



The Limits of Human Nature

Author(s): Keith Horton

Source: *The Philosophical Quarterly*, Vol. 49, No. 197, (Oct., 1999), pp. 452-470

Published by: Blackwell Publishing for The Philosophical Quarterly

Stable URL: <http://www.jstor.org/stable/2660496>

Accessed: 01/07/2008 23:46

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=black>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## THE LIMITS OF HUMAN NATURE

BY KEITH HORTON

### I. INTRODUCTION

It has become increasingly common recently, both in philosophical circles and in the wider culture, to construe human nature as setting some pretty stringent limits to moral endeavour. In making such claims, the following two forms of argument are often pressed into service. First, certain reasons are given for entertaining a norm of the following form:

N. Agents in such and such circumstances should  $\phi$ .

Then certain considerations concerning human nature are invoked to ground either the claim that

1. It is impossible for such agents to  $\phi$

or that

2. It would be counter-productive in moral terms for such agents to  $\phi$ .

If either (1) or (2) is true, (N) is taken to be defeated. I shall not question the validity of either of these forms of argument, but I shall argue against some influential recent employments of them which, in my opinion, seriously underestimate our capacities for rigorous moral endeavour.

In arguments of these forms, considerations concerning human nature function as a *de facto* constraint on moral norms. It is not that those considerations ground or constitute normative factors which are to be weighed against whichever normative considerations support the norm in question. They are, rather, just so many matters of fact that have to be taken into account – perhaps with regret, if we think that things would in some sense be better if human nature was not like this. This is not, of course, the only

role that considerations concerning human nature can play in a moral theory. Some philosophers think that such considerations can be used to *ground* normative propositions. I shall have nothing to say about such views here. And even if we restrict ourselves to forms of argument in which considerations concerning human nature function as a *de facto* constraint, there are other ways in which they might do so besides (1) and (2). For example, it might be true that

3. It would be highly demanding for such agents to  $\phi$ .

The question of whether (3) would defeat (N) is much more controversial than the question of whether (1) or (2) would. There are, of course, connections between (3) and (1) or (2). At some point, presumably, demandingness shades into impossibility. And following a certain norm might, in some cases, be counter-productive *because* doing so would be highly demanding. But in such a case it is (2) that would be decisive. Whether (3) or other similar claims would provide an *independent* reason to reject (N) is a question that I shall not discuss here.

There are a number of reasons why arguments in which considerations concerning human nature function as a *de facto* constraint have become increasingly prevalent in contemporary moral philosophy. One is the thought that any moral theory must take human nature into account in some way, that it would be absurd to develop a moral theory in sublime indifference to such considerations, combined with a suspicion about the idea that such considerations might be relevant to grounding moral norms. If such considerations are not to ground moral norms, then it may not be clear what work they are to do, if not to constrain them. Another reason is provided by recent developments in evolutionary psychology, which some philosophers take as uncovering constraints that putative moral norms have to meet. And these developments may reinforce the suspicion about human nature as the ground of moral norms: if the particular nature we have is just the contingent result of certain evolutionary pressures, it is far from obvious why it should be given any kind of normative status.

But perhaps the most important reason for the increasing prevalence of such arguments is a concern internal to moral thinking itself about the apparently unlimited demands that may be placed on agents by certain moral theories. This concern is particularly acute in the case of individualistic forms of consequentialism, that is, forms of consequentialism that tell *each* of us to do what would have the best consequences, given (reasonable expectations about) what others will actually do.<sup>1</sup> To take a much discussed

<sup>1</sup> See D. Parfit, *Reasons and Persons* (Oxford UP, 1984), p. 30. In this paper I shall be exclusively concerned with individualistic forms of consequentialism.

example, it might appear that, given the empirical circumstances in the world, consequentialism yields the following norm:

N1. Relatively well-off agents should give at least half of their incomes to international aid agencies.

But arguments using forms of (1) and (2) are widely taken to defeat (N1). Some philosophers take this as a reason to reject consequentialism; others respond by attempting to develop forms of consequentialism that accommodate these arguments. After all, they say, consequentialism does not ask agents to do what they cannot do or what would be counter-productive in consequentialist terms for them to do. In fact, many who take this line argue that, when the limitations of human nature are taken fully into account, consequentialism yields something pretty close to common-sense morality.<sup>2</sup>

Sometimes the claim is that it would be impossible or counter-productive for relatively well-off agents in general to act on (N1), even supposing that consequentialists somehow had their fingers on powerful instruments of social control. Thus some philosophers discuss the merits and demerits of using techniques of behaviour modification to get people to act more benevolently.<sup>3</sup> Also relevant in this context is the frequently made claim that publicly advocating a high standard would itself be counter-productive in consequentialist terms.<sup>4</sup> At other times, the focus is again on such agents in general, but the question of social control is bracketed. Frequently, however, the representative case is taken to be that of a sincere, explicitly consequentialist agent: even someone who is really committed to living in a way that will have the best consequences for all should not, it is claimed, try to act on (N1). To do so would be to try to go beyond the limits of human nature, and such hubris would invite ruin – or, at least, would be counter-productive in consequentialist terms.

This strikes me as rather extraordinary. It is a familiar and reasonable idea that human nature sets *some* kind of limits to the kinds of lives that human beings can lead, but that it should be so significant as to make it better for a sincere consequentialist agent, in the empirical conditions of the world today, to stick close to common-sense morality, rather than acting on (N1), is quite another thing. At the same time, it is a difficult claim to argue about with any determinacy, in part because we lack the kinds of hard data

<sup>2</sup> Prominent examples of consequentialists who take this sort of line include Richard Hare, Derek Parfit, Peter Railton, Frank Jackson and Roger Crisp. I discuss some of their arguments in more detail in §§II and III below.

<sup>3</sup> See, for example, J.L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977), p. 133; J. Griffin, *Value Judgement: Improving our Ethical Beliefs* (Oxford UP, 1996), pp. 88–9.

<sup>4</sup> See Mackie pp. 133–4; B.A.O. Williams, *Ethics and the Limits of Philosophy* (Harvard UP, 1985), p. 212 fn. 7; R. Crisp, *Mill on Utilitarianism* (London: Routledge, 1997), p. 125.

concerning human nature which, it might seem, are necessary to settle the issue. Without such hard data, it may appear that we are reduced to relying on 'hunches' about human nature.<sup>5</sup> But I hope to be able to put some pressure on the claims of those who are pessimistic about our capacities for rigorous moral endeavour by highlighting factors that they have neglected, and also by suggesting some reasons for this neglect.

It is not only consequentialists who construe human nature as setting some fairly stringent limits to moral endeavour. One can see why consequentialists should be especially anxious to prosecute this claim: without it, consequentialism is likely to become hugely demanding. But many non-consequentialists take this claim as *one reason among others* to reject norms like (N<sub>1</sub>). If what I say is correct, then, though I focus throughout on consequentialism, what I say will have a bearing on them too. In that case, there had better be good *normative* reasons for rejecting norms like (N<sub>1</sub>), rather than reasons based on the limits of human nature. In fact I think that non-consequentialists too should be worried about norms like (N<sub>1</sub>), though I shall not try to substantiate this claim here.

## II. IMPOSSIBLE ACTIONS

Given that it is supposed to be impossible or counter-productive even for a well intentioned consequentialist to follow (N<sub>1</sub>), I shall focus on the case of such an agent. (I shall say something about what people in general are capable of, and of what 'we' can get them to do, in §IV.) Anna is in her twenties, has a fairly well paid job, and lives with her boyfriend Steve, whom she plans to marry. She does not try to subject every decision to consequentialist calculation, but her consequentialism is not, on the other hand, completely self-effacing. Every now and then she takes a step back from her everyday concerns and asks whether her current way of life is optimal in consequentialist terms. This is one of those occasions. At present she gives 5 per cent of her income to aid agencies, which is already very generous by the standards of common-sense morality. But now she is wondering whether she should follow (N<sub>1</sub>). This would lead to a radical change in her way of life. Perhaps she feels it would be best to keep her job, in order to have lots of money to give, but every other aspect of her life would be deeply affected: her ordinary day-to-day pleasures and comforts, her projects, and, no doubt, her close personal relationships too.

<sup>5</sup> The term 'hunches' is used by Griffin, whose abandonment of utilitarianism seems to be due in large part to the fact that, as he puts it (p. 158 fn. 13), 'my hunch about human nature has changed'.

The most commonly made claim is that such a change would, in the long run, be counter-productive in consequentialist terms. I shall discuss this in the next section. But it will be helpful to begin with another, stronger claim:

1a. It is impossible for Anna even to start acting on (N<sub>1</sub>).

To start acting on (N<sub>1</sub>) would be to perform certain actions: explaining to those close to her what she is doing, reorganizing her affairs, signing a direct debit form, and so on. Why might this be impossible?

Well, one kind of claim about human nature that might support (1a) is that, once formed, our dispositions have a certain *inflexibility*. As part of a normal upbringing, one's motivational structure comes to embody a certain partiality, and it may be that, as James Griffin puts it (p. 85),

That partiality then becomes part of one; it is not something that one can psychologically enter into and exit from at will. It involves becoming a certain sort of person.

If 'at will' means 'from one moment to the next', then what Griffin says is certainly true. It is simply not in our repertoire to stand aside from our most firmly entrenched dispositions, on each occasion of choice, and do what we feel would be best impartially. And it may well be that certain kinds of dispositions, once formed, cannot be altered. If, for example, because of an unfortunate childhood, Anna has grown up to be distrustful of other people, there may be little she can do to change this.

But this is, for Anna, no ordinary moment of choice. This is one of those occasions when she is asking whether her everyday dispositions are optimal in consequentialist terms. This attitude seems to give her at least a partial distance from those everyday dispositions. And the policy she is considering, that of following (N<sub>1</sub>), is quite unlike a policy of being more trustful of others. There may be nothing she can do right now to start being more trustful of others, but she can make a start on following (N<sub>1</sub>) simply by signing a direct debit form. It is, in part, the fact that signing a direct debit form appears, in one sense, to be such an easy thing to do that makes (1a) seem implausible.

But perhaps this appearance is misleading. Perhaps if Anna reflects fully on what acting on (N<sub>1</sub>) will mean for her, she will find that she is incapable of doing it. This might be particularly plausible if acting on (N<sub>1</sub>) would force her to give up the projects and relationships that mean most to her. It might not be immediately obvious why this might be so. If, for example, she is British or American and earns an average income, then giving away half of that income will reduce her to a level at which millions of her compatriots already live, and many of them seem to have satisfying personal projects and relationships. But perhaps it may still be the case that *she* could not continue

to pursue the projects and relationships to which *she* is committed if she gave away half of her income. It is difficult to steer a sensible course here between overestimating the significance, for an agent's projects and relationships, of mere money, on the one hand, and the kind of excessive high-mindedness that talk of 'mere' money might presuppose, on the other; and much will depend on the particular circumstances of the individual case. But let us assume, for the purposes of argument, that Anna's most important commitment is to her partner Steve, and that giving away half of her income would rupture that relationship. If Steve and Anna pool their financial resources, Anna's giving away half of her income would have significant effects on his life: they would have to move into a cheaper flat, have less expensive holidays, and so on. Perhaps Steve would resent this. Perhaps he would construe it as an attack on him. Or perhaps he would view Anna's decision with the same kind of bewilderment as that with which he might have viewed a religious conversion on her part, with the sense that the person he had fallen in love with had become a different person.

If this is true, then another claim about human nature may be invoked to support (1a). Here the thought is not merely that dispositions in general tend to be relatively inflexible, but that certain commitments in particular ground *psychological incapacities*. When Anna reflects that radically changing her way of life would mean leaving Steve, she may find that she *cannot* do it.

There certainly are such psychological incapacities, but the issue needs careful handling. The first thing to say is that if Anna deliberates about whether to give away most of her income, she must, at least initially, take it to be something that she can do. This just follows from the nature of deliberation: we do not deliberate about things that we believe we cannot do.<sup>6</sup> So if Anna is able to undertake, at least on occasion, the kind of consequentialist deliberation that questions her whole way of life, it looks as if she must treat giving away half or more of her income as something she can do.

But perhaps she is, in fact, incapable of deliberating seriously about whether to give away most of her income, given her belief that doing so will end her relationship with Steve. Or perhaps she is capable of beginning a deliberation about this, but then she *discovers* in deliberation that she cannot do it. Such a discovery might foreclose deliberation: having discovered this, the option of giving away most of her income might no longer seem a real possibility, and deliberation would come to an end.

Let us take this second possibility first. What Anna discovers in deliberation is something about, or something that follows from, the present state of

<sup>6</sup> Cf. Aristotle, *Nicomachean Ethics* III iii, 1112a 18–33.

certain of her dispositions. But the present state of those dispositions is, from the perspective of her consequentialist deliberation, the very thing in question. She is asking whether those dispositions are, in consequentialist terms, optimal ones. *Ex hypothesi* the answer is negative. So it looks as though Anna cannot take such a discovery as authoritative. She can say to herself 'Given my dispositions just as they are, I cannot give away most of my income. But those dispositions are not optimal: I will act against them.'

It is difficult to know what to say about this kind of deliberation in which our agent treats certain of her own dispositions as one further thing to take into account when deciding what to do. Normally, of course, deliberation is not like this. Normally one's deliberations do not *take into account* one's dispositions: they are simply an *expression* of one's dispositions. Perhaps this deliberation which takes into account certain of one's dispositions is itself an expression of a further disposition. Or perhaps it is an exercise of practical reason, in a sense in which such an exercise is not construed merely as the expression of a further disposition. Fortunately we do not need to try to resolve these difficult issues here. All we need is the fact that, from Anna's point of view, the discovery that certain of her dispositions render her incapable of giving away half of her income need not be taken as authoritative. In the absence of the kind of deliberation which treats those dispositions as just one further thing to take into account, that might indeed have been the case. But such a discovery need not exhaust her deliberative resources: she can say to herself 'If I had not conducted this very deliberation, which treats those dispositions as one further thing to take into account, I would have been incapable of giving away half of my income. But when I conduct this further deliberation, I find room to act against those dispositions.'

It would certainly be very hard for Anna to act 'in spite of herself'. Whether she manages to do so will depend on the motivational capacities she has at her disposal, and this in turn will depend on the *moral sources* available to her: those considerations which convey a sense of why it is fundamentally *important* to do what one believes one morally ought to do.<sup>7</sup> In the ordinary case, of course, the mere belief that one ought to do something may be enough to motivate one to do it. But we are considering rather a special case. Given an apparent incapacity, it may be necessary to reflect deeply and seriously on why the course of action apparently ruled out is called for. Such reflection may empower agents to do what they would not have been able to do had they merely had the bare thought that they should do something, divorced from any sense of why it really matters that they do that thing.

<sup>7</sup> I borrow the term 'moral sources' from Charles Taylor, *Sources of the Self* (Cambridge UP, 1989); see esp. ch. 4.

The issue of moral sources is even more important for the question of whether one is able to *sustain* one's motivation over a long period of time than it is for the question of what one is able to do in one fateful moment of inspiration or self-control, and so I shall say more about it in the next section. But for now, I think one can see that it may also make a difference to the question broached above of whether Anna might be able even to *begin* a serious deliberation about whether to change her way of life radically. The option of signing a direct debit form giving away half or more of her income might appear to her in a kind of thin, fantastical way, in the same kind of way as one might say to oneself, for example, 'Well, I could just leave everything today and get the next plane to Nepal to become a Buddhist monk'. There is a gap between such idle fantasizing and serious deliberation, but that gap might be closed, in Anna's case, by reflection on whatever moral sources are available to her, on the kind of considerations, such as, for example, the reality of the suffering at stake, that might make vivid her sense that this is not the stuff of fantasy, but something of over-riding importance.

We may compare in the light of this what Derek Parfit (pp. 31–7) says about another example. Clare is a consequentialist, but she has allowed herself to develop dispositions that sometimes lead her to perform acts which she believes to be wrong, acts whose outcome is less good than the available alternatives. The consequentialist justification for this is that she reasonably believes that her having such dispositions will lead to better outcomes overall than her having the disposition of a 'pure do-gooder'. In Parfit's first case (p. 32), she must decide whether to 'give her child some benefit, or give much greater benefits to some unfortunate stranger. Because she loves her child, she benefits him rather than the stranger.' She believes this to be wrong. In her defence she says that she acts wrongly because she loves her child, and that it would be wrong for her to cause herself to lose this love. To the objection that benefiting her child was still a voluntary act, she replies 'I could have acted differently. But this only means I *would* have done so if my motives had been different.'

If, like Clare at this point, we assume psychological determinism, this is correct. But, first, it is not how things could appear to her in the context of deliberation. If she deliberates about what to do, then, setting aside the points above about psychological incapacity, it must appear to her that she can either benefit her child or the stranger, and that it is up to her which to choose. And second, even a psychological determinist need not say that her motives are fixed *independently* of deliberation. She might discover an apparent psychological incapacity, but the process of deliberation might remove that incapacity, in the kind of way sketched above. She might say to herself 'I know I have strong motives to favour my child, and unless I take

care those motives will lead me to do the wrong thing. But I will make myself do the right thing.' What would have been impossible for her might become possible through her conducting this very deliberation. This is consistent with the claim that her motives determine what she does: a certain kind of deliberation might *affect* her motives.

In Parfit's second case, Clare has to choose between saving her child's life and saving the lives of several strangers. She saves her child, which she believes is the wrong thing to do. Parfit appeals here (p. 34) to a claim about human nature:

We could imagine other motives that would have made the outcome even better. But such motives are not causally possible, given the facts of human nature. Since Clare loves her child, she would have saved him rather than several strangers. We could imagine that our love for our children would 'switch off' whenever others' lives are at stake.... But it is in fact impossible that our love could be like this. We could not bring about such 'fine-tuning'.

I agree that we could not 'switch off' our love in such cases. But that is not necessary: what is necessary is that Clare should act on another motive she has – to make the outcome better in consequentialist terms. No doubt that would be excruciatingly hard. Perhaps for some agents it really would be impossible: a lot would depend on the moral sources sustaining the agent's commitment to consequentialism. But it is at least clear that, *qua* consequentialist, she should *try* to make herself sacrifice her child, employing whatever deliberative resources are available to her to try to inspire or coerce herself to do what she believes she should do. (This is to assume, as Parfit does, that sacrificing her child would make outcomes better, all things considered, including any effects on her character and thus on her later actions.) What she is capable of doing cannot be determined *independently* of such efforts, even if psychological determinism is true. And Clare should also be aware of the possibility that premature capitulation to an apparent incapacity may just be an expression of what she simply prefers to do, or what on non-consequentialist grounds she thinks it is acceptable to do.

Parfit concedes that, if psychological determinism is false, Clare can sacrifice her child. But he says, first (p. 33), that it would still be very difficult for Clare to do this, and, second (p. 36), that she could not '*always* act like a pure do-gooder *without* having a pure do-gooder's disposition'. The first claim is obviously true, but is a version of (3) (see p. 453 above), rather than of (1) or (2). The second claim is probably also true, so it may still be true that consequentialists should not be pure do-gooders. But allowing themselves to develop non-do-gooding dispositions need not altogether remove their capacity to act against those dispositions on *certain* occasions, when the stakes are very high, and they have time to take a step back and reflect.

Parfit's second case involving Clare is about as hard a case as one could conceive. The case of Anna is less extreme, but I do not mean to underestimate the difficulties involved here either. It *would* be very hard for her to change radically her way of life, particularly if this involved leaving Steve, but I hope to have shown that we should take great care before granting that it is *impossible* for her to do this. In deliberation and reflection, she might be able to harness powerful resources of motivation, and I think that it is largely due to the neglect of such resources that facile assumptions about the limits of human capacity are made. This is particularly so for the question of whether it is possible to *sustain* a commitment to demanding norms like (N<sub>1</sub>), a question to which I now turn.

### III. COUNTER-PRODUCTIVE POLICIES

Even if (1a) is false, it might still be true that

1b. It is impossible for Anna to *sustain* her commitment to (N<sub>1</sub>).

If (1b) is true, it might also be true that

2a. It would be counter-productive in moral terms for Anna to change her way of life radically.

Why might (1b) be true? The argument here once more starts from the claim that acting on (N<sub>1</sub>) would be likely to undermine an agent's personal projects and relationships. I have already suggested in §II that we should not be too quick to endorse this claim, but let us grant it once more for the sake of argument in Anna's case. The next move is to invoke another claim about human nature, which I shall call the *demoralization thesis*:

DT. Undermining one's personal projects or relationships is likely to lead to some kind of psychological degeneration or demoralization – to (in different versions) apathy, cynicism or selfishness.

And then the next claim is that once agents are reduced to this state they will not be motivated to do *anything* for others, and so (1b) will be true.

If this is so, then (2a) might also be true – Anna's contribution to the good, through the course of her whole life, might be less than if she had kept to a more modest level of giving (though given the empirical circumstances of the world, the envisaged decline may have to be pretty rapid for the sums to come out right). But what reason is there to think that the demoralization thesis (DT) is true?

Sometimes (DT) appears to be based merely on a kind of everyday observation of human life. There seems to me to be nothing wrong in principle

with such a source of claims about human nature – it may be that, at present at least, such observation is the best we have to go on – though there are, of course, dangers connected with it. For one thing, there may be a tendency to seek out data that confirm what we already believe, or that serve our interests, and to ignore conflicting data. And there may also be the temptation to construe local patterns particular to our culture as universal regularities in human nature.

Sometimes, however, the demoralization thesis is supported by considerations drawn from evolutionary psychology, as by Griffin:

Natural selection has made us intensely self-concerned and concerned for a few others, especially offspring. How much scope for impartiality does our genetic heritage leave us?<sup>8</sup>

Here we seem to me to be on much more dubious ground. One problem with the appeal to evolutionary considerations is that the relation between genes, on the one hand, and adult psychology and behaviour, on the other, has to be a very complex one – and the relation between genes and *possible* psychologies and forms of behaviour more complex still. Worries about naïve forms of preformationism and biological determinism lead some evolutionary writers to retreat to the language of ‘propensities’ and ‘predispositions’; but in what sense precisely can a given ‘predisposition’ or ‘propensity’ render us *incapable* of a certain way of life? In the absence of a proper account of these issues, it is hard to escape the thought that evolutionary considerations tend to be used illegitimately to reinforce a prior sense that a large degree of selfishness is unavoidable.<sup>9</sup> As Susan Oyama says, ‘Genetic and biological ... are often effective synonyms for *inevitability*, *unchangeability* and *normality*’.<sup>10</sup>

The appeal to the demoralization thesis (DT) is, when used by consequentialists, often mixed in with other considerations – in particular, with the claim that consequentialists should not constantly employ an explicitly consequentialist decision-procedure, and with attempts to defend different versions of consequentialism. But these issues need not concern us here: acting on (N1) need not entail the use of any particular decision-procedure – all Anna needs to do is to sign the direct debit form and then live with the consequences. She does not constantly have to be thinking about it. And

<sup>8</sup> Review of Kagan’s *The Limits of Morality*, *Mind*, 99 (1990), pp. 128–31, at pp. 129–30.

<sup>9</sup> For a sensible treatment of these matters, see D. Scoccia, ‘Utilitarianism, Sociobiology and the Limits of Benevolence’, *Journal of Philosophy*, 87 (1990), pp. 329–45, though he too concedes without argument the claim that ‘there is nothing one can do significantly to increase the strength of one’s benevolence energies’ (pp. 332, 338).

<sup>10</sup> S. Oyama, ‘Essentialism, Women and War: Protesting Too Much, Protesting Too Little’, in D. Hull and M. Ruse (eds), *The Philosophy of Biology* (Oxford UP, 1991), pp. 414–26, at p. 422.

the different versions of consequentialism are of less interest here than the fact that if *any* (non-collective) versions of them are to yield anything like common-sense morality, they will have to help themselves to strong versions of (DT). The following three examples may be taken as representative.<sup>11</sup>

The first is from an influential paper by Peter Railton.<sup>12</sup> Railton's model consequentialist agent, Juan, leads a life in which he devotes a lot of resources to his relationship with his wife, resources which could apparently do more good if distributed to international aid agencies. How is such devotion justified in consequentialist terms? By appeal to (DT):

Given the ways that Juan can affect the world, it may be that if he were less devoted to Linda his overall contribution to human well-being would be less in the end, perhaps because he would become cynical and self-centred (p. 121).

Although Railton puts these claims tentatively, they must be true in quite a strong form if Juan's way of life is to be justified in consequentialist terms.

Frank Jackson is similarly pessimistic about the capacity of individuals to sustain dedication to the needs of distant strangers, because of factors which he thinks 'no doubt have an evolutionary explanation'.<sup>13</sup> He admits that 'there are exceptions to this generalization about human psychology', and mentions Mother Teresa as an example, but offers no explanation of why they are able to do what the rest of us, apparently, cannot do. He just says (p. 481) that 'they do not seem to be dependent on the kind of close personal relationships that are essential to keep most of us from being outrageously selfish'.

Roger Crisp is another consequentialist who believes that the demoralization thesis (DT) is true in a sufficiently strong form to yield an approximate concurrence between consequentialism and common-sense morality, though in his case the main danger is perceived to be one of apathy rather than of selfishness. His model consequentialist agent

will establish close personal relations with a few people, less close personal relations with a larger group and, one would expect, still have something left to give for others as a whole. The value of close relationships will not be lost, nor will he or she be left apathetic.... But demands could not be made beyond the point at which self-defeatingness sets in. In the case of a single individual, brought up in the ordinary way, this point would arrive fairly quickly.<sup>14</sup>

<sup>11</sup> See also H. Sidgwick, *The Methods of Ethics* (London: Macmillan, 1907), esp. p. 434, for an important early version of this claim.

<sup>12</sup> P. Railton, 'Alienation, Consequentialism and the Demands of Morality', in S. Scheffler (ed.), *Consequentialism and its Critics* (Oxford UP, 1988), pp. 93–133.

<sup>13</sup> F. Jackson, 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics*, 101 (1991), pp. 461–82, at p. 480.

<sup>14</sup> R. Crisp, 'Utilitarianism and the Life of Virtue', *The Philosophical Quarterly*, 42 (1992), pp. 139–60, at pp. 157–8.

If what Railton, Jackson and Crisp say is true, then it may indeed be unwise for Anna to set about radically changing her life. But is it true?

It will be instructive to compare the case of (1b) with that of case (1a). The case of (1b) is certainly unlike that of (1a) in one way: Anna cannot just choose what kind of psychology to have, in the way in which she can, normally, choose what to do.<sup>15</sup> It can be possible, as we saw above, to act against even a very well entrenched disposition, and, if one does so consistently, it might even be possible to transform that disposition. But, first, success can never be guaranteed, despite one's best efforts, and, second, even if one does manage to transform one particular disposition, it is hard to see what the overall effect on one's psychological economy will be. One's dispositions form a complex web, and the attempt to change certain of them, whether successful or not, will have repercussions, perhaps negative ones, for others. To ignore all this would be to succumb to one kind of temptation: that of an illusion of total rational control, as if one could remake one's entire psychology, without residue, according to some ideally rational blueprint.

But the case of (1b) is like that of (1a) in two other ways. In the first place, Anna cannot regard her dispositions entirely as something that is just *given*, rather than as something she shapes. She has certain powers of self-transformation, and, where that fails, of self-control. To ignore this would be to succumb to another kind of temptation: of a kind of craven fatalism, all the more tempting when her present way of life is the one she simply prefers. And, second, those powers of self-transformation and self-control cannot be determined *independently* of the question of what moral sources are available to her. What she is capable of, over the whole course of her life as well as in the context of particular decisions, will depend on the kind of strength or inspiration she is able to derive from these moral sources.

As I said in §II, I think a large part of the reason for the prevailing pessimism about human nature is the neglect of this question of moral sources. Perhaps one can hardly expect opponents of consequentialism to have a vivid conception of the moral sources that might sustain a commitment to demanding versions of the theory, but the odd thing is that consequentialists too seem to have little sense of this issue. I think that there are at least three reasons for this. First, contemporary consequentialists have inherited a long-standing problem with the articulation of moral sources: the rather flat kind of value theory held, at least until very recently, by most consequentialists has tended to render it mysterious under what aspect consequentialists are supposed to find rigorous moral endeavour choice-worthy or required.<sup>16</sup>

<sup>15</sup> Cf. Aristotle, *Nicomachean Ethics* III v, 1113b 6–14.

<sup>16</sup> On this issue, see Taylor, e.g., pp. 31–2, 332–3, 515–17.

Second, the temper of our age renders many of us suspicious of 'high-minded' moral talk, to such a degree that we seem to be losing much of the traditional vocabulary for articulating why it might be important that we act in certain ways. Aristotelian virtuous agents are supposed to act 'for the sake of the *καλόν*' (e.g., *NE* III vii, 1115b 12–13). I am not sure just what the word '*καλόν*' would have meant to a contemporary of Aristotle, but I am sure that the usual contemporary translations, 'noble' and 'fine', do not have the right kind of resonance for us. Such words, in this context, sound dated and even slightly ludicrous to us, so that we struggle to find the vocabulary to express what makes certain actions and ways of life choiceworthy. Similar points might be made about the words *base*, *worthy*, *unworthy* and *earnest*, all of which were apparently used without irony or psychological distance until fairly recently.

And third, there is the widespread desire to show that consequentialism does not, when everything is taken into account, diverge too much from common-sense morality: the prospects of such a demonstration are hugely enhanced, in the kind of way that we have seen, if hard-headed claims about human nature are combined with neglect of the question of moral sources. Many consequentialists seem ready to concede that too much divergence from common-sense morality, particularly in terms of demandingness, would constitute a potentially decisive objection to the theory, but the underplaying of moral sources helps to defuse this objection. This puts them in the deeply paradoxical position of having a vested interest in underplaying the very sources that might make adherence to their own theory, in its more demanding forms, possible.

Are there, then, moral sources available to consequentialists that might make a more rigorous form of moral endeavour possible? Well, I do not think that any of the standard *arguments* for consequentialism are likely to be much help. If Anna is asking herself whether she is likely to be able to sustain her commitment to (N1), it is unlikely that she will feel that her belief that consequentialism is, say, entailed by our linguistic dispositions, or that it follows from a thesis about the nature of reasons, or that it represents the best systematization of our common-sense ideas about morality, will make much difference. It seems to be too easy to acquire such beliefs for the mere having of them to make much of a difference, in the face of other well entrenched motivations. What is also needed is some sense of why it is *important* to act on (N1). From what kind of source might such a sense of importance be derived, one that might be powerful enough to sustain Anna's motivation to act on (N1) throughout her life?

One source that has been explored by some writers in the literature is a vivid awareness of the suffering of those who would be affected by acting on

(N1). Information about the plight of the needy in the Third World often strikes us, as Shelly Kagan says, like ‘something we read about in a story book’.<sup>17</sup> We do not usually have a vivid sense of the *reality* of such suffering, that it is really happening, in the very same world we live in today, and that it is, in certain cases, almost indescribably awful. Taking steps to increase our awareness of such suffering might be expected to make some difference.<sup>18</sup> But even Kagan, who explores this approach at great length, concedes that something else, something focused on the *agent’s* life, may be needed. Like many consequentialists, he sees that ‘something else’ as being some recognition of the agent’s *interests*, and he argues that the agent’s interest in having the kind of ‘integrity’ which follows from living in accordance with a justified moral system is one that only consequentialism can satisfy.<sup>19</sup>

There are various problems with this: one obvious problem is whether it is indeed the case that only consequentialism can be justified. But more important here is the fact that Kagan conceives such ‘integrity’ as a prudential value, and indeed says (p. 392) that it is ‘one value among others – and it can be outweighed’. What seems to me rather to be needed is for the agent to articulate a description under which a life of rigorous moral endeavour is, at the least, *choiceworthy*, or, at most, *required*, without the presumption that the criteria by which such a way of life emerges as choiceworthy or required have anything to do with the agent’s interests – and, further, a sense that such criteria *cannot*, at least in any straightforward way, be weighed alongside other values. It might be, as Kagan suggests, a sense that only this kind of life can be justified, or, perhaps better, it might be a sense that to continue to live in a little bubble removed from the suffering of others would be shallow and mean, or a failure of one’s human potential, or, at root, a failure to take one’s *own* life seriously: the manifestation of a decisionless coasting along without really stopping to ask what one is doing. Although such reflection would be focused on the quality of an agent’s *own* life, it would not, in any ordinary sense, be about what is in his *interests*.

Perhaps consequentialists could also explore the moral sources articulated by other moral theories, to see if any of them could be put to consequentialist use. The Kantian idea that, in going against the dictates of reason, one makes oneself ‘a plaything of the mere inclinations and hence a thing’<sup>20</sup>

<sup>17</sup> S. Kagan, *The Limits of Morality* (Oxford: Clarendon Press, 1989), p. 295.

<sup>18</sup> Cf. Kant, *The Metaphysics of Morals*, tr. M. McGregor (Cambridge UP, 1991), vi 457.

<sup>19</sup> See P. Singer, *Practical Ethics* (Cambridge UP, 1986), ch. 12, for another example of the assumption that what is needed to make up any motivational shortfall in relation to consequentialism is a demonstration that rigorous moral endeavour can be shown to be in the agent’s interests.

<sup>20</sup> Kant, *The Metaphysics of Morals* vi 420.

might be one example; the Aristotelian idea that, in flouting the requirements of virtue, one gains nothing, might be another – though it is a difficult matter to determine just how much the effectiveness of these Kantian and Aristotelian ideas depends on particular views about the *content* of reason, views that consequentialists do not, of course, share.

Just how effective reflection on such moral sources would be is something that it is hard to specify with any determinacy, but I hope it is at least clear for now that what an agent is capable of cannot be determined independently of this question. And yet most consequentialists discuss (1b) without broaching such questions. It is as if they have only a ‘pale’ belief (to use Kagan’s term) in their own theory, without any sense of why it is *important* to do what the theory says.

Can Anna, then, sustain her commitment to (N1)? What one might hope is that she will not succumb to the illusion of total rational control, but neither will she view her own life as something that merely *happens*. She will take what possibilities for self-transformation are available to her, and, where that fails, for self-control, in both cases bolstered by a thorough exploration of moral sources (one of which might indeed be the sense that it is, fundamentally, *up to her* how to live). However, she will also recognize that certain of her dispositions may remain unalterable, or that trying to alter them would be too costly and they thus need merely to be *managed*. But if she has sufficient self-knowledge, she will know that she may be tempted to put certain dispositions in this category just because she does not want to try to change them, and if she has sufficient self-respect, she will choose not to do this. Given all of this, it seems to me to be very hard to say with any confidence that (1b) is true.

One thing that emerges from this is that a lot depends on the quality of the deliberation and reflection that Anna conducts, the view she takes of her own life and actions, the moral sources she explores, and so on. Railton has his model consequentialist, Juan, say to himself [the numbering is mine]:

- (i) Look, it’s a better world when people can have a relationship like ours – and nobody could if everyone were always asking themselves who’s got the most need....
- (ii) Anyhow, I know that you can’t always put family first.... But still, you need that little circle. People get burnt out or lose touch, if they try to save the world by themselves. The ones who can stick with it and do a good job of making things better are usually the ones who can make that fit into a life that does not make them miserable. (iii) I have not met any real saints lately, and I do not trust people who think they *are* saints.<sup>21</sup>

<sup>21</sup> At this point in the paper (p. 111), it is true, Juan is the ideally non-alienated moral agent, rather than specifically a consequentialist, but Railton later (p. 114) says that ‘Juan, it might be argued (if the details were filled in) is a sophisticated consequentialist’.

(i) is irrelevant unless Juan is a collective consequentialist – he should be asking himself what it would be best for *him* to do, given realistic assumptions about what others will in fact do. In (ii) Juan uses the kind of phrase, ‘save the world by themselves’, that is characteristically used by conservatives and cynics to lampoon the efforts of reformers. If ‘saving the world by himself’ is the description under which Juan construes the alternative to his present way of life, then failure will certainly be guaranteed, but why should he so construe it? Why not construe the alternative under the description, say, ‘making life better for many people who are suffering terribly’, or ‘living in a way that does not try to shut out the reality of the lives of others’?

Then we have the generalization about people getting ‘burnt out’, another version of the demoralization thesis. Clearly some people get burnt out and some people do not: Juan needs to consider the causes in both cases. Could it be that some of those who get burnt out do construe their project in megalomaniac, world-saving, terms? Could it be that those who do not get burnt out have a strong grip on moral sources? (This may help to explain the ‘exception problem’ which Jackson mentioned, but to which he did not offer any answer.) What moral sources might serve to sustain *Juan’s* motivation? Without a serious examination of these questions, the remark about people getting burnt out looks like a handy rationalization.

It is not entirely clear what (iii) is doing here. Apparently the thought is that devoting more resources to the needy would require Juan to regard himself as a kind of ‘saint’, and that this would make him untrustworthy. Who knows where this thought comes from? It looks again as if Juan is helping himself to the kind of charge used by conservatives and cynics to try to excuse their own selfishness. (For an unloaded invocation of the idea of saintliness, see Hare’s claim that ‘each of us ... has to ask himself what is the level of saintliness of which he is likely to be capable, and strive for that’.<sup>22</sup> It might seem surprising, given this, that Hare too thinks that we should stick pretty close to common-sense morality, but any sense of bewilderment is somewhat reduced when we remember that the only ultimate justification that he can find for such striving is a prudential one.<sup>23</sup>)

In general, the tone throughout the passage I have cited from Railton is defensive, and there is no suggestion of any moral sources that may enable Juan to sustain his commitment to the needy. If this is the quality of Juan’s deliberation, then it may indeed be counter-productive for him to change his way of life radically. But that claim would then have something of the character of a self-fulfilling prophecy. What Juan needs to do is, first, to stop portraying a more demanding moral life in biased, prejudicial terms;

<sup>22</sup> R.M. Hare, *Moral Thinking* (Oxford: Clarendon Press, 1981), p. 201.

<sup>23</sup> See Hare ch. 11; cf. Taylor p. 87.

second, to explore the moral sources available to him to sustain his motivation; third, to beware of defensive rationalizations designed only to justify what he either wants to do, or on non-consequentialist grounds feels that it is justifiable for him to do. If he were to do all this, perhaps he would come to believe that he could do much more.

#### IV. OTHER MORAL AGENTS

I have focused on the example of a particular consequentialist agent, in part because the philosophers whose views I am opposing think that (N1) is defeated even for such agents. But I also chose to focus on a particular agent because somehow it is more tempting to construe human nature as setting stringent limits to moral endeavour when thinking about people in general than when thinking about particular individuals. And I have tried to portray things from Anna's point of view, because it is also more tempting to construe human nature in this way even when thinking about particular individuals from the outside, as it were, than when one is occupying the first-person point of view. It is much easier to believe that, say, acting on (N1) is beyond the capacities of people in general than really to believe that you yourself *cannot* do it. In taking up Anna's point of view, I hoped to overcome this kind of temptation.

But nearly all of what I have said applies to people in general too. I have not made Anna idiosyncratic in any way. Her main peculiarity is that she already has a strong grasp of the kind of claims made on her by other people in desperate need. In her case, this sense might have been developed because she is a consequentialist. It might not, of course: the direction of influence might even have been the other way round – that she became a consequentialist because of a prior sense of the moral importance of the needs of other people. But the point I want to make here is that none of what I have said rests on her being a consequentialist. It would make a difference if being a consequentialist enabled one to drink from powerful, distinctive moral sources, but it seems that nothing could be further from the truth. The moral sources that I have briefly sketched have nothing specifically to do with consequentialism at all. And it would make a difference if it was necessary to be a consequentialist in order to have a strong sense of the moral importance of the needs of others. But this is clearly false.

It appears, then, that we should say of people in general much the same as we have said of Anna, subject to the proviso about having this strong sense of the moral importance of the needs of others. We should say that *if* people in general come to share this sense, and *if* they are able to tune into

powerful moral sources, then they too are likely to be able to sustain rigorous moral endeavour. Nothing about *human nature* prevents this. Cultural factors, broadly speaking, are much more significant: the prevailing sense of the moral importance of the needs of others, the particular arguments that are considered to be effective and the moral sources available.

This has important implications for the social role of philosophers, because philosophers do have some influence on these cultural factors. Philosophers who believe in the moral importance of the claims of the distant needy can play an important role by persuasively laying out reasons for believing this, by refuting defensive rationalizations, and by articulating moral sources. This last task is particularly important. For such sources to be effective it will actually have to cross the mind of someone like Anna – as, apparently, it did not strike Juan – that there might indeed *be* sources that are able to support the commitment to demanding norms like (N1). Whether such an idea does strike her, or others like her, will depend in part on the prevailing ideas in the culture in which she lives. If that culture makes the demoralization thesis seem obvious and unquestionable, perhaps bolstered by a bit of popular evolutionary psychology, then the possibility of exploring moral sources that may sustain her commitment to demanding norms may simply never occur to her. In this kind of way, pessimism about our capacities for rigorous moral endeavour may become at the social level, as it sometimes is at the individual level, a kind of self-fulfilling prophecy.<sup>24</sup>

*University of Reading*

<sup>24</sup> I would like to thank, first, Michael Lacewing and Ellie Mason for talking some of these questions through with me as I was writing the paper; and, second, anonymous referees at *The Philosophical Quarterly*, Simon Blackburn and, especially, Brad Hooker for their helpful comments on the penultimate draft.