Minds, Ethics, and Conditionals

Themes from the Philosophy of Frank Jackson

Ian Ravenscroft

CLARENDON PRESS · OXFORD

Henderson, D. and Horgan, T. 2000. 'What is a priori and what is it good for?', *Southern Journal of Philosophy* 38 (Spindel Supplement): 51–86.
——2001. 'The *a priori* isn't all that it is cracked up to be, but it is something', *Philosophical Topics* 29 (The Philosophy of Alvin Goldman): 219–50.
Horgan, T. 1993. 'The austere ideology of folk psychology', *Mind and Language* 8: 282–97.
——2001. 'Contextual Semantics and Metaphysical Realism: Truth as Indirect Correspondence'. In M. Lynch (ed.), *The Nature of Truth: Classic and Contemporary Perspectives* (Cambridge, MA: MIT Press).
Horgan, T. and Timmons, M. 1991. 'New wave moral realism meets moral twin earth', *Journal of Philosophical Research* 16: 447–65. Reprinted in J. Heil (ed.), *Rationality, Morality, and Self-Interest* (New York: Rowman & Littlefield, 1993).
——1992a. 'Trouble for New Wave Moral Semantics: The "Open Question Argument" Revived', *Philosophical Papers* 21, 153–75.
——1992b. 'Troubles on moral Twin Earth: Moral queerness revived', *Synthese* 92: 221–60.
——1996a. 'From moral realism to moral relativism in one easy step', *Critica* 28: 3–39.
——1996b. 'Troubles for Michael Smith's metaethical rationalism', *Philosophical Papers* 25: 203–31.
——2000a. 'Copping out on moral Twin Earth', *Synthese* 124: 139–52.
——2000b. 'Nondescriptivist cognitivism: Prolegomenon to a new metaethic', *Philosophical Papers* 29: 121–53.
——2002. 'Conceptual Relativity and Metaphysical Realism', *Philosophical Issues* (Realism and Relativism) 12: 74–96.
——2006a. 'Morality without moral facts'. In J. Drier (ed.), *Contemporary Debates in Moral Theory* (Oxford: Blackwell).
——2006b. 'Cognitivist expressivism'. In T. Horgan and M. Timmons (eds.), *Metaethics after Moore* (Oxford: Oxford University Press).
——2006c. 'Expressivism, yes! Relativism, no!'. In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol 1. (Oxford: Clarendon Press).
Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. (Oxford: Clarendon Press.)
Lewis, D. 1966. 'An argument for the identity theory', *Journal of Philosophy* 63: 17–25.
——1970. 'How to define theoretical terms', *Journal of Philosophy* 67: 427–46.
——1972. 'Psychophysical and theoretical identifications', *Australasian Journal of Philosophy* 50: 249–58.
——1980. 'Mad pain and martian pain'. In N. Block (ed.), *Readings in the Philosophy of Psychology*, vol. 1 (Cambridge, MA: Harvard).
Smith, M. 1994. *The Moral Problem* (Oxford: Blackwell).
Snare, F. 1980. 'The diversity of morals', *Mind* 89: 353–69.
Timmons, M. 1999. *Morality without Foundations: A Defense of Ethical Contextualism* (Oxford: Oxford University Press).

10

Consequentialism and the Nearest and Dearest Objection

MICHAEL SMITH

1. Bloggs's story

Imagine that Bloggs is faced with a choice between giving a benefit to his child, or a slightly greater benefit to a complete stranger. The benefit is whatever the child or the stranger can buy for $100—Bloggs has $100 to give away—and it just so happens that the stranger would buy something from which he would gain a slightly greater benefit than would Bloggs's child. Let's stipulate that Bloggs believes this to be, and let's stipulate, as well, that he believes that the consequences of his actions are otherwise identical. He chooses to give the benefit to his child. What do we learn about Bloggs from his choice? We learn that Bloggs cares more about his child than he does about complete strangers. Nor is anyone likely to be surprised by this, for it just goes to show that he is much like the rest of us. He gives preferential treatment to his nearest and dearest when he acts, those with whom he has a special relationship, much as we do.

Now imagine that we ask Bloggs to justify his choice. Suppose he says that he did what was best, and that it was the best thing for him to do because the benefit went to *his child*. What do we learn about Bloggs from his attempt to provide a justification for giving the benefit to his child? Assuming that this is supposed to be the most basic value relevant to his choice we learn that, in addition to caring more about his child than he does about complete strangers, he thinks that, in so doing, he cares about what is of fundamental value. As he sees things, there is a distinctive form of value—'relative' value, as it has come to be called (Parfit 1984: 27)—realized in his giving the smaller benefit to his child. That is what is signalled by his citing the fact that the benefit accrues to *his* child: this is the special relationship just mentioned. Moreover Bloggs

thinks, correctly, that this kind of value wouldn't be realized if he instead gave the greater benefit to the stranger.

We can represent Blogg's universalized belief, the belief from which he derives his belief about the value of giving the benefit to his child, as follows (Smith 2003).

R: $\forall x$ (x's child's enjoying a benefit is $good_x$)

R entails that there is $value_{Bloggs}$ in Bloggs's children enjoying benefits, and it also entails that there is $value_{someone\ else}$ in someone else's children's enjoying benefits. But it leaves it open that there is no $value_{Bloggs}$ in someone else's children's enjoying benefits, and no $value_{someone\ else}$ in Bloggs's children enjoying benefits. Assuming that R is true, and assuming that the question relevant to the justification of choice is what has $value_{chooser}$, Bloggs's choice would thus seem to be justified.

Finally, imagine we challenge Bloggs's justification. 'You must surely agree', we say, 'that there is value of another form, namely neutral value, that attaches to the benefits enjoyed by your child and the stranger. For just imagine that you didn't bear any special relationship to either your child or to the complete stranger. Wouldn't you then assign at least some value to the benefit each receives, and wouldn't the amount of value you assign covary with the level of the benefit? If so, then you will surely agree that there is more value of this form in giving the greater benefit to the complete stranger. So how can you think that the value that attaches to giving the benefit to your child justifies your giving the smaller benefit to him when you must also agree that there is more value in giving the greater benefit to the complete stranger?'

Bloggs's response, let's suppose, is to concede that there is value of another form—'neutral' value, as it has come to be called (Parfit 1984: 27)—realized in giving the larger benefit to the complete stranger, and that there is more value of this form realized in giving the larger benefit to the complete stranger than there is in giving the smaller benefit to his child. But he insists that the additional relative value realized in giving the smaller benefit to his child more than compensates for that loss of neutral value. Indeed, we can suppose that Bloggs thinks that if the loss in neutral value had been much larger, but the gain in relative value had remained much the same, then the additional relative value would not have compensated for the loss. His justification is thus specific to the circumstances in which he makes his choice. His claim is simply that, in these circumstances, the gain in relative value compensates for the loss of neutral value.

We can represent the content of the universalized belief from which Bloggs derives his beliefs about the neutral value of the benefits that accrue to both the stranger and his child as follows (Smith 2003).

N: $\forall x\ \forall y$ (Benefits enjoyed by the maximal aggregate composed inter alia of y are $good_x$)

N entails that benefits enjoyed by the aggregate composed of Bloggs's child and the complete stranger are both $good_{Bloggs}$, and it presumably entails that the greater the level of the benefit that that aggregate enjoys, the $better_{Bloggs}$. In other words, unlike R, N doesn't discriminate between a benefit enjoyed by Bloggs's child and a benefit enjoyed by a complete stranger. Each of these benefits contribute to the level of benefit enjoyed by aggregates of which they are both a part, and so each has $value_{Bloggs}$.

In these terms we can now represent the way in which Bloggs weighs the values posited by R and N. Though, in virtue of the truth of N, Bloggs reasons that it would be $better_{Bloggs}$ to give the benefit to the complete stranger, he also reasons that there is, in virtue of the truth of R, additional $goodness_{Bloggs}$ in the benefit enjoyed by his child. In the circumstances, he thinks that this additional $goodness_{Bloggs}$ tips the balance, making it the case that there is more $goodness_{Bloggs}$ overall—that is, more $goodness_{Bloggs}$ in virtue of the truth of both R and N—in the benefit enjoyed by his child than there is in the benefit enjoyed by the complete stranger. As a result, he concludes that it is $better_{Bloggs}$ overall for him to give the benefit to his child.

Finally, suppose we tell Bloggs that he really shouldn't give the smaller benefit to his child; that he should give the larger benefit to the complete stranger none the less. What might Bloggs say in response? One obvious response would be that we are mistaken, as our view of what he should do seems to be based solely on the neutral values at stake, whereas there are relative values at stake too, values of which we are either ignorant or which we are willfully ignoring. In either case, he might say, we should think again. This would be to object to what we say he should do on the grounds that it is false.

However suppose we convince him that we aren't willfully ignoring relative values, and that it is he who is mistaken: there are no relative values, and hence only neutral values are at stake. Another perhaps less obvious response, but one which he might well give at this point, is that it would ruin his life, and perhaps his child's as well, if he had to systematically ignore his special relationship with his child in making such choices: that consistently doing what we say he should do is inconsistent with his living a worthwhile life, and perhaps with his child's living a worthwhile life too. This would not be to object to what

we say he should do on the grounds that it is false. Indeed, it isn't clear that this is any sort of objection to what we say he should do. Bloggs seems rather to be taking it for granted that he should do what we say he should do. He is simply putting on record the personal cost to him of his doing so, and the cost to his child.

2. Big 'C' consequentialism and the nearest and dearest objection

Bloggs's story brings into sharp relief the debate over big 'C' consequentialism: that's big 'C', as opposed to small 'c', consequentialism. Big 'C' consequentialism makes two crucial claims. First, it makes a substantive claim about the nature of value. It says that all values are neutral (Parfit 1984). Second, it makes a conceptual claim about the nature of obligation. It says that facts about what we ought to do can be analysed in terms of facts about which of the various things that we can do will maximize value (we will return to the details of this analysis in a moment).

Unsurprisingly, since big 'C' consequentialism entails that there are no relative values, it follows that, in Bloggs's choice situation, there is no additional relative value realized in his giving the smaller benefit to his child to justify his doing that rather than giving the greater benefit to the complete stranger. Big 'C' consequentialism thus holds that Bloggs acts wrongly when he gives the $100 to his child. Having said that, however, we must immediately add that what Bloggs does isn't at odds with small 'c' consequentialism. Small 'c' consequentialism agrees with big 'C' consequentialism's conceptual claim about the nature of obligation. It simply denies big 'C' consequentialism's substantive claim about the nature of value. Whereas big 'C' consequentialism holds that all values are neutral, small 'c' consequentialism is silent about the nature of values. Values might all be neutral, or they might all be relative, or some might be neutral while others are relative. To repeat, small 'c' consequentialism, unlike big 'C' consequentialism, takes no stand on this substantive issue about the form of the values.

Bloggs implicitly committed himself to small 'c' consequentialism when he attempted to justify giving $100 to his child. For he argued that, when we take into account the value that derives from both N and R, we see that his giving the smaller benefit to his child is actually $\text{better}_{\text{Bloggs}}$ than giving the larger benefit to the complete stranger. What Bloggs appeals to is thus the following principle connecting indexed value with obligation:

(O) $\forall x$ (x ought to ϕ iff ϕ-ing is best_x)

This is, in effect, the claim that the justification of choice is a matter of what has $\text{value}_{\text{chooser}}$ alluded to above.

Frank Jackson joins this debate over big 'C' consequentialism in his 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection' (Jackson 1991). Here is the opening passage of Jackson's paper.

> Our lives are given shape, meaning and value by what we hold dear, by those persons and life projects to which we are especially committed. This implies that when we act we must give a special place to those persons (typically our family and friends) and those projects. But, according to consequentialism classically conceived, the rightness and wrongness of an action is determined by the action's consequences considered impartially, without reference to the agent whose actions they are consequences of. It is the nature of any particular consequence that matters, not the identity of the agent responsible for the consequence. It seems then that consequentialism is in conflict with what makes life worth living. (Jackson 1991: 461)

Jackson's suggestion is thus that big 'C' consequentialism 'is in conflict with what makes life worth living' because it 'would, given the way things more or less are, render the morally good life not worth living'.

Unfortunately, Jackson doesn't ever spell out what exactly this conflict is supposed to amount to. Indeed, the only other time he explicitly mentions the nature of the conflict is right at the end of the paper when he sums up.

> My concern ... has been to reply to the objection that consequentialism would, given the way things more or less are, render the morally good life not worth living. I take this to be the really disturbing aspect of the nearest and dearest objection. Consequentialists ... cannot live with the conflict with a life worth living, given the way things more or less are. That would be to invite the challenge that their conception of what ought to be done had lost touch with *human* morality. (Jackson 1991: 482)

Before considering the details of his response, we therefore need to bring out what, as he sees things, the conflict between big 'C' consequentialism and living a life worth living is supposed to be.

The reason for doing this is that, at first blush at any rate, Jackson looks to be making a version of the second response that we imagined Bloggs making above. But, much as we said in that case, this response doesn't seem to constitute an objection to big 'C' consequentialism at all. It seems rather to be an observation about the personal cost of doing what the theory tells an agent to do. Indeed, it is hard to see how the mere possibility of a conflict between an agent's living up to the obligations posited by a moral theory and that agent's living a life worth living could constitute an objection to that moral theory. This is because every moral theory, or anyway every

plausible moral theory, will tell people that they should sometimes act in ways that will make their own lives not worth living. All that is required is that they have the bad luck to find themselves in the right (wrong?) kind of circumstances.

Jackson in fact considers this reply to the objection with which he is concerned at the very beginning of his paper.

> One way to reply ... would be to break the implicit connection between acting morally and living a life worth living. Doing what is morally right or morally required is one thing; doing what makes life worth living is another. Hence, runs the reply, it is no refutation of a moral theory that doing as it enjoins would rob life of its shape and meaning. This is a chilling reply and I will say no more about it. (Jackson 1991: 461)

But there is nothing especially chilling about a reply that points out an obvious consequence of every plausible moral theory.

Imagine a case in which an agent finds himself faced with a choice between either submitting to torture and subsequent death, so making his life not worth living, or his bringing about ... and here substitute whatever your favourite moral theory deems to be an even worse outcome than this agent's torture and subsequent death. It doesn't matter whether your favourite moral theory says that what makes outcomes worse is their realizing less neutral value, or their realizing less relative value, or their realizing less of some weighted sum of value of both kinds. So long as the agent has to choose between making his own life not worth living and bringing about something that is even worse, by the theory's own lights—and, to repeat, every plausible moral theory will say that there is a worse outcome than one agent's living a life that is not worth living—then every plausible moral theory will tell that agent to make his own life not worth living. This will simply be an application of O.

The upshot is that, if there is to be an objection to big 'C' consequentialism based on the fact that it is in conflict with an agent's living a worthwhile life, then the objection will have to be that the conflict with big 'C' consequentialism is in some way much more direct and systematic than the conflict we have just noted. I take it that Jackson thinks that the conflict between big 'C' consequentialism and living a life worth living is indeed more direct and systematic, and that this explains why he finds the biting the bullet on the objection to be such a chilling prospect. But is it plausible to suppose that there is such a direct and systematic conflict?

Derek Parfit famously argues that it is indeed plausible. He suggests that big 'C' Consequentialism—or 'C', as he calls it—is 'indirectly collectively self-defeating'.

> Call [a moral theory] T *indirectly collectively self-defeating* when it is true that, if several people try to achieve their T-given aims, these aims would be worse achieved.
>
> On all or most of its different versions, this may be true of C. C implies that, whenever we can, we should try to do what would make the outcome as good as possible. If we are disposed to act in this way, we are *pure do-gooders*. If we were all pure do-gooders, this might make the outcome worse. This might be true even if we always did what, of the acts that are possible for us, would make the outcome best. The bad effects would come, not from our acts, but from our disposition. (Parfit 1984: 27)

As I understand it, the nearest and dearest objection that worries Jackson is best understood as a particular manifestation of big 'C' consequentialism's being indirectly collectively self-defeating. The objection can be brought out by imagining a third response Bloggs might have made to our suggestion that he is mistaken and should give the $100 to the complete stranger.

Bloggs might say, 'Suppose we all consistently try and succeed in doing what big "C" consequentialism tells us to do, which is to maximize neutral value. This would require that the desire for what is neutrally valuable—that is, the desire for benefits, independently of who receives them—is the strongest of our desires. But if this was the strongest of our desires then the lives we would each end up leading would be much worse—that is, there would be less neutral value in each of our lives, less benefits, considered as a whole—than there would have been if we had had different desires, and so acted differently. This is because our each having, as our strongest desire from time to time, the desire that benefits accrue to our nearest and dearest, is a crucial factor in causing benefits both to ourselves and to our nearest and dearest. Yet if we sometimes have, as our strongest desire, the desire to give benefits to our nearest and dearest, then it isn't the case that we always have, as our strongest desire, the desire that there be as many benefits as possible for people. So our leading lives that are worth leading—that is, our leading lives in which neutral value is maximized—is plausibly inconsistent with our always performing acts that maximize neutral value. So big "C" consequentialism's account of which acts we should perform must be mistaken.'

The crucial assumption Bloggs makes in giving this response is that, as Parfit puts it, effects can 'come, not from our acts, but from our disposition.' This seems to be what Jackson has in mind too when he says, in the passage quoted at the outset, that our 'lives are given shape, meaning and value by what we hold dear, by those persons and life projects to which we are especially committed.' The fact that we sometimes have, as our strongest desire, the desire that benefits accrue to our nearest and dearest, is an important causal factor in producing benefits both to ourselves and to our nearest and dearest independently of the

acts that such a desire produces (Adams 1976). But in that case it follows that if we all try and succeed in doing what big 'C' consequentialism tells us that we should do, which is to perform acts that maximize neutral value, then the outcome may well be worse in big 'C' consequentialism's own terms. The outcome would have been better if we had been differently motivated, and so had failed to do what big 'C' consequentialism tells us to do. Big 'C' consequentialism thus appears to condemn its own account of what we ought to do. For it tells us that things may be worse when we try and succeed in making them better.

3. Jackson's reply to the nearest and dearest objection

Jackson's initial reply to the nearest and dearest objection, so understood, takes issue with the premiss that big 'C' consequentialism tells us to act in ways which are such that, if we consistently acted in those ways, we would lead lives that are not worth living. His reason for rejecting this premiss turns on his view that, according to most plausible formulation of big 'C' consequentialism, what agents should do is not act so as to maximize neutral value—that's what we assumed in constructing the argument against big 'C' consequentialism—but rather act so as to maximize *expected* neutral value. I will comment on the plausibility of this view presently. For the moment, however, let me spell out how, as Jackson sees things, accepting this alternative formulation of big 'C' consequentialism provides us with (the beginnings of) a reply to the nearest and dearest objection.

As Jackson points out, we generally have a much better idea of what will benefit those who are known to us, and only the vaguest idea of what will benefit complete strangers. For example, if I give my child $100, I have a pretty good idea of what he will spend it on, but I have relatively little idea of what a complete stranger would spend $100 on. Moreover, even when we do know what a complete stranger will spend $100 on, the causal pathway from any action that we perform to the delivery of that benefit is usually somewhat more unpredictable in the case of a complete stranger than it is in the case of one well known to me. I place $100 in my child's hand, and, because I know him and his habits, I am confident that he will keep it and spend it in ways that will benefit him. But I am rarely in a position to identify a complete stranger to whom I could provide a substantial benefit by putting $100 in his hand. I rely on an aid agency to identify such strangers for me, and then it becomes somewhat obscure how much of what I give to the aid agency, if anything, ends up in the hands of the complete stranger.

Generally speaking, then, it follows that limited creatures like ourselves will have much higher confidence that our actions will produce benefits for those well known to us, and much lower confidence that our actions will produce benefits for complete strangers. But if this is right then, even assuming that benefits are of neutral value only, it follows that acting in ways that produce smaller benefits for those well known to us, rather than greater benefits for complete strangers, will, generally speaking, have greater expected neutral value. To be sure, on each such occasion there will be an alternative action available to us that would produce more neutral value in fact, namely, the action of providing the greater benefit to a complete stranger. But since we have only a very low expectation that any particular act we could perform is one which would provide a greater benefit to some complete stranger, the expected neutral value of providing a smaller benefit to those well known to us will, in most cases, be much greater than the expected neutral value of providing a greater benefit to a complete stranger.

Jackson notes a potential problem with this initial reply to the nearest and dearest objection.

> It might well be objected that we can distinguish *two* nearest and dearest objections, and that I have replied to only one of them. One objection is, "How can consequentialists make sense of the fact that there is a relatively small group of people whose welfare plays a special role in our lives, given the agent-neutral nature of consequentialism's value function?" Our reply was that ... right value translates into right action ... through an agent's beliefs, and that when this is appreciated, empirical facts about our cognitive powers and situation make it plausible that our actions should be highly focussed much of the time. The other objection is, "How can consequentialists make sense of it being the *particular* small group of people that it mostly is?" Perhaps consequentialism can make sense of there being a small group, but why the small group of family, friends, fellow citizens, and the like that it so often is? (Jackson 1991: 478)

As I said, Jackson's initial reply to the nearest and dearest objection is that it presupposes an implausible conception of what, according to big 'C' consequentialism, we should do. But what the potential problem shows is this isn't the case. For suppose that we substitute his preferred conception. This makes no difference at all. Someone who consistently maximizes *expected* neutral value, even someone with limits on their knowledge like those mentioned, is, for all we've been told, *still* someone whose strongest desire is the desire for what is of neutral value. Someone who consistently maximizes expected neutral value is therefore *still* someone who fails to have the desires that make their life worth living. For they do not have, as their strongest desire from time to time, the desire to provide benefits to their family, friends, fellow citizens, and the like.

Jackson thus goes on to supplement his initial reply with the further suggestion that additional empirical facts explain why the particular small group on which we focus is the small group of family, friends, fellow citizens, and the like.

> [I]n deciding what to do here and now an agent must take account of what he or she will do in the future, and that involves taking very seriously questions of character. Do I have the persistence that will be called for, will I remain sufficiently enthusiastic about the project to put in the time required, will I be able to retain a sufficiently impartial outlook, will I be able to avoid the various temptations that will arise, and so on and so forth? ... [A]s a rule we do better for reasons of character (that no doubt have an evolutionary explanation) with projects that involve family and friends rather than strangers. This is simply because we are much less likely to lose the enthusiasm required to see the project through to a successful conclusion when the project benefits people we have a particular affection for. (Jackson 1991: 480)

As Jackson sees things, given that we are enjoined to maximize expected neutral value, these two empirical considerations—the limits of our knowledge, and the limits of our self-control—combine to focus our attention on the benefits that we can provide for our nearest and dearest.

Jackson's reply suggests that when we told Bloggs's story at the outset we went badly wrong when we stipulated that he believes that, were the complete stranger to receive the $100, he would buy something from which he would gain a slightly greater benefit than would Bloggs's own child. Jackson's point is, in essence, that we thereby stipulated something empirically implausible. To be empirically plausible we would have to imagine that Bloggs is far less confident about the level of benefit a complete stranger will gain from giving $100 to him than he is about the level of benefit that would accrue to his child from being given $100. So when we ask which of the actions available to Bloggs will maximize expected benefit the answer is going to turn out to be giving the $100 to his child.

If Jackson is right then we see that Bloggs's third response to the suggestion that he should give $100 to the complete stranger turns on a false claim. Bloggs's response was supposed to be that big 'C' consequentialism condemns itself: by big 'C' consequentialism's own lights, things may be worse when we try and succeed in making them better. The false premiss in Bloggs's argument for this conclusion is the premiss that if we sometimes have, as our strongest desire, the desire to give benefits to our nearest and dearest, then it isn't the case that our strongest desire is always the desire that there be as many benefits as possible for people. For, given the limits of our knowledge and self-control, it turns out that the acts that maximize expected benefits for people generally *are* the acts that maximize expected benefits for our nearest and dearest. There is therefore no inconsistency in having a pair of desires tied as our strongest desires. Our actions can all be overdetermined. For these desires—the desire that our nearest and dearest benefit, and the desire that the people in general benefit—don't require us to act in different ways, not given our limited knowledge and self-control. So even if we all try and succeed in doing what big 'C' consequentialism tells us to do we can still have, within ourselves, the crucial desire whose possession is necessary for our living a worthwhile life.

What should we make of this reply? Big 'C' consequentialism is a moral theory. I take it that this means that it is supposed to tell us what makes acts right and wrong, not just in the actual world, but in every possible world. Now what the nearest and dearest objection purports to show is that there is something wrong with big 'C' consequentialism by the theory's own lights; for, to repeat, the theory itself seems to tell us that the outcome is worse if we all try and succeed in making it better. But it is hard to see the relevance of the empirical considerations Jackson adduces to this objection. To be sure, if the people mentioned in the crucial premiss of the nearest and dearest objection are people like you and me, people with limited knowledge and self-control, then the premiss of the objection is false. But in that case we should just assume that the people mentioned in the premiss of the objection don't have (say) such limited self-control. For the premiss of the argument is in that case true, and the conclusion then apparently follows.

Nor is this a fanciful reply. Jackson himself admits that there are people in the actual world who are much more self-controlled than the rest of us, people like Mother Theresa and Ralph Nader (Jackson 1991, p.481). They seem both willing and able to act in ways that increase expected neutral value for a group that isn't the small group of their family, friends, fellow citizens, and the like, and they seem willing and able to do so notwithstanding the fact that they make their own lives worse than they could have been as a result. So let's just imagine a community of people who are all more like Ralph Nader and Mother Theresa. When the people in this community all try and succeed in making the outcome better then, by big 'C' consequentialism's own lights, the outcome is worse than it would have been if they had been differently motivated, and so had acted differently. If this is an objection to big 'C' consequentialism at all, then the mere possibility of such a community seems to be all that's required to bring that objection to the fore.

It therefore seems to me that Jackson's reply to the nearest and dearest objection misses its mark. It does not help with that objection to reformulate

big 'C' consequentialism as the doctrine that we should maximize, not neutral value simpliciter, but expected neutral value.

4. Parfit's reply to the nearest and dearest objection

So far we have simply taken it for granted that Blogg's third response to our suggestion that he should give the $100 to the complete stranger constituted an objection to big 'C' consequentialism. If the members of a community could all try and succeed in making the outcome better, by big 'C' consequentialism's own lights, and yet the outcome is worse, again by the theory's own lights, than it would have been if they had been differently motivated, and so had acted differently, then, we have assumed, this does indeed constitute an objection to big 'C' consequentialism. But does it really?

Like Mill, Sidgwick and others, Parfit insists that we distinguish sharply between the different things that can be evaluated in big 'C' consequentialist terms (Mill 1861, Sidgwick 1907, Adams 1976, Hare 1981, Railton 1984). One question we can ask is which act an agent should perform. In big 'C' consequentialist terms, this amounts to the question, 'Of the various acts that an agent could perform, which is the act that will produce the greatest neutral value?' Another question we can ask is which desires an agent should possess. In big 'C' consequentialist terms, this amounts to the question, 'Of the various desires that an agent could have, which are the desires whose possession by him will produce the greatest neutral value?' And yet another question we can ask is which life an agent should live. In big 'C' consequentialist terms, this amounts to the question, 'Of the various lives an agent could live, which is the life the living of which by him will produce the greatest neutral value?' And so we could go on (Parfit 1984 pp.28–9; see also Railton 1988, Brink 1989, Pettit and Smith 2000).

Parfit points out that so long as the items up for evaluation have different consequences from each other, big 'C' consequentialism's answers to these different questions will be logically independent of each other. Notwithstanding the intimate connection between the desires people have and the acts that those desires produce, it therefore follows that, since their desires can have effects independently of these acts, big 'C' consequentialism may tell people to perform certain acts, but also tell them to have desires which are quite different from those that they would have if they were to perform all of those acts. The same is true of the lives that people lead and the acts that they perform in leading their life. Since too these have different effects, it follows that big 'C' consequentialism may tell people to perform certain acts, but also tell them to lead a life which is quite different from the life that they would lead if they did perform all of those acts. Far from condemning itself, big 'C' consequentialism thus simply applies its principle of evaluation perfectly consistently to anything and everything that we might want to evaluate.

To my mind this provides us with a powerful and decisive reply to the version of the nearest and dearest objection that worries Jackson. The reply concedes the premiss of the objection. Big 'C' consequentialism does indeed tell us to act in ways which are such that, if we consistently acted in those ways, we would lead lives that are not worth living. For what the reply denies is that it follows from this that big 'C' consequentialism tells us to lead such lives. On the contrary, the reply goes, big 'C' consequentialism tells each of us to live that life, of those we could live, which is such that, by living that life, we maximize neutral value. The premiss of the nearest and dearest objection thus doesn't support the conclusion. It gives us no reason to believe that big 'C' consequentialism condemns itself. Big 'C' consequentialism doesn't condemn itself. Rather, to repeat, it requires us to apply its principle of evaluation consistently to everything that we might want to evaluate—actions, desires, lives—independently of our application of it to anything else.

At one point Jackson considers this line of reply to the nearest and dearest objection, but rejects it.

> I am not here denying the correct and important point that some particular action may be wrong in consequentialist terms and yet spring from a character which is right in consequentialist terms. I am denying that the point helps with the essentials of the nearest and dearest objection. For the consequences of having a character which gives a special place in one's affections and concerns to those persons who are closest to one are, in the main, consequences of the manifestations of such a character, that is, of the actions which are especially directed to the needs of those closest to us. Hence, a consequentialist justification of such a character presupposes a consequentialist justification of those actions—which returns us to the very question raised by the nearest and dearest objection. (Jackson 1991: 479)

But I am not sure how Jackson can say what he goes on to say, given the qualifier 'in the main'. The crucial point is that, since the consequences of our actions and our character are non-identical, no conclusion about the character we ought to have, or the life we ought to lead, can be drawn from big 'C' consequentialism's answer to the question 'Which actions ought we to perform?' This completely undermines the nearest and dearest objection because that objection, in effect, attempts to draw just such a conclusion.

5. Jackson's preferred formulation of big 'C' consequentialism

As we saw, Jackson makes a good deal of the fact that big 'C' consequentialism requires us to act so as to maximize not neutral value simpliciter, but rather expected neutral value. Though, as we have seen, formulating the doctrine in these terms doesn't help with the nearest and dearest objection, the argument he gives for formulating big 'C' consequentialism in this way is worth considering on its own merits.

The argument proceeds by way of a discussion of two examples which are supposed to help us choose between the competing formulations. There are many formulations to consider, but Jackson restricts himself to considering three. According to one, the formulation which we gave initially, big 'C' consequentialism tells us that we ought to act so as maximize neutral value. According to the second, it tells us that we ought to act so as to have the best chance of maximizing neutral value. And according to the third, the formulation which Jackson prefers, big 'C' consequentialism tells us that we ought to act so as maximize expected neutral value.

As I said, Jackson's argument for formulating big 'C' consequentialism in terms of expected neutral value proceeds by way of a discussion of examples. Here is the first example.

> Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B, and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way she can tell which of the two is the perfect cure and which the killer drug. What should Jill do?
>
> The possible outcomes we need to consider are: a complete cure for John, a partial cure, and death. It is clear how to rank them: a complete cure is best, followed by a partial cure, and worst is John's death ... But how do we move from that ranking to a resolution concerning what Jill ought to do? The obvious answer is to take a leaf out of decision theory's book and take the results of multiplying the value of each possible outcome given that the action is performed, summing these for each action, and then designating the action with the greatest sum as what ought to be done. In our example there will be three sums to consider, namely:
>
> *Pr*(partial cure/drug A taken) x *V*(partial cure) + *Pr*(no change/drug A taken) x *V*(no change);
>
> *Pr*(complete cure/drug B taken) x *V*(complete cure) + *Pr*(death/drug B taken) x *V*(death); and
>
> *Pr*(complete cure/drug C taken) x *V*(complete cure) + *Pr*(death/drug C taken) x *V*(death).
>
> Obviously, in the situation as described, the first will take the highest value, and so we get the answer that Jill should prescribe drug A. (Jackson 1991: 462–3)

What this example is supposed to show is that it would be implausible to formulate big 'C' consequentialism as the doctrine that we ought to act so as to maximize neutral value. For neutral value is at a maximum when John is completely cured, so Jill's prescribing drug A, which will partially but not completely cure him, is, by this criterion, definitely *not* the thing that we ought to do. This leaves us with only two alternatives: prescribing either drug B or drug C. But, Jackson says, 'We would be horrified if she prescribed drug B, and horrified if she prescribed drug C.' (Jackson 1991, p.466) So, Jackson concludes, it can hardly be plausible to suppose that big 'C' consequentialism tells us that we ought to maximize neutral value.

Here is Jackson's second example.

> As before, Jill is the doctor and John is the patient with the skin problem. But this time Jill has only two drugs, drug X and drug Y, at her disposal which have any chance of effecting a cure. Drug X has a 90% chance of curing the patient but also has a 10% chance of killing him; drug Y has a 50% chance of curing the patient but has no bad side effects. Jill's choice is between prescribing X or prescribing Y. It is clear that she should prescribe Y, and yet that course of action is not the course of action most likely to have the best results. (Jackson 1991: 467)

What this second example is supposed to show is that it would be implausible to formulate big 'C' consequentialism as the doctrine that we ought to act so as to have the best chance of maximizing neutral value. For Jill's prescribing drug X has a 90% chance of maximizing neutral value—that is, of bringing about a complete cure—whereas prescribing drug Y only has a 50% chance of having this result. Yet what Jill plainly should do is prescribe drug Y. Jill should prescribe drug Y, notwithstanding the fact that it does not have the greatest chance of producing the best result.

Taken together, what the two examples are supposed to show is that big 'C' consequentialism is best formulated as the doctrine that we ought to act so as to maximize expected neutral value. For this formulation supposedly gives us the right answers in both examples. It tells us that Jill ought to prescribe drug A in the three drugs example, and it tells us that she ought to prescribe drug Y in the two drugs example. But while I admit that this sounds superficially plausible, it seems to me that, on further reflection, we see that Jackson's use of the two examples to support the expected value formulation of big 'C' consequentialism is flawed in a quite fundamental way. Insofar as it is a doctrine

about what we ought to do, big 'C' consequentialism is best formulated as the doctrine that we ought to act so as to maximize neutral value (see also Parfit 1984, Railton 1984, Brink 1989). This means that I disagree with what Jackson says about the three drugs example.

We get a hint of the problem with Jackson's argument very early on when he explains how we can represent what big 'C' consequentialism tells agents to do in terms of what a variant on standard decision theory tells them that they ought to do.

> [W]e can think of consequentialism's value function as telling us what, according to consequentialism, we ought to desire. For a person's desires can be represented—with, of course, a fair degree of idealization—by a preference function which ranks states of affairs in terms of how much the person would like the state of affairs to obtain, and we can think of consequentialism as saying that the desires a person ought to have are those which would be represented by a preference function which coincided with consequentialism's value function. The other ingredient in the decision-theoretic account of what consequentialism says a person ought to do, the agent's subjective probability function, is an idealization of the agent's beliefs. Hence, the decision-theoretic account is one in terms of what the person ought to desire and in fact believes. (Jackson 1991: 464)

The suggestion is thus that what big 'C' consequentialism tells agents to do is what they ought to do according to a variant on standard decision theory, the variant in which we combine what they in fact believe with the desires which big 'C' consequentialism tells them they ought to have.

There is a major problem with this suggestion, however. For we need to provide an interpretation of the 'ought' that makes the crucial sentence in this passage—'we can think of consequentialism as saying that the desires a person ought to have are those which would be represented by a preference function which coincided with consequentialism's value function'—come out true. There are three interpretations to be considered: I will call these the moral, the formal and the critical interpretations of the 'ought'. Let me begin with the moral interpretation, just to set it to one side.

What the quoted sentence says, on the moral interpretation, is that big 'C' consequentialism tells us that the desires a person ought (morally) to have are those that coincide with big 'C' consequentialism's value function. But this is evidently false. As we saw in our discussion of Parfit's response to the nearest and dearest objection, just as big 'C' consequentialism says that people ought (morally) to perform those acts, of those they could perform, that maximize (expected) neutral value, it says that people ought (morally) to have those desires, of those they could have, which are such that their possession of those

desires maximizes (expected) neutral value. This straightforwardly conflicts with Jackson's claim in the quoted sentence.

Suppose big 'C' consequentialism tells us that only one thing is of value, namely pleasure. What Jackson says in the quoted sentence, on the moral interpretation of the 'ought', is that big 'C' consequentialism tells us that people ought (morally) to have the desire for pleasure, whereas, to repeat, what big 'C' consequentialism in fact tells us is that people ought (morally) to have that set of desires whose possession maximizes (expected) pleasure. This may be a set of desires that consists of one member, the desire for pleasure, but, equally, it may not be. Everything depends on whether possession of the desire for pleasure maximizes (expected) pleasure. So on the moral interpretation of the 'ought', the quoted sentence isn't obviously true at all.

In any event, the moral interpretation of the sentence is plainly wrong-headed. What Jackson is trying to do is to formulate big 'C' consequentialism's account of what we ought (morally) to do, so we can hardly use the big 'C' consequentialist principle itself in formulating that account. If we already have a big 'C' consequentialist principle, why not just apply it directly to acts? As Jackson himself puts it in discussing a related idea:

> [P]erhaps we are being offered a *variant* on consequentialism according to which an action is to be judged not directly but via the status, judged consequentially, of the character which gives rise to it, but then we appear to be landed with a dubious compromise reminiscent of rule utilitarianism. If consequences are the key in one place, why not across the board? (Jackson 1991: 479)

Let's therefore put the moral interpretation to one side. This leaves us with two alternative interpretations: what I've called the formal, and the critical interpretations.

The idea behind the formal interpretation of the 'ought' is that, just as belief has the formal aim of truth, from which it follows that people ought (in this formal sense) to have true beliefs or knowledge, so desire has, as its formal aim, the good, from which it follows that people ought (in this formal sense) to desire what is, in fact, good. If, for example, pleasure is good, then, the suggestion goes, people ought (formally) to desire pleasure. So, just as big 'C' consequentialism's claim that people ought (morally) to have those beliefs whose possession maximizes neutral value is no objection to the claim that people ought (formally) to have true beliefs, so, according to this alternative interpretation of the quoted sentence, big 'C' consequentialism's claim that people ought (morally) to have those desires whose possession maximizes neutral value is no objection to the claim that people ought (formally) to desire

what is, in fact, good. It is the latter interpretation that is crucial in formulating big 'C' consequentialism decision-theoretically. Or so the suggestion goes. People ought (morally) to do what they ought to do according to the variant on decision theory in which we replace their actual desires with the desires that they ought (formally) to have.

The problem with this formal interpretation of the 'ought' in the quoted sentence, however, is that though it makes the quoted sentence come out true, it also makes much more glaring Jackson's lack of even-handedness in explaining how we can represent what big 'C' consequentialism tells people to do in terms of what they ought to do according to a variant on standard decision theory. For if, in formulating the variant on standard decision theory, regular desires are to be replaced by desires which meet their formal aim, then shouldn't regular beliefs also be replaced by beliefs that meet their formal aim? In other words, shouldn't regular beliefs be replaced with knowledge? Jackson notes this possibility himself at one point:

> [I]n addition to distinguishing what a person in fact desires from what he or she ought to desire, we also distinguish what a person in fact believes from what a person ought to believe ... Hence, it might well be suggested that we should recover the consequentialist answer to what a person ought to do from the value function via what that person ought to believe rather than from what he or she in fact believes. (Jackson 1991: 464)

But if we were to do this then, given that what a person ought (formally) to believe is what is true—since a person ought (formally) to have knowledge—it would follow that, according to big 'C' consequentialism, people ought to maximize neutral value, not maximize expected neutral value. For this is what agents ought to do according to the variant on standard decision theory that replaces an agent's ordinary desires and beliefs with desires and beliefs that meet their formal aim. Yet this is the formulation of big 'C' consequentialism that was supposedly refuted by the three drugs example. Something therefore seems to have gone badly wrong. To anticipate what is to come, if the three drugs example does indeed tell against this formulation of big 'C' consequentialism, then surely we must suppose that it tells equally against both the suggestion that we replace ordinary beliefs by beliefs that meet their formal aim, and ordinary desires with desires that meet their formal aim.

As I said, Jackson himself notes the possibility that we should formulate big 'C' consequentialism decision-theoretically by replacing an agent's desires and beliefs with desires and beliefs that both meet their formal aim. He doesn't, however, think that this idea is very plausible. In other words, he denies that we should be even-handed in our treatment of belief and desire when we

formulate big 'C' consequentialism decision-theoretically. He explains why in the course of giving a response to what he calls an 'annoying complication'.

> ... I need to note an annoying complication. I have been arguing for an interpretation of consequentialism which makes what an agent ought to do the act which has the greatest expected moral utility, and so is a function of the consequentialist value function and the agent's probability function at the time. But an agent's probability function at the time of action may differ from her function at other times, and form the probability function of other persons at the same or other times. What happens if we substitute one of these other functions in place of the agent's probability function at the time of action? The answer is that we get an annoying profusion of 'oughts' ...
>
> I think that we have no alternative but to recognize a whole range of oughts—what she ought to do by the lights of her beliefs at the time of action, what she ought to do by the lights of what she later establishes ..., what she ought to do by the lights of one or another onlooker who has different information on the subject, and, what is more, what she ought to do by God's lights, that is, by the lights of one who *knows* what will and would happen for each and every course of action ... I hereby stipulate that what I mean from here-on by 'ought,' and what I meant, and hope and expect you implicitly took me to mean when we were discussing the examples, was the ought most immediately relevant to action, the ought which I urged to be the primary business of ethical theory to deliver. When we act we must perforce use what is available to us at the time, not what may be available to us in the future or what is available to someone else, and least of all what is available to a God-like being who knows everything about what would, will and did happen. ... (Jackson 1991: 471–2)

Jackson's explanation of why we shouldn't be even-handed in our treatment of belief and desire is that we can't be even-handed in this way if we are trying to elucidate the 'ought most immediately relevant to action'.

This suggests the third possible reading of the 'ought' in the crucial sentence that appears in Jackson's account of how big 'C' consequentialism can be represented decision-theoretically. When he says 'we can think of consequentialism as saying that the desires a person ought to have are those which would be represented by a preference function which coincided with consequentialism's value function', perhaps he intends the 'ought' to be read as the 'ought most immediately relevant to action.' Now, as I understand it, the 'ought' that is most immediately relevant to action is the 'ought' that underwrites rational criticism of what agents do: it is, as I shall say, what they ought (critically) to do. When we act, our aim is to live up to our responsibilities as rational agents, thereby avoiding rational criticism. In order to do this we must exercise all of those rational capacities we possess whose exercise is required for rational action. For, so long as we do this, we do all that we can; and, since it cannot be supposed that we ought (critically) to do something that we cannot do, we do all that we ought (critically) to do.

I take it that this is why Jackson is so scathing about the idea of an 'ought' that is based on God-like knowledge. Since no one could have such knowledge, given that we are trying to elucidate what agents ought to do in the sense of 'ought' that is 'most immediately relevant to action'—that is, given that we are trying to elucidate what agents ought (critically) to do—it is irrelevant what agents would do if they had God-like knowledge. What is relevant is rather what they would do if they were to fully exercise such, admittedly limited, belief-forming capacities as they have. Let me illustrate the ways in which this idea can be developed with some examples.

Suppose an agent desires to walk into a room, and he has two beliefs about how he might achieve this. He can walk through door 1, which he is certain will get him into the room, or he can walk through door 2, which he is fairly confident will get him into the room, though he is not absolutely certain: he thinks it might just be a door to a cupboard. Given that the strength of agents' instrumental desires should be a function of the strength of their desires for ends and their level confidence about the means to their ends, it follows that this agent should have a stronger instrumental desire to open door 1 than to open door 2. In decision-theoretic terms, he will maximize expected utility by opening door 1. This is therefore what he would do if he fully exercised his capacity for instrumental rationality. In this sense, it is what he ought (critically) to do. But of course, for all that, the agent might not exercise these capacities fully. For example, he might put his desire to walk into the room together with his belief about door 2, but simply not put it together with his belief about door 1. If this is what he does then he will presumably have an instrumental desire to open door 2, and no instrumental desire to open door 1. What he will do is open door 2, and, in so doing, he will act intentionally, but irrationally. He will fail to do what he ought (critically) to do.

If this is the right way to understand the 'ought' that Jackson is trying to elucidate, then we can generalize. When an agent acts, and in so doing fully exercises all of the rational capacities he possesses whose exercise is required for rational action—that is to say, when, in acting, he does everything that he ought (critically) to do—then he will presumably exercise other rational capacities as well. For example, just as he will act on instrumental desires that elude criticism in terms of norms of instrumental rationality, he will act on means-ends beliefs that elude criticism in terms of epistemic norms. His means-end beliefs will be formed on the basis of sound reasoning from evidence, rather than on the basis of unsound reasoning, or whimsy, at least to the extent that he has the capacity to engage in such reasoning. The mere fact that an agent fully exercises such rational capacities as he possesses in the formation of his means-end beliefs will, of course, be no guarantee that his means-end beliefs are true, for he may lack certain crucial rational capacities, or his evidence may be misleading. Human agents are not God-like. All it guarantees is that, if the agent's means-end beliefs are false, then at least it will not be his fault. For in acting he will have done all that he can, and all, therefore, that he ought (critically) to do.

Moreover when, in acting, an agent does everything that he ought (critically) to do, he will also have desires for ends that elude rational criticism. Of course, radical Humeans deny that desires for ends are fit objects of rational criticism. But this certainly isn't the view of those, like Jackson, who defend the modest form of internalism which holds that judgements of value and desires for ends stand in the following rational relation (Jackson and Pettit 1995):

> MI: Reason requires that ∀x (If x judges that p is $good_x$, then x desires that p)

For, according to MI, those who possess the capacity to have desires for ends for the things that they judge to be good, but who fail to exercise that capacity, are also rationally criticizable (Smith 1994). They fail either to have the desires for ends that they ought (critically) to have, or they fail to make the value judgements that they ought (critically) to make. Someone who, in acting, does everything that he ought (critically) to do thus doesn't just act so as to realize their desires for ends, but also acts so as realize what he judges to be good.

Finally, as this example suggests, an agent who acts having done everything that he ought (critically) to do, will also be an agent whose value judgements elude rational criticism. That is to say, much like his means-end beliefs, his value judgements will be formed on the basis of sound reasoning from evidence, rather than on the basis of unsound reasoning, or whimsy, at least to the extent that he is capable of such reasoning. Again, as with his means-end beliefs, the mere fact that he fully exercises such rational capacities as he possesses in making his value judgements will be no guarantee that his value judgements are true. He may lack certain crucial capacities, or his evidence may be misleading. All it guarantees is that, if they are false, then it is not his fault. In acting he will have done all that he can, and all, therefore, that he ought (critically) to do.

To sum up, when agents act, they are liable for rational criticism if, in acting, they fail to exercise all of the rational capacities they have whose exercise is required for rational action. If we diagram the various elements potentially

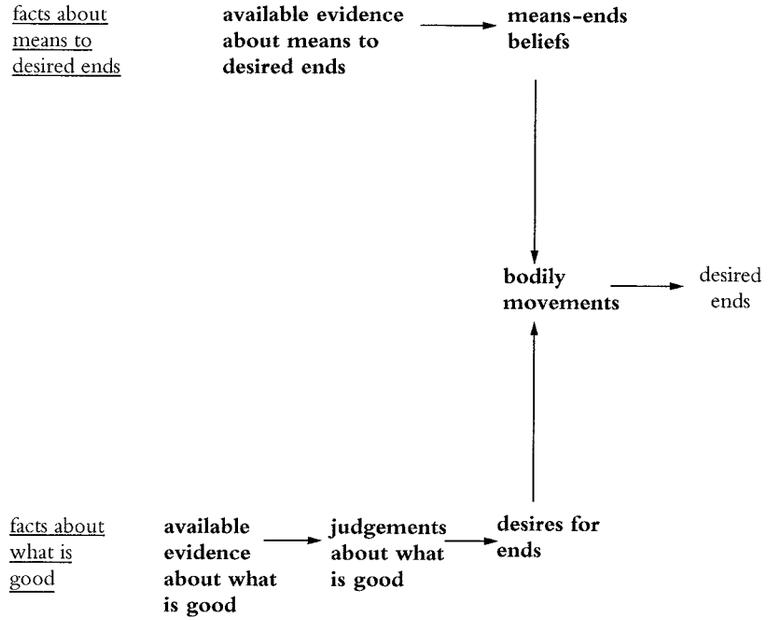

Figure 10.1. Rational Action

involved in rational action, then the places where these capacities can be exercised are represented by the arrows and plus signs relating the elements in bold (see also Smith 2004).

As rational agents, most of us have capacities of some sort to respond to available evidence in the formation of both our means-end beliefs and our value judgements; to respond to our value judgements in the formation of our desires for ends; and to respond to our desires for ends and means-end beliefs when we subsequently move our bodies. An agent who does everything that he ought (critically) to do when he acts is thus one who fully exercises these capacities. But of course, since an agent cannot guarantee that the evidence available to him isn't misleading, there is no reason to suppose that, merely by exercising his rational capacities, he can guarantee that his means-end beliefs or value judgements are true. Nor, for that same reason, can an agent guarantee that the ends he desires will result from his moving his body. There is thus no general presumption that agents ought (critically) to act in accordance with the true and the good, because their doing so is not, in general, under their rational control.

I have laboured this point because it shows the extent to which Jackson misunderstands what is required if his goal is indeed to elucidate the 'ought most immediately relevant to action'—that is, what we ought (critically) to do. True enough, we rationally criticize agents if they fail to act on appropriate means-end beliefs: that is, if they fail to exercise their capacity for instrumental rationality. What they ought (critically) to do is to act having exercised this capacity. But we equally rationally criticize agents if they act on means-end beliefs that are formed on the basis of sloppy reasoning, or whimsy. In other words, since expectations are rationally criticizable, so too are actions performed on the basis of expectations. What agents ought (critically) to do is to act having exercised their capacity for epistemic rationality. And we equally rationally criticize agents if they act on desires for ends that do not accord with their judgements of value, and agents who act on value judgements that are formed on the basis of sloppy reasoning or whimsy. Glaringly, however, we no more rationally criticize agents for failing to do what is in fact good than we rationally criticize them for failing to act on means-end beliefs that are in fact true.

This helps us understand what is really going on in Jackson's examples. Consider again Jill, faced with a choice between giving John drugs A, B, and C, but this time let's suppose that Jill disagrees with us about the relative value of the outcomes, and let's also suppose that this disagreement in no way reflects badly on her as a rational subject. Perhaps she was simply raised in a context in which all the evidence supported value judgements that we don't ourselves agree with. As Jackson tells the story, we are supposed to be horrified if she gives John drug B, and equally horrified if she gives John drug C, because in doing so she fails to maximize expected value. To maximize expected value she must give John drug A. But what his discussion of the example buries is the fact that, to the extent that our horror reflects our assessment of Jill as a rational agent—that is, to the extent that our horror is based on our belief that Jill fails to exercise such rational capacities as she has—the values we take to be relevant will have to be Jill's values, not our own. If, relative to her own assignment of values, giving John drug B, or drug C, maximizes expected value then we would rationally criticize her for failing to give John either drug B, or drug C. This is what she ought (critically) to do.

Conversely, if we suppose that in assessing Jill's conduct we are allowed to replace Jill's values with some other system of values—suppose, for example, that in the spirit of Jackson's preferred decision-theoretic formulation of big 'C' consequentialism, we replace her values with what big 'C' consequentialism

tells us to be good—then, when we assess Jill's conduct, we are plainly no longer interested in assessing her as a rational agent. The 'ought' claims we make are not claims about what Jill ought (critically) to do; they are not 'ought' claims that are most immediately relevant to action. They are 'ought' claims of the other kind, claims about what Jill ought (formally) to do.

The upshot is that, to the extent that we feel horrified merely by Jill's failure to act in accordance with what big 'C' consequentialism deems to be really good—to the extent that we find something to lament about that, independently of whether her own judgements of value accord with big 'C' consequentialism, and, if they do, whether this can be traced to sloppy reasoning or whimsy on her behalf—so, by parity of reasoning, we should feel equally horrified by her failure to act on means-end beliefs that are true. In assessing whether she ought to give drug A, or drug B, or drug C, we therefore have no choice but to replace her desires by desires for what big 'C' consequentialism deems to be good, and her means-end beliefs with beliefs about what really are means to the ends that are good. We have no choice but to adopt a God-like point of view. But if we do this then, of course, we reach the conclusion that Jill ought to give John that drug, whichever it is, that in fact effects a complete cure: drug B, or drug C, whichever it is. In other words, we reach the conclusion that what Jill ought to do is maximize neutral value.

To sum up, there are two salient ways in which we can represent what an agent ought (morally) to do in terms of what they ought to do according to a variant on decision theory. According to one, what we are interested in is what the agent ought (critically) to do. This is the 'ought' that is most immediately relevant to action. What we do in this case is ask what the agent ought to do according to the variant on decision theory in which she has the desires and beliefs that she would have if she fully exercised her rational capacities. Such an agent will maximize expected value, but the beliefs and values in question will both be closely related to her own. There will be no general presumption that such an agent will value what is good, or believe what is true. According to the other, what we are interested in is what the agent ought (formally) to do. What we do in this case is ask what the agent ought to do according to the variant on decision theory in which her desires and beliefs are both replaced by desires and beliefs that meet their formal aim. The agent desires what is good and believes what is true. Such an agent will therefore simply maximize value. But it is hard to see what 'ought' we would be elucidating by lacking even-handedness in the way Jackson suggests, that is, by asking what the agent ought to do according to that variant on decision theory in which an agent has his actual beliefs, but his desires are replaced by desires for what is good. It therefore seems to me that Jackson fails to give a convincing argument for the expected neutral value formulation of big 'C' consequentialism. We should suppose that big 'C' consequentialism tells agents to maximize neutral value simpliciter, not maximize expected neutral value.

6. The epistemological version of the nearest and dearest objection

In this final section, I would like to return to consider another version of the nearest and dearest objection, a version I mentioned at the beginning, but put to one side. The objection I have in mind is epistemological (Smith 2001). It is the version of the objection that Bloggs has in mind when he gives the first of the responses we considered to our suggestion that he should give the $100 to the stranger.

According to John Rawls, we all come to moral philosophy with various firmly held convictions about what is value and what is not, and about which actions are right and which are not, and what we want to know is whether these convictions are justified (Rawls 1951). Rawls's famous suggestion is that there is a common-sense procedure by which we test these convictions. This is the reflective equilibrium procedure. We test our convictions by trying, as best we can, to bring them into some sort of system. Justification is a matter of surviving in such a systematization. More precisely, his suggestion is that we test our convictions by first coming up with a hypothesis about what is of fundamental value, or what the fundamental principles are. In doing this, we invoke the standard criteria for constructing theories. We formulate a hypothesis that is simple, elegant, powerful, and adequate to the task of explaining why the convictions with which we began are true. A moral theory is, essentially, just such a hypothesis.

Sometimes, when we formulate such a hypothesis we find that it is sufficiently simple, elegant, powerful and adequate to the task of explaining why some subset of our firmly held convictions are true that it gives us the confidence to reject, as mistaken or misguided, those convictions with which we began with which it is inconsistent. But sometimes the reverse is true. A hypothesis might have the wrong mix of theoretical virtues—it might be simple, say, but insufficiently adequate to the task of explaining our firmly held convictions—and so not inspire the confidence required to get us to reject the convictions with which it is inconsistent. In the former case we

reject our convictions as unjustified. In the latter case we reject our hypothesis as incredible.

As I understand it, there is an epistemological version of the nearest and dearest objection to big 'C' consequentialism, a version that is best understood in terms of this Rawlsian reflective equilibrium procedure. Big 'C' consequentialism is a hypothesis that is supposed to give system to our various firmly held convictions about what is of fundamental value and what the fundamental principles are. According to big 'C' consequentialism, stating the fundamental principles is easy, as there is only one such principle, and it is analytic: we ought to act so as to maximize value. The hard part, the part where big 'C' consequentialism really does amount to a substantive hypothesis, is to give an account of the values that are to be maximized. Though big 'C' consequentialism is consistent with a range of possibilities in this regard, it places one significant global constraint on the form of all the values: values are one and all neutral in form, which is to say that they are expressed in principles like N, rather than in principles like R.

Big 'C' consequentialism's hypothesis that all values are neutral is very simple. But what the nearest and dearest objection brings out is that the simplicity of the hypothesis doesn't inspire sufficient confidence in us to reject our firmly held conviction that some values are relative. We hear a story like Bloggs's and we are supposed to find ourselves dismissing his attempt to give a justification of his conduct, but we can't dismiss it. We share Bloggs's conviction that there are relative values at stake and that they can be realized by his giving the $100 to his child. Our confidence in the claim that there are relative values is thus greater than our confidence in big 'C' consequentialism's hypothesis that all values are neutral, notwithstanding the fact that that hypothesis purchases an abundance of the theoretical virtue of simplicity. In this way what the nearest and dearest objection brings out is that big 'C' consequentialism is simply incredible. This, as I understand it, is the epistemological version of the nearest and dearest objection.

Right at the very end of his paper Jackson briefly considers, and rejects, the nearest and dearest objection in its more epistemological cast.

> One objection to consequentialism is that it conflicts with firmly held moral convictions, in particular concerning our obligations toward our nearest and dearest. It may be urged that my reply to this objection is seriously incomplete. For we can reasonably easily describe a possible case where the factors I mentioned as providing a justification in consequentialist terms for favouring one's nearest and dearest do not apply, and yet, according to commonsense morality, one should favour, or at the least it is permissible to favour, one's nearest and dearest. My concern, though, has been to reply to the objection that consequentialism would, given the way things more or less are, render the morally good life not worth living. I take this to be the really disturbing aspect of the nearest and dearest objection. Consequentialists can perhaps live with the conflict with commonsense morality, drawing for instance on the notorious difficulties attending giving a rationale for its central features. But it seems to me that they cannot live with the conflict with a life worth living, given the way things more or less are. That would be to invite the challenge that their conception of what ought to be done had lost touch with *human* morality. (Jackson 1991: 482)

Jackson thus doesn't seem to be much impressed with the epistemological version of the nearest and dearest objection. But though only mentioned in passing, it is important to note that Jackson's response to the epistemological version of the nearest and dearest objection is less than convincing.

Jackson tells us that there are 'notorious difficulties attending giving a rationale' for the 'central features' of a non-big 'C' consequentialist hypothesis. He provides a hint as to nature of these 'notorious difficulties' by footnoting Shelley Kagan's *The Limits of Morality* (1989). But what Kagan argues, and argues decisively, is that it is difficult to make sense of absolute constraints on behaviour, constraints of the kind that are supposed to be required by deontological restrictions. In small 'c' consequentialist terms, this requires not just the hypothesis that some values are relative—for example, the restriction on my harming people requires the disvalue of the outcome of someone's being harmed by me, which does indeed require formulation in a principle of the same form as R—but also that this outcome has *infinite* disvalue. Kagan's argument is, in essence, that such infinite disvalues cannot be combined in any plausible way with uncertainty about their realization. But conceding that Kagan is right about this, as it seems to me we should, doesn't tell in favour of the big 'C' consequentialism's hypothesis that all value is neutral. There is an alternative hypothesis in between these two, namely, the hypothesis that in addition to neutral values there are relative values, but that all such values are *finite*.

As I understand it, this is all that the nearest and dearest objection assumes. The objection is not that the relative value that attaches to a benefit to Bloggs's child, in virtue of its being a benefit to his child, is such as to place an absolute prohibition on his bringing about neutral value. The objection is rather that, though in circumstances in which the gain in relative value is small enough and the loss of neutral value large enough, the gain in relative value would not outweigh the loss of neutral value, in circumstances like those described in Bloggs's story where the gain in relative value is large and the loss of neutral value small, the gain in relative value is great enough to

outweigh the loss in neutral value. This is not the hypothesis that the relative values that attach to the weal and woe of our nearest and dearest are such as to give rise to deontological restrictions. But it is still a hypothesis that is inconsistent with the big 'C' consequentialist hypothesis that all value is neutral.

It is perhaps worth adding that there is supposedly another, and much more well-known, difficulty with giving a rationale for the central features of a non-big 'C' consequentialist hypothesis. The difficulty is that brought out by G. E. Moore in his famous argument against ethical egoism (Moore 1903). Moore argues that since the concept of value is not an indexed concept—in other words, since there is no such thing as $\text{goodness}_{\text{Bloggs}}$, but only goodness (unsubscripted)—we must reformulate both N and R. Principles that ascribe agent neutral value, principles like N, must all be reformulated using an unindexed concept of goodness, and principles that purport to ascribe relative value, principles like R, must all be abandoned. They can be given no coherent formulation at all once we reject the assumption that the concept of value is an indexed concept. But if the concept of relative value is, quite literally, incoherent, then that provides us with a decisive reason to reject the epistemological version of the nearest and dearest objection.

As I said, Moore's argument turns on whether, according to the best analysis of the concept of value, that concept is indexed or unindexed. Unsurprisingly, given that he takes the concept of value to be simple and unanalysable, Moore thinks that the concept is unindexed. But these days no one takes seriously the idea of there being a metaphysically simple property of goodness. Much the most widely held theory of value is rather some version of the dispositional theory (Smith 1994). According to the dispositional theory, to say that p is good is to say (roughly speaking) that we would desire that p if we had an ideal desire set. However, if something along these lines is the right analysis of the concept of value then it turns out that that concept is indexed after all (Smith 2002). Goodness is indexed to that group of individuals whose ideal desires are the truth-makers of evaluative claims: all goodness is $\text{goodness}_{\text{that group of individuals}}$. Contrary to Moore, it is therefore at least coherent to suppose that there are both neutral values and relative values (Smith 2003). For if we had an ideal desire set then it is possible we would desire that all people benefit, independently of whether or not they bear any special relationship to us, and it is also possible that we would have an independent desire that our own children benefit. In this way we can see that it is at least coherent to suppose that both N and R are true.

The upshot is that Jackson's response to the epistemological version of the nearest and dearest objection, the response that follows Kagan, misses its mark, and so too does Moore's more well-known response to the epistemological version of the objection. The epistemological version of the nearest and dearest objection is therefore still very much on the table. This is not, of course, a serious objection to any of Jackson's principle concerns in 'Decision Theoretic Consequentialism and the Nearest and Dearest Objection'. But it does mean that, at the end of the day, we are yet to be told why we shouldn't dismiss big 'C' consequentialism on the grounds that it is simply incredible.

Princeton University

References

Adams, R. 1976. 'Motive utilitarianism'. In *Journal of Philosophy* 73: 467–81.

Brink, D. 1989. *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).

Hare, R. 1981. *Moral Thinking* (Oxford: Oxford University Press).

Jackson, F. 1991. 'Decision theoretic consequentialism and the nearest and dearest objection', *Ethics* 101: 461–82.

Jackson, F. and Pettit, P. 1995. 'Moral functionalism and moral motivation', *Philosophical Quarterly* 45: 20–40.

Kagan, S. 1989. *The Limits of Morality* (Oxford: Clarendon Press).

Mill, J. S. 1861: *Utilitarianism* (London: Fontana Library, 1962).

Moore, G. E. 1903: *Principia Ethica* (Cambridge: Cambridge University Press).

Parfit, D. 1984: *Reasons and Persons* (Oxford: Oxford University Press).

Pettit, Philip and Michael Smith 2000. 'Global consequentialism'. In B. Hooker, E. Mason and D. Miller (eds.), *Morality, Rules, and Consequences: A Critical Reader* (Edinburgh: Edinburgh University Press).

Railton, P. 1984. 'Alienation, consequentialism, and the demands of morality'. Reprinted in S. l. Scheffler (ed.), *Consequentialism and its Critics* (Oxford: Oxford University Press, 1988).

——. 1988. 'How thinking about character and utilitarianism might lead to rethinking the character of utilitarianism'. In P. French, T. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy: Volume XIII. Ethical Theory: Character and Virtue* (Notre Dame: University of Notre Dame Press).

Rawls, J. 1951. 'Outline of a decision procedure for ethics'. *Philosophical Review* 42: 177–97.

Sidgwick, H. 1907. *The Methods of Ethics* (London: Macmillan).

Smith, M. 1994: *The Moral Problem* (Oxford: Basil Blackwell).
——2001. 'Immodest consequentialism and character', *Utilitas* (Special Issue on Consequentialism and Character) 13: 173–94.
——2002. 'Exploring the implications of the dispositional theory of value', *Philosophical Issues: Realism and Relativism* 12: 329–47.
——2003. 'Neutral and relative value after Moore', *Ethics* (Centenary Symposium on G. E. Moore's *Principia Ethica*) 113: 576–98.
——2004. 'The structure of orthonomy'. *Philosophy* 55: 165–93.

11

The 'Actual' in Actualism[1]

JULIA DRIVER

The work of Frank Jackson has been important to at least two central debates in consequentialist ethical theory—those are the debates (1) between possibilism and actualism and (2) between objective consequentialism and expectabilism (or a variety of subjective consequentialism).[2] Suppose that we define the right action as that action which maximizes the good. Some writers, such as Michael Slote, have argued that this straightforward criterion is underdetermined. (See Slote 1992: 239–48.) Are we to maximize 'actual' good or 'expected' good? That is the debate between the objective consequentialist and the subjective consequentialist (with a caveat to be discussed later). There is also the issue of whether or not the agent who is deliberating considers what would be best given what will happen as opposed to what could or can happen. That is the debate between the actualist and the possibilist in determining relevant options for the moral agent to consider in deliberation. This essay explores differing answers to both of these questions, and then explores one strategy for answering both—a strategy which has been very much influenced by the work of Frank Jackson. Though I agree with Jackson on actualism I will disagree with him on expectabilism. My claim is that the definition of 'right action' is clear and not at all underdetermined—the right action just is the action that maximizes the good, the actual good. But what is often confusing is that the semantics of right is confused with an issue in the epistemology of right—that is the issue of determining how we are to go about doing the best that we can properly.

[1] I thank Roger Crisp, Mike Ridge, Walter Sinnott-Armstrong, and Roy Sorensen for helpful comments on an earlier draft of this paper.

[2] It's worth pointing out, however, that the actualism/possibilism debate has significance for moral theory far beyond consequentialism. Any theory which holds that weighing options is crucial to moral decision-making will need to confront the issue.