STRUCTURES OF NORMATIVE THEORIES[*]

The Monist 76 (1993) 22-40

Introduction

Normative theorists like to divide normative theories into classes. One special point of focus has been to place utilitarianism into a larger class of theories which do not necessarily share its view about what is alone of impersonal intrinsic value, namely, individual human well-being, but do share another structural feature, roughly its demand that each person seek to maximize the realization of what is of impersonal intrinsic value. The larger class is distinguished from its complement in two apparently different ways. Let us look briefly at these two ways.

On the one hand, there is the distinction between normative theories which are agent neutral, and those which are agent centered. A theory is agent neutral if it gives to everyone the same advice or aims. According to an agent neutral theory, your (theory-given) aims are better fulfilled exactly when mine are. By contrast, an agent centered theory gives to us at least some prescriptions or advice or aims which include indexicals. Sometimes the things an agent centered theory advises me to do will conflict with the things the same theory advises you to do. Your theory-given ends can be achieved only at the expense of mine.

The classic and paradigmatic agent centered normative theory is egoism. Egoism tells me to pursue my good and you to pursue yours. The classic and perhaps paradigmatic agent neutral theory is utilitarianism. It tells each of us to promote the greatest happiness of the greatest number. Things are going better from your perspective exactly when they are going better from my perspective, according to utilitarianism.

On the other hand, there is the distinction between normative theories which evaluate everything, whether it be a policy, an action, a law, or a trait of character, according to the value of the consequences it produces, and normative theories which at least sometimes evaluate some things according to features other than their consequences. If two actions have the same consequences, then a consequentialist theory doesn't care at all which action is performed. A nonconsequentialist theory may distinguish between the rightness of two actions even if they would have exactly the same consequences.

Some theorists[1] conflate these distinctions into one. But this appears to be a mistake. There are uncontroversial examples of agent centered theories which evaluate actions solely in terms of their

consequences; hedonistic egoism is one of these. Whether there are any agent neutral theories that are not consequentialist is not so clear, because some recent work in moral philosophy suggests that *any* plausible moral theory can be recast as a consequentialist theory. John Broome argues, for instance, that a theory is, as he says, "teleological" if the instructions it gives us about how to act and evaluate can be characterized by a *better than* relation that meets certain formal constraints.[2] Agent centered theories can certainly meet these constraints. They merely specify an agent centered *better than* relation. Broome does not go so far as to say that any coherent normative theory must be characterizable in this way, but every remotely reasonable theory he considers does meet the constraints.

The main strategy for "consequentializing" any given moral theory is simple. We merely take the features of an action that the theory considers to be relevant, and build them into the consequences. For example, if a theory says that promises are not to be broken, then we restate this requirement: that a promise has been broken is a bad consequence. Notice that the weighting is not yet specified. If the theory under consideration includes an aboslute side constraint against promise-breaking, then we have the consequentialist version give a lexically prior negative weight to promise-breaking.[3]

This sort of consequentializing strategy is sometimes called "gimmicky," for example by Robert Nozick in *Anarchy, State, and Utopia.* But Vallentyne, Sen, and Broome argue compellingly that it is not gimmicky but in fact is the most natural way to taxonomize theories[4]. I believe that attention to the distinction between agent neutral and agent centered structures of moral theory is in any case far more fruitful than trying to draw a bright line between the class of strictly consequentialist theories (those whose specification of relevant kinds of consequence make no reference, explicit or implicit, to times other than the time of the state of affairs that constitutes the consequence) and its complement. Certainly Scheffler and Kagan, to name two prominent recent theorists in just this area, have in fact focused on the distinction between agent centered and agent neutral theories, even if they mistakenly *call* agent neutral theories "consequentialist." In this paper I want to address two questions about the agent neutral/agent centered distinction. First, is the conflation of that distinction with the consequentialist/nonconsequentialist distinction anything more than a harmless misnomer? Does it lead to substantive mistakes in moral theorizing? And second, what grounds do we have for choosing between the two kinds of structure?

Disarming Two Arguments

Conflating agent neutrality with consequentialism is not merely a misnomer. It can lend a specious cogency to some arguments against agent centered views, or against agent centered elements of common sense morality. I will consider two of these, and argue that their force dissipates when we make it explicit that the divide that concerns us is the one between agent centered and agent neutral theories.

A. The Stigma

A stigma attaches to the rejection of consequentialism, and pointing it out tends to drive us (or some of us) from an otherwise comfortable position within common sense morality. The process of stigmatization goes roughly as follows.

The consequentialist asks whether we don't, in fact, take the happiness of others to be a good thing. Of course, we do, how could one not admit that it is a good thing that others are happy? Then, the consequentialist asks, are there circumstances wherein one ought not to promote the (all considered) good, under which we ought to prefer less good to more? Common sense morality seems to say that there are. But how could this be? The rules that constrain us from doing more good need strong justification to overcome the paradoxical air of requiring us to do less good than we might.

It then appears that the burden cannot be discharged. If we try to explain the importance of the rules by showing how much good they will do, then the consequentialist has us in his net. If we assert blandly that the following of rules is of intrinsic importance, then the consequentialist doubts it but recommends that we place rule-following in our evaluation of consequences. We common sense moralists find that move unsatisfactory. We are not willing to twist a child's arm even if the information we could thereby get from his grandmother might help us to save two other children from arm-twisting[5]. We are then left in a very uncomfortable position. We can only repeat: "We *must* follow these rules." J. J. C. Smart[6] calls that attitude "rule worship," which of course is not an argument, but does seem to express the discomfort we feel at sticking at such a dogmatic spot. On the one side of deliberations is the good; we ourselves admit that it is good, and feel thus committed to favoring its advancement. On the other side are mere rules. Our rejection of consequentialism leaves an ugly stigma.

But this really is all wrong, I think; we are deceived, and deceived by the conflation of two kinds of distinction. The simple answer we may now give is that every moral view is consequentialist, that we common sense moralists as much as anyone are out to maximize the good. Of course, our understanding of the good may be an agent centered one, whereas the typical

challenger has an agent neutral understanding, but this contrast will have to be engaged by some other argument. We don't have to be embarassed by the charge that we are ignoring the good, because the charge is just false.[7]

B. Sidgwick's Stratagem

In his *Methods of Ethics*, Henry Sidgwick examines three kinds of metaethical theory. One is intuitionism, but for my purposes the relevant dialectic is between the other two, which Sidgwick calls varieties of hedonism. One is "Egoistic Hedonism," or just egoism: the view that it is right for me to pursue my own happiness. The other is "Universalistic Hedonism," or utilitarianism: the view that it is right for me to pursue that happiness of all. Sidgwick thinks that these two together form the basic kinds of theories which vie for our allegiance. He takes seriously the threat posed to his favorite, utilitarianism, by the radical egoistic alternative, and thinks it is crucial to be able to give reasons against egoism and for utilitarianism.

It is well known that Sidgwick wanted to derive utilitarianism. He writes, "by considering the relation of the integrant parts to the whole and to each other, I obtain the self-evident principle that the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other...."[8] Given how important it is for Sidgwick to be able to maintain that utilitarianism is the correct view, he devotes surprisingly little space to arguing against egoism. He agrees roughly with Mill, that no proof of utilitarianism can be given, strictly speaking, but that some considerations can be proffered that may tend to determine the intellect. He only has one real stratagem against the egoist, though, and it is only half a stratagem, for it will only work if he can get the egoist to come half way.

> If the Egoist strictly confines himself to stating his conviction that he ought to take his own happiness or pleasure as his ultimate end, there seems to be no opening for any line of reasoning to lead him to Universalistic Hedonism as a first principle; it cannot be proved that the difference between his own happiness and another's happiness is not *for him* all important.[9]

The admission is admirably frank. The reader might have thought that Sidgwick would use his device of the point of view of the Universe to try to claim that the Egoist occupies an incoherent position in moral space, that any view so obviously and radically partial as egoism simply does not

count as a moral theory. But Sidgwick is aware that such a move would be unsatisfying:

> It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently "I" am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not being taken as fundamental in determining the ultimate end of rational action for an individual.[10]

Yet, even his half-stratagem seems to Sidgwick a significant victory. The egoist's coherence can only be maintained so long as he sticks to talking about what he ought to do.

> When, however, the Egoist puts forward, implicitly or explicitly, the proposition that his happiness or pleasure is Good, ... it then becomes relevant to point out to him that *his* happiness cannot be a more important part of Good, taken universally, than the equal happiness of any other person. And thus, starting with his own principle, he may be brought to accept Universal happiness or pleasure as that which is absolutely and without qualification Good or Desirable: as an end, therefore, to which the action of a reasonable agent as such ought to be directed.[11]

Sidgwick thinks that once we get the egoist to claim not only that he ought to pursue his own happiness but that his happiness is somehow Good[12] we can drive him from an unstable position into utilitarianism. I don't think that Sidgwick makes clear exactly how this is to be achieved. One tempting interpretation just cannot be right. It cannot be that Sidgwick is saying that once the egoist agrees to have in his moral outlook only agent neutral reasons, we can force him to utilitarianism. That cannot be what Sidgwick is saying because it is simply not an argument. Sidgwick has already noted that the only difference between egoism and utilitarianism is that the one is agent centered, and the other is agent neutral. So his stratagem cannot be to try to get the egoist to admit that agent neutrality is an essential feature of a reasonable normative theory. The argument would be the trivial point that if we can get the egoist to give up the one thing that distinguishes his view from utilitarianism, then he will be a utilitarian. But this looks more like an explanation of what we should be trying to do than a way of doing it.

Why does Sidgwick think that the key step is getting the egoist to make assertions about what is Good instead of sticking to statements about what he ought to do? My suggestion is speculative and somewhat complicated. I need to take a detour through a certain notion of objectivity, one which I believe may play a role in Sidgwick's thinking.

1. Objectivity

Let's say that a value is objective (for me) when I hold it in a way that outreaches my personal perspective. Let me explain.

Some values I hold in a strong way: I (actually, now) value the thing even under the (possible, future) circumstance in which I (would, will) fail to hold it. One might call such a value, a value that outreaches its own existence.[13] Of course, whether a value outreaches its own existence depends upon who holds it, and in what way. Thus, you might value listening to Mozart in that you take the experience to be enriching and ennobling. In that case, you actually value listening to Mozart even under the counterfactual (or future) condition that you don't like it. The experience is enriching, you want to say, even for people who do not see that it is. On the other hand, I might value listening to Mozart only in that I find it a diverting pastime. Then, I don't actually value it for circumstances in which I would not value it, since it would not provide the pleasant experience that it actually does for me. So, your valuation of Mozart outreaches its own existence, while mine does not. There is nothing disreputable about holding either kind of value. It's simply that some sorts of valuings are tied more tightly to the fact that we actually hold the particular value than others are.[14]

There is a sense in which whenever I hold a value that outreaches my current, actual perspective, I am objectifying it. I seem to be placing the value in the object instead of in myself. Consider again the value attributed to listening to Mozart by one who takes Mozart to be enriching and ennobling. Suppose such a person says,

(V) Listening to Mozart is a valuable activity.

The claim seems to put the value in the activity itself. For the Mozart lover wants to maintain that (V) persists in its truth even when evaluated in circumstances in which the valuer herself changes in radical ways. It is as if the value were suffused throughout the activity, staying there as the valuer changes. But suppose instead that (V) is put forward by the Mozart enjoyer, the listener who values Mozart only as a pleasant diversion. In that case, (V) is not true when evaluated in circumstances in which the listener does not enjoy Mozart. The truth of (V), the value of the activity, does not survive changes in the way the valuer values. In this sense, it is not objectified but remains within the valuer.

2. Explaining Sidgwick

We want to understand why, according to Sidgwick, the egoist is safe when hugging the "ought" judgment, but unstable when striking out to the "Good" judgment. I think we can find

some guidance in the metaphor of objectivity. Thus, an "ought" judgment does not objectify. Saying "I ought to pursue my own happiness" keeps my reasons securely inside me. But saying "My own happiness is Good" does objectify the value. It seems to place the value of, in this case, the Egoist's happiness, outside himself and in the happiness. Objectification suggests that the value is public, that it ought to be appreciable by anyone.

But, if this is in fact what Sidgwick was thinking (and I must emphasize that my reconstruction of his thought is highly speculative), then he was eliding two different kinds of objectivity. I used the term "objective" in the first place because I believe the kind I defined is sometimes confused with another kind. A value is objective in my sense, if it outreaches its own existence. But in the sense that Sidgwick would need, "objective" must mean something very like "agent neutral." It must mean something like, "appreciable to anyone as a reason."

So even if the Egoist is willing to commit himself to valuing his own happiness in a way that outreaches the existence of the psychological state of valuing, he is not thereby committed to claiming that his happiness has a special value above and beyond that of others which everyone can appreciate from their own perspective. A good thing, too, for the egoist wants to maintain that the value of a person's happiness is appreciable from that person's perspective in a way that is not accessible to anyone else.

It is interesting to note that Kagan, too, appeals to a notion of objectivity.

> To say that from the moral standpoint one outcome is objectively better than another, is to say that *everyone* has a reason to choose the better outcome.... It is, in this sense, an *agent-neutral* reason: one that is equally generated for all moral agents, regardless of, e.g., their particular concerns or other interests. This agent-neutral reason is, of course the pro tanto reason to promote the good.[15]

As Kagan uses the expression, an objective value is the same as an agent neutral one. If we are tempted to say that calling something "good" constitutes attributing to it an objective value, then we will be able to execute the same move that Sidgwick tried to pull off. And, as I noted above, Kagan does equate having a consequentialist view with having an agent neutral one. I don't mean to say that Kagan is guilty of a fallacy of equivocation. He uses "good" and "consequentialism" and "objective" quite consistently throughout. But I do think that the combination of these expressions lends an intuitive credibility to the idea that a moral theory that evaluates actions by evaluating their consequences has to be agent neutral, and this idea in turn makes it difficult to understand how any moral view could be plausibly agent centered.

So unclarity about what consequentialism amounts to may lead us to endorse bad arguments against agent centered moral conceptions. When we level our attention at the contrast between agent centered values and agent neutral ones, we are less apt to be drawn into the agent neutral way of thinking. Of course, we may already have strong intuitive attachments to agent neutrality, and none of what I've said so far calls that attachment into question.

One might well wonder what sort of consideration might decide between agent neutrality and agent centeredness. Each type of theory can, I think, be made out to be quite coherent and to seem quite plausible. Sidgwick's pair of examples, egoism for an agent centered view and utilitarianism for an agent neutral one, provide a good case in point. They show that even once a theorist has settled on a conception of the good, there remains the question of whether that good is to be incorporated into the theory as agent centered or agent neutral. A similar point can be made about rights, I think. We generally think of rights as embodying an agent centered element of morality (or of the law). But, given an account of which rights there are, and what it is of importance that they protect or promote, we can still wonder whether they shouldn't be stirred into moral theory in an agent neutral way instead–as "goal rights", in Sen's terms[16], for example.[17]

The right way to address the choice between agent centered and agent neutral structures is not, I think, by trying to show that one or the other is incoherent or indefensible, but by ascending to a more abstract level of theory. It is likely that which kinds of normative theories can be justified or plausibly maintained depends on what kind of justification one employs. In the last part of this paper, then, I want to look at how three kinds of metaethical theories are related to agent neutrality and agent centeredness at the normative level.

Metaethical Justifications and Normative Structure

A. Ideal Observer Theories

Ideal Observer metaethics are sometimes called "impartial spectator" theories. Impartiality is the hallmark of agent neutral morality. So we might expect the Ideal Observer theorist to be a natural ally of agent neutrality. But in fact, there is no such alliance nor ought there to be.

The point is that although the Ideal Observer is supposed to be impartial, the impartiality is not of the right sort to guarantee an agent neutral judgment. Roughly, the impartiality of an observer might be of two types:

(a) That the observer cannot approve of partiality;

(b) That the observer can approve of partiality, but must do so impartially, so to speak.

While (a) does exert pressure toward agent neutrality, (b) does not. Roderick Firth, for one, clearly intended (b) as part of his characterization of the Ideal Observer. Firth says that he does not wish to rule out by his analysis "the moral theory (held by Ross and others) that the rightness or wrongness of an act is determined in part by irreducible obligations arising directly from certain personal relationships...."[18] Impartiality in the relevant sense is akin to universalizability[19].

It is not obvious that an Ideal Observer could not consistently judge favorably agent centered behavior. Ideal Observers need not be impartial in that respect–or better, they *do* have to be impartial, but even an impartial person can approve of partiality.[20]

Then the Ideal Observer metaethic will not *per se* decide between agent neutral and agent centered moral theories. Of course, any particular Ideal Observer theory will eventually have to come to grips with the agent neutral/agent centered contrast, if it is to offer any practical guidance in choosing a first order theory. But the agent neutrality and agent centeredness will be a point of dispute among Ideal Observer theorists just as it is among moral theorists in general.

B. Contract Theories

A likely candidate for generating agent centered moral reasons is contract theory. According to contractualists, moral rules are those that would be agreed to under certain conditions (with the conditions varying from version to version). Since contract theory generates moral reasons from the interests that people already have, it seems natural that they should tend to produce agent centered moral conceptions.

Suppose the question is as T. M. Scanlon puts it: What rules could not be reasonably rejected by parties concerned to reach agreement?[21] There are likely to be some areas of life for which no rule passes the test. Any rule in these spheres could be reasonably rejected by someone (that is, given any rule we could find someone who could reasonably reject it). If this is right, then a contractualist theory will include permissions, for permissions arise when requirements do not fill all of logical space.

But so far we don't know whether the permissions would be agent neutral or agent centered. For example, a theory might contain a general standing requirement to produce as much physical pleasure as possible, but also a permission to choose the less hedonic act if the more hedonic would reduce the pleasure of the least happy person. This theory is strict consequentialist. It is agent neutral. And it contains a permission. Would the permissions generated by a contractualist metaethic be agent neutral or agent centered?

Let's distinguish *Ideal* Contractualist theories from *Pragmatic* ones. By an Ideal theory I mean one which specifies idealized conditions as background for the contract. The most interesting for my purpose is a veil of ignorance *a la* Rawls. A Pragmatic theory does not idealize. It looks to agreement among agents as they in fact are, with the beliefs they actually have.

Shelly Kagan offers a line of reasoning to show that contract theories that use the device of the veil of ignorance will not produce agent centered elements.

> Each member of the agreement is rationally concerned to support those rules which have the greatest likelihood of promoting his interests. Given his ignorance of his actual position in society, however, he is unable to single out those rules that would most favor his actual position. The best he can do, then, is to support a system according to which the average person in society is to be made as well off as possible. More precisely, he will favor a system according to which the level to which each person has his interests satisfied is to be made, on average, as high as possible.[22]

Kagan seems to be right about Ideal Contractualism. Notice that it is not necessary to assume that the contracting parties are self-interested. No matter what interests they are supposed to have, the veil of ignorance erases the centeredness of their perspectives. When a normative theory is generated by agreement among parties with the same perspective, it must turn out to be an agent neutral theory. (Or so it seems–I will question this claim below.)

For more pragmatic versions of contractualism, though, the result is different. Let's suppose that each party to the moral contract has a strong primary concern with his or her own family. Assessing proposed rules, they look to see which will best protect and advance those interests. Suppose also that each believes in the same set of "Basic Interests": interests the meeting of which takes lexical priority in her values over other interests. Finally, suppose that each party also has the following empirical beliefs. (1) Each believes that without interference her family will be able to provide for its own Basic Interests. (2) Each believes that some other families will *not* be able to provide for their own Basic Interests, and that if she is required to assist those families, she may

have to sacrifice her own family's Basic Interests.

Here is one proposal that will look attractive. Its rules include a general agent neutral requirement to maximize the satisfaction of interests, but also, tempering that requirement, (a) a restriction against invading the Basic Interests of some to promote those of others; (b) a permission not to sacrifice the Basic Interests of one's own family, even to promote a greater number of Basic Interests of others.

Notice that if this proposal were assessed from behind a veil of ignorance, there would be an alternative which would look more attractive: the purely agent neutral scheme which assigned a lexically higher value to Basic Interests and a lexically lower value to other interests, and required everyone to promote these values in lexical order. It wouldn't make sense to hold out for the agent centered permission (b), for the reasons that Kagan gives. By opting for the agent centered permission, I favor myself *if* I turn out to be one of the more advantaged, but I hurt myself more, or more likely, if I turn out to be one of the less advantaged. But without the veil of ignorance, the indexical beliefs (1) and (2) take hold. I am not worried that the agent centered permission will harm my family's interests, since I believe that I will not be one of the least advantaged. If all the parties have such an indexical belief, they will contract for the first, agent centered scheme over the second, agent neutral one.

So far it looks like things fall out this way. Ideal Contractualism, with a veil of ignorance in place, produces agent neutral normative theories. Pragmatic Contractualism, without a veil, can at least under some assumptions produce agent centered ones. But let me now explore a way that even Ideal versions of contractualism may support moral theories with agent centered elements.

What is shown by Kagan's argument is that the aim of any given party in selecting among moral theories from behind a veil of ignorance should be to maximize the average satisfaction of whatever interests are taken to be at stake. But there is a subtle difference between this conclusion and the claim that the moral theory the parties should choose is an agent neutral one. Couldn't it turn out that the adoption of an agent centered theory would maximize average satisfaction of interests? Certainly, an agent neutral theory is more amenable to statement in terms of average satisfaction, but that's not the same as its *producing* more.

It is a favorite claim of critics of utilitarianism that accepting its excessive demands is incompatible with living a good, worthwhile, or meaningful life. Writing about Gandhi (who was

not, of course, a utilitarian), George Orwell said that "for the seeker after goodness there must be no close friendships and no exclusive loves."[23] "If one is to love God, or to love humanity as a whole," Orwell wrote, "one cannot give one's preference to any individual person.... To an ordinary human being, love means nothing if it does not mean loving some people more than others." Gandhi himself "on three occasions... was willing to let his wife or a child die rather than administer the animal food prescribed by the doctor.... This attitude is perhaps a noble one, but, in the sense which–I think–most people would give to the word, it is inhuman." Here the contrast is supposed to be between taking an unworldly perfection as one's ideal, and taking what Orwell calls a "humanistic" ideal, which he thinks *is* compatible with normal human love. But Orwell seems closer to the mark when he says that loving "humanity as a whole" is also incompatible with exclusive personal love. This insight is drawn out by Michael Stocker[24] and Susan Wolf[25] in arguing that devotion to an ideal of impersonal morality, and I would argue, to agent neutral moralities, robs us of many of the goods of ordinary life.

Surely this is too strong. Some people could live a rewarding and meaningful life fully in the service of others. Perhaps Gandhi himself is an example. Still, the thrust of the point must be granted. Not everyone, nor even most of us, could live very happy lives if every moment of our time were devoted in the most efficient utilitarian way. It appears, then, that if we accepted the requirement of maximizing average utility as the single element of our contractual morality, we would be committing ourselves to a very significant chance of a life of slavery to an ideal we could not find deeply satisfying.

I don't mean to be rehearsing Rawls' "strains of commitment" argument. The point is not that we would be unable to follow an excessively demanding moral theory, and would thus be led astray. Let's assume that we *could* follow it. The point is that from our position behind the veil of ignorance, we should not relish that outcome, for it would rob us of much of what we feel to be most important in our own lives.

The problem could be remedied by introducing permissions into the contracted scheme. They would clearly have to be agent relative permissions, for the only way to avoid the enslavement to the pursuit of the general happiness would be to allow me to pursue my projects and you to pursue yours. Following the moral psychologist, then, we could expect parties to reject a requirement that might prevent them from pursuing their most central commitments.

Of course, as Kagan remarks, by rejecting this requirement they take a certain risk. For they know they may find themselves in a position in which others could help them pursue their projects, but will not be required to do so. Behind a veil of ignorance, won't that possibility loom at least as large as the chance that a too-stringent requirement would stand in the way of personal projects? But the point is that accepting an excessively demanding moral scheme is in itself incompatible with having the kinds of commitments that are at stake. It is that being the sort of person who has an overriding commitment to promote an agent neutral good precludes the commitments that friendship and love, for example, require. It is this preclusion that it is reasonable to reject.

In summary, then, contractualism can motivate agent centered elements of a normative moral scheme in two ways. First, if it does not lower a veil of ignorance over the parties in the contractual situation, it can allow our natural agent centered concerns to find expression in the resulting morality. And second, even when a veil of ignorance is part of the machinery, it may turn out that agent centered permissions are a part of any moral scheme consistent with each party's need to form commitments and attachments which are in turn necessary elements in a good life.

C. Virtue Ethics

Let's turn finally to Virtue Ethics. For our purposes, Virtue theory claims that the proper way to conceive of right action is as that which would be done by a person with the right state of character. The virtues come first, in this way of thinking, and then we understand what is the right thing to do by appeal to them. By contrast, a utilitarian tries to locate the virtues by first working out a criterion of right action, and then investigating what sorts of personal characteristics best conduce to right actions (or perhaps more accurately, to good results).

What sort of moral structure will Virtue Ethics produce? We would be led astray if we thought that the main question is whether consequences are all that matters. Aristotle opens the Nicomachean Ethics with the commonsense notion that every action and choice aims at some end. He notes that in some cases the end is the action itself, while in others it is something outside of the action. And he thinks it is clear that the highest good, that for the sake of which everything else is done, is happiness, or living well. Now, if we thought that the central divide between kinds of moral theories were the distinction between consequentialist and nonconsequentialist views, we'd have to conclude that it is of primary importance to decide whether the correct account of happiness, for Aristotle, is an account of something that is strictly a consequence of our actions, or whether

happiness is something rather that may be a part of, or logically related to, the actions themselves.

But happily, we do not have to investigate that tangled question, not in the present context. For suppose that we decided that happiness consists at least in part of acting in a certain way. We would have to reject strict consequentialism. But the question would still remain: are we morally required always to act in such a way that the agent neutral good, namely, the happiness of human beings in general, is maximized? Or does morality give us special dispensation to favor our own happiness, or the happiness of those close to us? There is room, as we have seen, for an agent neutral theory which is not strictly consequentialist. Suppose, on the other hand, that we decided that happiness is a state of mind, which can be and always is strictly consequent upon human actions. We'd still be left with the question of whose happiness each of us ought to take to be the end of our own actions and choices.

So we will do better to look directly at the issue of agent centeredness. And virtue ethics turns out to be a friend, I think, of agent centered normative conceptions. The reason is that according to virtue theory, each of us is to be concerned in the first instance with our *own* virtue. My primary moral responsibility is to live my life according to virtue, and yours is to live your life according to virtue. I don't mean that according to virtue theory, a virtuous person will spend all of his time thinking about how to become more virtuous. Rather, the idea is that my virtue serves as an endpoint in the chain of justification that virtue theory offers. In Aristotelian terms, it is that for the sake of which things are done.

Suppose, for example, I ask the virtue theorist why I ought to send $25 to my alma mater. She says that it's because I promised to do so. Then I ask why I ought to keep my promise. She says that fidelity is a virtue. Let's suppose she convinces me that it is. I then ask why something's being a virtue gives me a reason to act according to it. She says that a virtuous life is my first responsibility, or my general goal. So the reasons for moral action are, according to virtue theory, grounded in a general aim of living as a virtuous person. It's *my* virtue that is of primary importance to me.

Another way to get at what I think is the same point draws on recent work of Gary Watson's.[26] Watson contrasts the ethics of virtue with perfectionism, which "enjoins us to promote the development and exercise of virtue, these being intrinsically good."[27] In its account of what is intrinsically good, perfectionism agrees with virtue theory. Both theories are, as Watson says, "areteic." Why does the perfectionist seem to have missed the point of virtue ethics? The

perfectionist agrees with the virtue theorist that a life lived in accordance with virtue is the primary bearer of value. But he then goes on to suggest that the conclusion is that we should all try to promote the living of virtuous lives in general. He makes the same move (or an analogous one) that Sidgwick makes in moving from Egoism to Utilitarianism. The mistake, according to the virtue theorist, is that the perfectionist thinks that because a virtuous life has value, it follows automatically that we all have moral reason to promote the living of virtuous lives in general. The conclusion may in the end turn out to be true, according to the virtue theorist, but only if it turns out that some virtue of beneficence is so strong as to control the actions of a virtuous person. The mistake is to think that we each have equal reason to be concerned with every person's virtue, whereas it is a hallmark of virtue theory that it assigns to each of us in the first instance a reason to be concerned with our own virtue.

This feature does not in itself guarantee an agent centered conception, for further investigation may show that you and I guard and promote our respective virtue by aiming at the same agent neutral goals. But room is left for agent centered elements. Consider Philippa Foot's remarks:

> [I]t seems clear that virtues are, in some general way, beneficial .... Nobody can get on well if he lacks courage, and does not have some measure of temperance and wisdom, while communities where justice and charity are lacking are apt to be wretched places to live.... But now we must ask to whom the benefit goes, whether to the man who has the virtue or rather to those who have to do with him? .... Courage, temperance and wisdom benefit both the man who has these dispostions and other people as well; and moral failings such as pride, vanity, worldliness, and avarice harm both their possessor and others, though chiefly perhaps the former. But what about the virtues of charity and justice? These are directly concerned with the welfare of others, and with what is owed to them; and since each may require sacrifice of interests on the part of the virtuous man both may seem to be deleterious to their possessor and beneficial to others....[28]

We might think, then, that some virtues will ground agent centered values, and others will ground agent neutral ones. The agent centered values will be grounded by those virtues which are primarily of benefit to the possessor, since what they dispose me to do may conflict, as our interests conflict, with what they dispose you to do. That is not to say that even these self-centered virtues fail to harmonize in some way, but the harmony will be more like that of a game of basketball, in which the teams have team-centered aims, than it will be like harmony of a group of people working for a common goal. The virtues which are primarily of benefit to the community or humanity in general seem apt to ground agent neutral values, since they are supposed to dispose us to promote a single common goal.

In the tradition of virtue theory, Foot takes very seriously the problem, for she thinks it is a problem, inherent in the fact that some virtues appear deleterious to their possessor. Morality would be deceitful, she thinks, if it recommended a kind of character to us which turned out to be to our detriment. So, for example, in her work on Nietzsche[29] and in her discussion of Thrasymachus' challenge to justice,[30] Foot struggles with the question of whether justice is indeed good for the just person. It is a virtue of virtue theory, it seems to me, that it is able to appreciate the powerful intuitive appeal of Thrasymachus' challenge, where some other approaches must write off such questions as conceptually confused.[31] In at least one version of Virtue Theory, then, it is crucial to be able to show of each virtue that it is beneficial to its possessor (in the context of possession of the other virtues). It is reasonable to believe, though it is not certain, that virtues derived and defended in this manner will ground agent centered moral prescriptions. Here is an example.

Suppose that a requirement for earning a promotion that you have long sought is that you take charge of a committee of important figures in your field. The thought of speaking in front of that group gives you chills and you can barely bring yourself to imagine accepting the post without breaking out in hives. Surely classic virtue theory is correct to say that courage is a virtue in this sense: that someone who has it will be better equipped to deal with situations just such as this.

But at the same time, a competitor for the promotion is having the same thoughts. Should your competitor try to persuade you to take the position? It is obvious that courage, at any rate, does not in any way require that your competitor try to get *you* to overcome your fears. There is nothing courageous about urging someone else to persevere in the face of adversity. Courage requires, rather, that your competitor overcome *her* fears and volunteer *herself* for the job. So courage gives you and your competitor different and indeed incompatible aims.

Of course, when a full virtue theory is spelled out, its final advice to both of you may be that the one of you best qualified for the promotion should volunteer. That advice might follow from the virtue of justice. But it is not obvious that virtue will require such self-sacrifice–for each of you has a great stake in the promotion, and the states of character that are to count as virtues must, at least on many views, be beneficial to the possessor.

Conclusion

Moral theory can progress by illuminating the variety of structures and contents that normative schemes can have, improving our map of the overall landscape. Or, it can progress by attempting to

reduce the variety of plausible or justifiable options, weeding out incoherent ones or trying to drive us to one spot or another by appealing to our settled convictions.

I believe that the contrast between agent neutral and agent centered theories, and not the contrast between consequentialist and nonconsequentialist theories, is a fundamental part of the most useful taxonomy. Of course, there are other fundamental dimensions which I have not considered at all, and which deserve treatment. But as we focus more sharply on the agent neutral/agent centered contrast, certain arguments intended to push us toward (so called) consequentialist schemes lose their cogency.

If we want to find sound reasons to choose between the two kinds of structure, we need to step back a level and ask what kinds of reasoning we should use as our source of justification. I am not advocating a foundationalist moral epistemology with metaethics at the base, or urging that we have to choose a methodology independent of our pre-theoretic moral convictions and stand by its deliverances come what may. My suggestions are entirely consistent with the general method of reflective equilibrium. What's important, if I'm right, is that attention *within* the method of reflective equilibrium to conceptual apparatus at the level of metaethics is necessary if we are to find good reasons for choosing among normative theories. In any case, my aim in this paper is not polemical, but clarificatory and, I hope, suggestive.

Brown University

NOTES

[*] For incalculable assistance sorting through the ideas that I present in this paper, I am indebted to Cliff Landesman, Jimmy Wales, John Broome, and Peter Vallentyne, all of whom corresponded with me extensively via electronic mail. For the old fashioned kind of assistance, I thank David Estlund and Dan Brock, and last Felicia Ackerman, who pointed me to the Orwell references.

[1]See for example Shelly Kagan, *The Limits of Morality* (Oxford: Oxford University Press, 1989), esp. p. 8; Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982), esp. p. 1-3; and Thomas Nagel, in "The Limits of Objectivity," *The Tanner Lectures on Human Value* 1980 (Salt Lake City: University of Utah Press), esp. p.119, and also in "Autonomy and Deontology," in Samuel Scheffler (ed.), *Consequentialism and Its Critics* (Oxford: Oxford University Press, 1988), p. 143.

[2]See Broome's *Weighing Goods*, (New York: Blackwell, 1991), esp. chapter one.

[3]This strategy is not new. It has been offered by Amartya Sen, in "Evaluator Relativity and Consequential Evaluation," *Philosophy and Public Affairs* 12 (1983), among other places.

A broad understanding of the idea of a consequence, designed to include in the consequences of an action features of the agency that brings it about, was developed by Lars Bergstrom in *The Consequences of Action* (Stockholm: Almqvist and Wiskell, 1966). Peter Vallentyne made a persuasive case for including among teleological theories those which allow such "past regarding" aspects of consequences as whether a promise has been broken into the domain of evaluation, in "Teleology, Consequentialism, and the Past," *Journal of Value Inquiry* 22 (1988) pp. 89-101. Others have at least noticed the possibility of drawing all morally relevant features of an action into the consequence, for example Charles Taylor, "The Diversity of Goods," in A. Sen and Bernard Williams (eds.), *Utilitarianism and Beyond* (Cambridge, England: Cambridge University Press, 1982), p. 144, and J. J. C. Smart, in Smart and B. Williams, *Utilitarianism For and Against* (Cambridge, England: Cambridge University Press, 1973), p. 27.

[4]See note 3 above, and also Vallentyne's "Gimmicky Representations of Moral Theories," *Metaphilosophy* 19 (1988) pp. 253-63.

[5]See Nagel, "Autonomy and Deontology," p. 126, for this example.

[6]*Utilitarianism: For and Against*, p. 6.

[7]I believe that Philippa Foot is arguing along these lines in "Utilitarianism and the Virtues," in *Consequentialism and Its Critics*, pp. 224-42.

[8]Henry Sidgwick, *The Methods of Ethics* (Indianapolis: Hackett, 1981), p. 382.

[9]Sidgwick, p. 420.

[10]Sidgwick, p. 498.

[11]Sidgwick, p. 420-1.

[12]Notice that Sidgwick does not generally capitalize "Good," but does here.

[13]I hope the reader will see why I put the pairs of operators in parentheses. If not, see Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970), esp. pp. 47-56 and pp. 90-124; and R. M. Hare, *Moral Thinking* (Oxford: Oxford University Press, 1981), esp. pp. 87-106, to fill out the ways in which the relation of present to future are relevantly similar to the relation of actual to possible. As it happens, both of these accounts are derived, I believe, from Sidgwick.

[14]Derek Parfit uses approximately this notion of holding a value objectively, in *Reasons and Persons*, (Oxford: Oxford University Press, 1984), p. 151. Michael Smith pointed out how the notion of a desire conditional on its own temporal persistence can be extended to the notion of a desire conditional on its own modal persistence, in a seminar at Princeton in 1986.

[15]*The Limits of Morality*, p. 61.

[16]Amartya Sen, "Rights and Agency," in *Consequentialism and Its Critics*, pp. 187-223.

[17]As far as I know, Ronald Dworkin has not come down on one side or the other. In *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1985), p. 108, he says that it is a controversial question. It strikes me as somewhat odd that the author of *Taking Rights Seriously* presents no clear view about whether we have agent centered or agent neutral responsibilities to avoid violation of rights. But in any case, it is obvious that one can have a fairly fully developed theory of rights without answering the structural question!

[18]Roderick Firth, "Ethical Absolutism and the Ideal Observer," in John Hospers and W. Sellars (eds.), *Readings in Ethical Theory* (2nd ed.; Englewood Cliffs, New Jersey: Prentice Hall, 1970), p. 215.

[19]I am thinking especially of R. M. Hare's idea of universalizability in *Moral Thinking*.

[20]Could an Ideal Observer who approved of partiality, but was herself thoroughly impartial, approve of her own behavior? It seems not. I doubt that this self-disapproval would be any defect in the metaethic, but see Michael Stocker "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy* 73 (1976), pp. 453-66. Marcia Baron usefully distinguishes between the kinds of impartiality I am talking about in "Impartiality and Friendship," *Ethics* 101 (1991), pp. 836-57, esp. pp. 842-44.

[21]"Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, pp. 103-28.

[22]*The Limits of Morality*, pp. 41-2. See also John Harsanyi, "Morality and the Theory of Rational Behavior," in *Utilitarianism and Beyond*, pp. 39-62.

[23]"Reflections on Gandhi," in Orwell's *Collected Essays* (London: Secker & Warburg 1961), p. 455.

[24]"The Schizophrenia of Modern Ethical Theories."

[25]"Moral Saints," *Journal of Philosophy* 79 (1982), pp. 419-39.

[26]"On the Primacy of Character," in Rorty and Flanagan (eds.), *Identity, Character, and Morality* (Cambridge, Mass.: MIT Bradford, 1990), pp. 449-69.

[27]"On the Primacy of Character," p. 457.

[28]"Virtues and Vices," in her *Virtues and Vices* (Berkeley: University of California Press, 1978), pp. 2-3.

[29]"Nietzsche: The Revaluation of Values," in *Virtues and Vices*, pp. 81-95.

[30]"Moral Beliefs," in *Virtues and Vices*, pp. 110-31.

[31]As in H. A. Prichard's famous essay, "Does Moral Philosophy Rest on a Mistake?" *Mind* 21 (1912), pp. 21-37; also see John Hospers, *Human Conduct*, (New York: Harcourt, Brace, & World, 1961).