insofar as it does rest on false belief, then we can accommodate it under (ii), in the way already discussed.)

If an agent really is uninterested in pursuing what he needs; and this is not the product of false belief; and he could not reach any such motive from motives he has by the kind of deliberative processes we have discussed; then I think we do have to say that in the internal sense he indeed has no reason to pursue these things. In saying this, however, we have to bear in mind how strong these assumptions are, and how seldom we are likely to think that we know them to be true. When

we say that a person has reason to take medicine which he needs, although he consistently and persuasively denies any interest in preserving his health, we may well still be speaking in the internal sense, with the thought that really at some level he *must* want to be well.

However, if we become clear that we have no such thought, and persist in saying that the person has this reason, then we must be speaking in another sense, and this is the external sense. People do say things that ask to be taken in the external interpretation. In James' story of Owen Wingrave, from which Britten made an opera, Owen's family urge on him the necessity and importance of his joining the army, since all his male ancestors were soldiers, and family pride requires him to do the same. Owen Wingrave has no motivation to join the army at all, and all his desires lead in another direction: he hates everything about military life and what it means. His family might have expressed themselves by saying that *there was a reason for Owen to join the army*. Knowing that there was nothing in Owen's S which would lead, through deliberative reasoning, to his doing this would not make them withdraw the claim or admit that they made it under a misapprehension. They mean it in an external sense. What is that sense?

A preliminary point is that this is not the same question as that of the status of a supposed categorical imperative, in the Kantian sense of an 'ought' which applies to an agent independently of what the agent happens to want: or rather, it is not undoubtedly the same question. First, a categorical imperative has often been taken, as by Kant, to be necessarily an imperative of morality, but external reason statements do not necessarily relate to morality. Second, it remains an obscure issue what the relation is between 'there is a reason for A to...' and 'A ought to...' Some philosophers take them to be equivalent, and under that view the question of external reasons of course comes much closer to the question of a categorical imperative. However, I shall not make any assumption about such an equivalence, and shall not further discuss 'ought'.[2]

In considering what an external reason statement might mean, we have to remember again the dimension of possible explanation, a consideration which applies to any reason for action. If something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action. Now no external reason statement could *by itself* offer an explanation of anyone's action. Even if it were true (whatever that might turn out to mean) that there was a reason for Owen to join the army, that fact by itself would never explain anything that Owen did, not even his joining the army. For if it was true at all, it was true when Owen was not motivated to join the army. The whole point of external reason statements is that they can be true independently of the agent's motivations. But nothing can explain an agent's (intentional) actions except something that motivates him so to act. So something else is needed besides the truth of the external reason statement to explain action, some psychological link; and that psychological link would seem to be belief. A's believing an external reason statement about himself may help to explain his action.

[2] It is discussed in chapter 9, below.

External reason statements have been introduced merely in the general form 'there is a reason for A to...', but we now need to go beyond that form, to specific statements of reasons. No doubt there are some cases of an agent's ϕ-ing because he believes that there is a reason for him to ϕ, while he does not have any belief about what that reason is. They would be cases of his relying on some authority whom he trusts, or, again, of his recalling that he did know of some reason for his ϕ-ing, but his not being able to remember what it was. In these respects, reasons for action are like reasons for belief. But, as with reasons for belief, they are evidently secondary cases. The basic case must be that in which A ϕ's, not because he believes only that there is some reason or other for him to ϕ, but because he believes of some determinate consideration that it constitutes a reason for him to ϕ. Thus Owen Wingrave might come to join the army because (now) he believes that it is a reason for him to do so that his family has a tradition of military honour.

Does believing that a particular consideration is a reason to act in a particular way provide, or indeed constitute, a motivation to act? If it does not, then we are no further on. Let us grant that it does – this claim indeed seems plausible, so long at least as the connexion between such beliefs and the disposition to act is not tightened to that unnecessary degree which excludes *akrasia*. The claim is in fact *so* plausible, that this agent, with this belief, appears to be one about whom, now, an *internal* reason statement could truly be made: he is one with an appropriate motivation in his S. A man who does believe that considerations of family honour constitute reasons for action is a man with a certain disposition to action, and also dispositions of approval, sentiment, emotional reaction, and so forth.

Now it does not follow from this that there is nothing in external

reason statements. What does follow is that their content is not going to be revealed by considering merely the state of one who believes such a statement, nor how that state explains action, for that state is merely the state with regard to which an internal reason statement could truly be made. Rather, the content of the external type of statement will have to be revealed by considering what it is to *come to believe* such a statement – it is there, if at all, that their peculiarity will have to emerge.

We will take the case (we have implicitly been doing so already) in which an external reason statement is made about someone who, like Owen Wingrave, is not already motivated in the required way, and so is someone about whom an internal statement could not also be truly made. (Since the difference between external and internal statements turns on the implications accepted by the speaker, external statements can of course be made about agents who are already motivated; but that is not the interesting case.) The agent does not presently believe the external statement. If he comes to believe it, he will be motivated to act; so coming to believe it must, essentially, involve acquiring a new motivation. How can that be?

This is closely related to an old question, of how 'reason can give rise to a motivation', a question which has famously received from Hume a negative answer. But in that form, the question is itself unclear, and is unclearly related to the argument – for of course reason, that is to say, rational processes, can give rise to new motivations, as we have seen in the account of deliberation. Moreover, the traditional way of putting the issue also (I shall suggest) picks up an onus of proof about what is to count as a 'purely rational process' which not only should it not pick up, but which properly belongs with the critic who wants to oppose Hume's general conclusion and to make a lot out of external reason statements – someone I shall call 'the external reasons theorist'.

The basic point lies in recognising that the external reasons theorist must conceive *in a special way* the connexion between acquiring a motivation and coming to believe the reason statement. For of course there are various means by which the agent could come to have the motivation and also to believe the reason statement, but which are the wrong kind of means to interest the external reasons theorist. Owen might be so persuaded by his family's moving rhetoric that he acquired both the motivation and the belief. But this excludes an element which the external reasons theorist essentially wants, that the agent should

acquire the motivation *because* he comes to believe the reason statement, and that he should do the latter, moreover, because, in some way, he is considering the matter aright. If the theorist is to hold on to these conditions, he will, I think, have to make the condition under which the agent appropriately comes to have the motivation something like this, that he should deliberate correctly; and the external reasons statement itself will have to be taken as roughly equivalent to, or at least as entailing, the claim that if the agent rationally deliberated, then, whatever motivations he originally had, he would come to be motivated to ϕ.

But if this is correct, there does indeed seem great force in Hume's basic point, and it is very plausible to suppose that all external reason statements are false. For, *ex hypothesi*, there is no motivation for the agent to deliberate *from*, to reach this new motivation. Given the agent's earlier existing motivations, and this new motivation, what has to hold for external reason statements to be true, on this line of interpretation, is that the new motivation could be in some way rationally arrived at, granted the earlier motivations. Yet at the same time it must not bear to the earlier motivations the kind of rational relation which we considered in the earlier discussion of deliberation – for in that case an internal reason statement would have been true in the first place. I see no reason to suppose that these conditions could possibly be met.

It might be said that the force of an external reason statement can be explained in the following way. Such a statement implies that a rational agent would be motivated to act appropriately, and it can carry this implication, because a rational agent is precisely one who has a general disposition in his S to do what (he believes) there is reason for him to do. So when he comes to believe that there is reason for him to ϕ, he is motivated to ϕ, even though, before, he neither had a motive to ϕ, nor any motive related to ϕ-ing in one of the ways considered in the account of deliberation.

But this reply merely puts off the problem. It reapplies the desire and belief model (roughly speaking) of explanation to the actions in question, but using a desire and a belief the content of which are in question. *What* is it that one comes to believe when he comes to believe that there is reason for him to ϕ, if it is not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? We were asking how any true proposition could have that content; it cannot help, in answering that,

to appeal to a supposed desire which is activated by a belief which has that very content.

These arguments about what it is to accept an external reason statement involve some idea of what is possible under the account of deliberation already given, and what is excluded by that account. But here it may be objected that the account of deliberation is very vague, and has for instance allowed the use of the imagination to extend or restrict the contents of the agent's S. But if that is so, then it is unclear what the limits are to what an agent might arrive at by rational deliberation from his existing S.

It *is* unclear, and I regard it as a basically desirable feature of a theory of practical reasoning that it should preserve and account for that unclarity. There is an essential indeterminacy in what can be counted a rational deliberative process. Practical reasoning is a heuristic process, and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion. To someone who thinks that reasons for action are basically to be understood in terms of the internal reasons model, this is not a difficulty. There is indeed a vagueness about 'A has reason to ϕ', in the internal sense, insofar as the deliberative processes which could lead from A's present S to his being motivated to ϕ may be more or less ambitiously conceived. But this is no embarrassment to those who take as basic the internal conception of reasons for action. It merely shows that there is a wider range of states, and a less determinate one, than one might have supposed, which can be counted as A's having a reason to ϕ.

It is the external reasons theorist who faces a problem at this point. There are of course many things that a speaker may say to one who is not disposed to ϕ when the speaker thinks that he should be, as that he is inconsiderate, or cruel, or selfish, or imprudent; or that things, and he, would be a lot nicer if he were so motivated. Any of these can be sensible things to say. But one who makes a great deal out of putting the criticism in the form of an external reason statement seems concerned to say that what is particularly wrong with the agent is that he is *irrational*. It is this theorist who particularly needs to make this charge precise: in particular, because he wants any rational agent, as such, to acknowledge the requirement to do the thing in question.

Owen Wingrave's family may not have expressed themselves in terms of 'reasons', but, as we imagined, they could have used the external reasons formulation. This fact itself provides some difficulty for the external reasons theorist. This theorist, who sees the truth of an external reason statement as potentially grounding a charge of irrationality against the agent who ignores it, might well want to say that if the Wingraves put their complaints against Owen in this form, they would very probably be claiming something which, in this particular case, was false. What the theorist would have a harder time showing would be that the words used by the Wingraves *meant* something different from what they mean when they are, as he supposes, truly uttered. But what they mean when uttered by the Wingraves is almost certainly *not* that rational deliberation would get Owen to be motivated to join the army – which is (very roughly) the meaning or implication we have found for them, if they are to bear the kind of weight such theorists wish to give them.

The sort of considerations offered here strongly suggest to me that external reason statements, when definitely isolated as such, are false, or incoherent, or really something else misleadingly expressed. It is in fact harder to isolate them in people's speech than the introduction of them at the beginning of this chapter suggested. Those who use these words often seem, rather, to be entertaining an optimistic internal reason claim, but sometimes the statement is indeed offered as standing definitely outside the agent's S and what he might derive from it in rational deliberation, and then there is, I suggest, a great unclarity about what is meant. Sometimes it is little more than that things would be better if the agent so acted. But the formulation in terms of reasons does have an effect, particularly in its suggestion that the agent is being irrational, and this suggestion, once the basis of an internal reason claim has been clearly laid aside, is bluff. If this is so, the only real claims about reasons for action will be internal claims.

A problem which has been thought to lie very close to the present subject is that of public goods and free riders, which concerns the situation (very roughly) in which each person has egoistic reason to want a certain good provided, but at the same time each has egoistic reason not to take part in providing it. I shall not attempt any discussion of this problem, but it may be helpful, simply in order to make clear my own view of reasons for action and to bring out contrasts with some other views, if I end by setting out a list of questions which bear on the problem, together with the answers that would be given to them by one who thinks (to put it cursorily) that the only rationality of action is the rationality of internal reasons.

1. Can we define notions of rationality which are not purely egoistic?
   Yes.
2. Can we define notions of rationality which are not purely means–end?
   Yes.
3. Can we define a notion of rationality where the action rational for A is in no way relative to A's existing motivations?
   No.
4. Can we show that a person who only has egoistic motivations is irrational in not pursuing non-egoistic ends?
   Not necessarily, though we may be able to in special cases. (The trouble with the egoistic person is not characteristically irrationality.)

Let there be some good, G, and a set of persons, P, such that each member of P has egoistic reason to want G provided, but delivering G requires action C, which involves costs, by each of some proper sub-set of P; and let A be a member of P: then

5. Has A egoistic reason to do C if he is reasonably sure either that too few members of P will do C for G to be provided, or that enough other members of P will do C, so that G will be provided?
   No.
6. Are there any circumstances of this kind in which A can have egoistic reason to do C?
   Yes, in those cases in which reaching the critical number of those doing C is sensitive to his doing C, or he has reason to think this.
7. Are there any motivations which would make it rational for A to do C, even though not in the situation just referred to?
   Yes, if he is not purely egoistic: many. For instance, there are expressive motivations – appropriate e.g. in the celebrated voting case.[3] There are also motivations which derive from the sense of fairness. This can precisely transcend the dilemma of 'either useless or unnecessary', by the form of argument 'somebody, but no reason to omit any particular body, so everybody'.
8. It is irrational for an agent to have such motivations?
   In any sense in which the question is intelligible, no.
9. Is it rational for society to bring people up with these sorts of motivations?
   Insofar as the question is intelligible, yes. And certainly we have reason to encourage people to have these dispositions – e.g. in virtue of possessing them ourselves.

I confess that I cannot see any other major questions which, at this level of generality, bear on these issues. All these questions have clear answers which are entirely compatible with a conception of practical rationality in terms of internal reasons for action, and are also, it seems to me, entirely reasonable answers.

[3] A well-known treatment is by M. Olson Jr. *The Logic of Collective Action* (Cambridge, Mass., 1965). On expressive motivations in this connexion, see S. I. Benn, 'Rationality and Political Behaviour', in S. I. Benn and G. W. Mortimore, eds., *Rationality and the Social Sciences* (London, 1976). On the point about fairness, which follows in the text, there is of course a very great deal more to be said: for instance, about how members of a group can, compatibly with fairness, converge on strategies more efficient than everyone's doing C (such as people taking turns).