Utilitarianism and Psychological Realism

SOPHIE RIETTI
University of Ottawa

Utilitarianism has frequently been criticized for lacking psychological realism, but what this means and why it is thought to matter varies. This article distinguishes and examines three main relevant kinds of appeals to psychological realism: (a) A minimalist, self-avowedly metaethically neutral and empirically based 'ought implies can' approach, exemplified by Owen Flanagan. (b) Arguments from psychological costs and flourishing, exemplified by Michael Stocker and Bernard Williams. (c) 'Thick' psychological realism, exemplified by Elizabeth Anscombe, where a conception of human nature does not simply provide constraints on value theory, but forms the substantive basis on which it builds. The main challenge raised for utilitarianism turns out to be metaethical, not a matter of empirical psychology. The question is not so much whether utilitarianism can accommodate (putative) descriptive facts of human psychology as such, but what normative weight these facts should be given and why.

INTRODUCTION

A common line of objection to utilitarianism in the recent philosophical literature holds that utilitarianism is problematic as a moral theory because of its perceived lack of psychological realism – the latter notion being often significantly vague. This criticism is primarily directed at the normative demands made (or assumed to be made) by utilitarians, but by extension also at their metaethical assumptions, and, most particularly, at their moral psychology, the supposed thinness of which is often implied to be the true culprit. As a result of its flawed and shallow moral psychology, the critics claim, utilitarianism makes demands on agents that are too high, and of the wrong kinds, with a range of damaging effects, not just to the well-being of agents, but also to their moral capacities and performance. This line of objection holds that people cannot, need not, or even *ought not* to try to do what utilitarians say people should do, for reasons that ultimately appeal to some notion of actual or potential human nature.

Thus, for instance, Bernard Williams[1] famously argues against utilitarianism because it alienates agents from their own projects and principles, which in turn are crucial for their maintaining integrity of agency, and for their flourishing. Similarly, Michael Stocker[2] argues

[1] Bernard Williams, 'A Critique of Utilitarianism', in Bernard Williams and J. J. C. Smart, *Utilitarianism: For and Against* (Cambridge, 1973), pp. 77–135.

[2] Michael Stocker, 'The Schizophrenia of Modern Ethical Theory', *The Journal of Philosophy* 73 (1976), pp. 453–66.

© 2009 Cambridge University Press
doi:10.1017/S0953820809990094
Utilitas Vol. 21, No. 3, September 2009
Printed in the United Kingdom

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

that utilitarianism imposes a 'schizophrenic' divergence between motivation and justification, and undermines personal relationships; and a whole host of critics of utilitarianism since its inception have argued against its consequentialism, its psychological hedonism, the impersonality of its impartial adding across persons, and its emphasis on maximization, on grounds that owe a good deal to notions of psychological realism.

Williams and Stocker have become quasi-canonical touchstones for the kind of objections to utilitarianism under consideration here. As I shall argue below, however, in this line of attack they find further, if strangely assorted, fellow travellers in authors such as Owen Flanagan[3] and Elizabeth Anscombe.[4] Like Stocker and Williams, if not quite for the same reasons, these authors are critical of much of the (until recently) mainstream approach of moral philosophy, utilitarianism centrally included. Also like Stocker and Williams, they raise issues of psychological realism against utilitarianism not purely on descriptive grounds, but also on evaluative and prescriptive ones, and typically in the name of alternative visions of what moral theories could and should be. Williams and Anscombe also, notably, both display a distinct unease with the notion of morality as such, preferring 'ethics' as an approach; insofar as this touches on the main issue at stake here, this distinction will be discussed below.

While, as stated, this line of objection to utilitarianism is an influential one, not just in the philosophical literature, but also among non-philosophers, it takes a variety of forms, and these have rather different implications for the issues at stake. They also make a considerable difference as to which kinds of utilitarianism will be deemed (most) problematic. I will be looking here at three main forms which the critique of utilitarianism from psychological realism can take. I argue that while these raise concerns which both utilitarians and moral philosophers more generally would do well to take seriously, the appeal to psychological realism, in order to carry weight, needs to be underpinned, first, by a more developed empirical psychology than may be currently available (or adaptable for the purposes of moral philosophy) and, second, by a fuller account of why these psychological factors should carry normative weight.

Moreover, a number of the psychological factors appealed to against utilitarianism – notably, the separateness of persons, and the special significance of our nearest and dearest – may also be significant distorting influences on moral reasoning, a point utilitarianism notes and aims to correct for; there are, I will argue, good reasons to worry

[3] Owen Flanagan, *Varieties of Moral Personality* (Cambridge, MA, 1991).
[4] G. E. M. Anscombe, 'Modern Moral Philosophy', *Philosophy* 33 (1958), pp. 1–19.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

about throwing the baby out with the bathwater in aiming to bring our moral demands closer to the supposed facts of human nature rather than vice versa.

A final initial aside: as most will know, utilitarianism rarely finds itself alone in the dock on the charges under discussion here, being typically accompanied by at least Kantian deontology, and occasionally by some variants of contractarian ethics. Depending on who is raising the charges, virtue ethics, particularly the Aristotelian kind, may find itself confusingly alternating between sharing the dock and being held up as a shining alternative to the accused. *Mutatis mutandis*, much of the following discussion will apply to other normative approaches too, but the immediate focus here is on applications to utilitarianism.

WHAT DOES THE APPEAL TO PSYCHOLOGICAL REALISM AIM FOR?

At its most basic, the demand for psychological realism, targeted at utilitarianism, implies something like this: the gap between the facts of human psychology and the demands of utilitarian morality is, on crucial if not all points, unbridgeable, or excessively, possibly self-defeatingly, difficult to bridge. The difficulties may be due to formal or empirical issues – it is primarily the latter which will be at stake here. The distinction may often be a fine one, though: as Andrew Sneddon[5] has argued, the arguments from human nature – typically focused on alienation – raised against utilitarianism by writers such as Williams and Stocker may be given either a formal or what he calls a psychological (and what I shall call an empirical) reading.

Thus, when Williams argues that utilitarianism undermines integrity (understood largely in terms of the individual's capacity to sustain commitment to own projects and values) and through that the very capacity for agency, this can be read as a claim that utilitarianism is formally self-defeating (utilitarianism requires *and* undermines effective agency), or as an empirical claim about human incapacity for sustaining effective agency if commitment to individual projects and principles is undermined. Similarly, when Stocker argues that a utilitarian attitude to the value of persons (i.e. that it is conditional, and non-intrinsic) is incompatible with the human good of love, which in turn is necessary for a happy, fulfilled human life, this can again be read as a claim that utilitarianism is formally self-defeating (it undermines a good that is essential for the human happiness it aims to promote), or as an empirical claim about the nature of human love and what strains it can take. I will tend here to take it, and will argue for this later

[5] Andrew Sneddon, 'Feeling Utilitarian', *Utilitas* 15 (2003), pp. 330–52.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

though not at great length, that the empirical psychological claims are the most crucial ones, and that they underpin the formal ones – that assumptions about what human agency and human love *are*, in an empirical sense, would be required for, and inform, the claim that utilitarianism's attitude towards it is self-defeating.

While a common focus is on the *level* of demands made by utilitarianism (it asks too much, for instance by demanding the sacrifice of personal projects and relationships for the sake of overall utility), closer examination, for instance of the examples from Stocker and Williams above, reveals that the target is almost as often the *kind* of demands made (it asks the wrong things), and the *grounds* adduced for them (it makes demands for the wrong reasons). The critics may themselves be willing to countenance quite high levels of demandingness, often of a perfectionist kind, as long as the content of and reasons for the demands are to their liking. Appeals to psychological realism, while they may undermine specific forms of demandingness associated with utilitarian positions, are often put forward by proponents of theories, such as neo-Aristotelian virtue ethics, that are themselves highly demanding, albeit not in quite the same ways or for the same reasons as utilitarianism.

A useful distinction can however be introduced here, adapted slightly from its original focus on high demands, from Samuel Scheffler.[6] When a moral theory makes particularly high demands on agents, Scheffler argues we have two basic options in responding to it:

First, we can hold that the demands need to be reduced in strength or scope – and, for our present purposes, that they may also need to be adjusted in kind, and not least in rationale, to something more psychologically realistic.

Second, we can hold that 'morality demands what it demands', in which case either 'so much the worse for morality' or 'so much the worse for us'. Susan Wolf[7] might exemplify the former option, Shelly Kagan[8] the latter.

The question here, then, is this: if we choose an option of the first kind – that morality should be changed – could this be on grounds of psychological realism? It is worth noting, again, that some of the authors who object to utilitarianism by appealing to psychological realism, such as Bernard Williams, seem drawn towards 'so much the worse for morality' as an alternative option if morality cannot adapt

[6] Samuel Scheffler, 'Morality's Demands and their Limits', *The Journal of Philosophy* 83 (1986), pp. 531–7.

[7] Susan Wolf, 'Moral Saints', *The Journal of Philosophy* 79 (1982), pp. 419–39.

[8] Shelly Kagan, *The Limits of Morality* (Oxford, 1991); see also the *précis* article 'Defending Options', *Ethics* 104 (1994), pp. 333–51.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

to their requirements. The broader issue at stake is this: if options of the first kind are preferable, is this on descriptive grounds, normative grounds, or both? If both, what if anything is the link between them? Is it the facts of human nature, whatever they may turn out to be, by themselves and *qua* facts, that give reason to oppose utilitarianism, or must they be given normative weight in order to make a difference?

What is it about the demands utilitarianism makes on agents that raises a problem, though? As indicated above, sheer demandingness may be much of the problem, but it is not all of it. Thus, for instance, Scheffler's discussion makes a distinction between three senses in which a moral requirement can be too demanding. First, demandingness may mean *stringency* – how much does the theory ask of agents in morally relevant contexts? Second, demandingness may be a matter of *pervasiveness* – how many, and which, contexts does the theory treat as morally relevant? Are any parts of human life exempt from its encroachments, or at least left open to individual discretion? Lastly, demandingness may be a matter of *overridingness* – do moral considerations trump (all) other kinds of considerations? A moral theory could be (too) demanding in any one or more of these senses. And if what it is demanding *about* is the wrong things for the wrong reasons, then of course so much the worse.

In terms of applications to psychological realism, stringency comes across as the most basic problem: does the theory ask more and other than we are psychologically capable of doing? Does it ask more and other than we can do and still thrive? Pervasiveness and overridingness seem primarily to exacerbate rather than create the problem – though they would certainly raise problems of their own, pervasive but non-stringent demands would be less of a concern than pervasive and stringent demands. (Non-stringent but overriding demands intuitively seem like an odd notion, though examples could presumably be found or invented.) The implicit resistance to psychological defeaters in both pervasiveness and overridingness would however be a likely concern to proponents of psychological realism. Pervasive demands might raise moral claims against an agent in more areas of his life than he is psychologically capable of handling, or at least, of handling and still thriving. Overriding demands might trump considerations the agent is not psychologically capable of setting aside, or, as before, that he is not capable of setting aside and still thriving.

When a theory makes particularly high demands, Scheffler argues, we can reduce them by targeting any of these kinds of demandingness: by rejecting stringency (Scheffler's favoured solution), by exempting at least some areas of human life from moral demands so as to limit pervasiveness (as suggested, variously, by Stocker, Nagel and

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

Williams), or by rejecting the notion that moral demands override all other considerations (e.g. Wolf).

As stated, though, what proponents of psychological realism object to may be a matter of *what kind of thing* is demanded as well as *how high* the demand is, but also, crucially, of the *grounds* for the demand. When Stocker objects to utilitarian (and deontological) accounts of why it is good to visit a friend in hospital, he objects because he thinks utilitarianism gets it wrong as to *why* it is good to do so. A utilitarian, as understood by Stocker, holds that what would justify visiting a friend in hospital is that doing so would promote utility-maximization (or at least not be incompatible with it), not the acting out of friendship, and not friendship's being an intrinsic good, or the inherent value of the friend as a person. So even where utilitarianism picks out the right thing to do (visit your friend), it does so for the wrong reasons, and the fact that utilitarianism is guided by wrong reasons (in Stocker's view) also makes it likely that utilitarian views about what to do will be wrong in other cases.

This is not simply a case, then, of objecting to utilitarianism making *high* demands (e.g. of the kind defended, variously, by Singer and Kagan), but to the *kind* of demands, and the *reasons* given for them, even where they are not high. The 'schizophrenia' Stocker thinks utilitarianism imposes is not simply a matter of setting the bar of what we ought to do higher than we are willing to jump: it is a disorder imposed even by the more easily met demands utilitarianism makes of us, and the carrier is utilitarianism's disdain, in its account of justification, for facts, some of them normatively weighty, about human nature.

HOW DOES PSYCHOLOGICAL REALISM AFFECT THE PICTURE?

While utilitarianism takes a number of different forms, some common points that most utilitarians share tend to be ones to which promoters of psychological realism will take exception. The basic principle of utility – that what is right is what will promote the greatest possible happiness of the greatest possible number – already suggests that utilitarianism will be demanding in all three senses distinguished by Scheffler. The demand to maximize makes for stringency; the fact that the demand applies, at least in principle, wherever utility can be maximized, makes it pervasive, and its being set as *the* moral consideration makes for its demands to be overriding. These points are overtly accepted by prominent advocates of utilitarianism such as Peter Singer and Shelly Kagan. Singer's famous article 'Famine, Affluence

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

and Morality'[9] explicitly raises, and inclines towards, the idea that the duty to maximize may require sacrifices of us up to the point where the marginal utility of the sacrifices declines – that our duties to try to relieve poverty and famine in the developing world, for instance, will take precedence until something of comparable moral significance will be lost by our attempts to fulfil the duty. Kagan's *The Limits of Morality* is a sustained defence of the view that the *pro tanto* reason to promote the good (where the good is understood in terms of overall utility, and to be maximized) cannot be defeated by appeals to agent-relative options.

In addition to being highly demanding in this sense, utilitarianism may be taken to be unrealistic in relation to our actual capacities – if we are not able to foresee the future well enough to perform the required calculations, for instance, the limitations on our knowledge, foresight and computational abilities may be psychological obstacles to our meeting the demands made by the principle of utility.[10] But these appeals to what is psychologically possible are also usually supplemented by other considerations, typically about the nature of human motivation and individuality: as separate persons, with particular attachments to particular others, and to our own projects, we are unlikely to be motivated by considerations of overall utility, or to desire to put these ahead of our individual goals and attachments, and our more agent-relative motivations are likely to present obstacles both to the will *and* to the ability to promote the overall good. Moreover, setting the demands too high may be demotivating, by setting agents up for failure, and also, though in a different way, by alienating them from their own projects and relationships, directing their actions instead towards more general goals.

The charge of alienation can also be reiterated even if the utilitarian responds by distinguishing what motivates an action, e.g. visiting a friend, from what justifies it. Indirect utilitarianism need not require that agents be directly motivated by the principle of utility: as long as the actions their actual motivations drive them to perform are compatible with the promotion of overall utility, both the actions and the motivations may be justified by this compatibility. But the principle of utility still operates as a background condition here in ways those who worry about alienation may continue to object to. If the conditions of entering, sustaining and ending personal relationships, for instance,

[9] Peter Singer, 'Famine, Affluence and Morality', *Philosophy and Public Affairs* 1 (1972), pp. 229–43.

[10] On some possible implications of this point, see e.g. James Lenman, 'Consequentialism and Cluelessness', *Philosophy and Public Affairs* 29 (2000) pp. 342–70 – also his 'Utilitarianism and Obviousness', *Utilitas* 16 (2004), pp. 322–5 and Elinor Mason, 'Consequentialism and the Principle of Indifference', *Utilitas* 16 (2004), pp. 316–21 in reply.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

are supposed to be subject to considerations of utility, so that you may still be required to abandon loved ones, or avoid entering into relationships with them in the first place, due to considerations of overall good, then the worry about alienation, while still indirect, is still there.[11]

In addition to these considerations, utilitarianism is often held to be problematic due to both its hedonism and its 'adding across persons'. People do not just want pleasant experiences, critics of utilitarianism argue (Nozick's famous 'experience machine'[12] thought experiment being a particular case in point), and even if they did, the *general* happiness is not one that anyone experiences. So utilitarians, critics hold, misunderstand both what the good for humans is, and who it is for, and because of these false assumptions, its normative demands are also flawed. Moreover, the form utilitarian impartiality takes, and its supposed riding roughshod over the separateness of persons, also means that the distinction between self and other, which some critics of utilitarianism take to be both psychologically and morally foundational, is effectively elided within the utilitarian approach.[13]

For now, these are notable as initial reasons why utilitarianism specifically would be open to charges of lacking psychological realism, both in its descriptive assumptions and in the normative weight it gives to facts of human psychology. However, I shall argue, there are also

[11] On this issue, see e.g. Peter Railton, 'Alienation, Consequentialism and the Demands of Morality', *Philosophy and Public Affairs* 13 (1984), pp. 134–71; Dean Cocking and Justin Oakley, 'Indirect Consequentialism, Friendship, and the Problem of Alienation', *Ethics* 106 (1995), pp. 86–111; Elinor Mason, 'Can an Indirect Consequentialist be a Real Friend?', *Ethics* 108 (1998), pp. 386–93; Keith Horton, 'The Limits of Human Nature', *Philosophical Quarterly* 49 (1999), pp. 452–70; as well as Sneddon, 'Feeling Utilitarian' – noting invocations of Mill throughout. Stocker ('Schizophrenia') arguably anticipates a great deal of the debate on indirect consequentialism and utilitarianism in his 'schizophrenia' argument – the coming apart of motivation and justification is in his view precisely the problem, and not a solution, to the gap between human nature and what certain moral theories, particularly impartialist ones, ask of agents. I will not try here to settle any of these issues with any definiteness – that would be a topic for a whole different article (at least), but note also that Stocker is arguing primarily against moral theories that imply *pervasive* disconnect between human motivating drives and moral justification, not against the need for correctives to our actual motivational structures. Stocker's view does however imply that, in principle at least, the reasons a theory provides for action should be suitable to *become* motives, or the theory is schizophrenic.

[12] Robert Nozick, *Anarchy, State and Utopia* (New York, 1974), pp. 42–5.

[13] This point is particularly emphasized by Michael Stocker – see 'Agent and Other: Against Ethical Universalism', *Australasian Journal of Philosophy* 54 (1976), pp. 206–20, and 'Emotional Identification, Closeness and Size: Some Contributions to Virtue Ethics', *Virtue Ethics: A Critical Reader*, ed. Daniel Statman (Edinburgh, 1997), pp. 118–28. Self–other asymmetry as an issue for moral theory is also discussed, with significantly less emphatic conclusions than Stocker's, in Michael Slote, 'Morality and Self–Other Asymmetry', *The Journal of Philosophy* 81 (1984), pp. 179–92. See also Clay Splawn, 'The Self–Other Asymmetry and Act-Utilitarianism', *Utilitas* 13 (2001), pp. 323–33, on the prospects of incorporating the distinction within a utilitarian ethical system.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

significant differences between ways in which appeals to psychological realism are made in the literature.

APPEALS TO PSYCHOLOGICAL REALISM: SOME DISTINCT FORMS

At this point, then, it will be useful to distinguish more clearly some of the different ways in which psychological realism is appealed to by critics of utilitarianism. In the next sections, I distinguish three main kinds of such approaches. The first (type A) is a minimalist, 'ought implies can' sense of psychological realism; the second (type B) appeals to considerations of costs to human flourishing, and the third (type C) uses a thick notion of human nature and the human good as a foundation of, not just a restricting condition on, its conception of the good.

The three vary in the degree to which their objections to utilitarianism draw on descriptive versus normative readings of human nature: even the minimalist account, though, gives some normative weight to (putative) facts of human nature. Insofar as it is the normative weight given to psychological facts, rather than the psychological facts themselves, that drive the objections to utilitarianism, this suggests that the main battleground between utilitarians and proponents of psychological realism in moral theory lies in the realm of metaethics and not in issues of empirical psychology.

(A) Minimalist appeals to psychological realism: 'ought implies can'

The first approach to psychological realism is a minimalist one, setting the (supposed) psychological facts only as a basic constraint on what moral theories can demand.

A paradigmatic instance of this form of appeal to psychological realism is Owen Flanagan's principle of minimal psychological realizability (PMPR):

> 'Make sure, when constructing a moral theory or projecting a moral ideal, that the character, decision processing, and behavior prescribed are possible, or are perceived to be possible, for creatures like us.'[14]

In its *descriptive* aspect, Flanagan holds, PMPR picks out an aspiration of almost all moral theories. In its *prescriptive* aspect, PMPR sets out a criterion for evaluating theories in terms of this aspiration. PMPR picks out an aspiration of moral theories, rather than a characteristic of them, because a moral theory which aspires to be psychologically realistic may fail to achieve this goal. In the next instance, of course,

[14] Flanagan, *Varieties*, p. 33.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

it may also fail on criteria more specific to moral theories, such as setting out plausible metaethical and normative principles – PMPR, for Flanagan, acts as a fundamental but massively underdetermining constraint on moral theories.

An additional reason why PMPR is underdetermining is Flanagan's further adducement of what he calls the Thesis of Multiple Realizability of moral psychologies (TMR): the set of realizable moral psychologies is infinitely large. This does not, of course, mean that all moral psychologies are possible (or PMPR would be a pointless principle). But it does mean that Flanagan can allow for variation by both nature and nurture, and that especially in view of the interaction between the two variables, and the range of variation within each of them, specifying some *one* shared human nature against which moral demands can be tested is an unlikely outcome. It should however also be noted that nature and nurture can both act as limits on variability, perhaps more surprisingly in the latter case: socialization, Flanagan argues, and environmental influences more broadly, can make some traits, motivations and behaviours psychologically impossible that were not so naturally. Moreover, 'natural' traits need not be universal, but may be highly idiosyncratic, and traits acquired by nurture could be universally shared.

In the light of TMR, especially, PMPR is 'not remotely sufficient' to fix the right moral theory, *but* it is still a necessary requirement. Natural inclinations, Flanagan argues, are 'a mixed bag from a normative point of view'; only when provided with a determinate character by nurture, and a 'social space' in which to operate, can they be said to be virtuous or vicious. The determinate shape any individual's moral personality comes to have is the outcome of an utterly natural set of processes, but these are primarily ontogenic and therefore temporally, socially and geographically parochial. 'There is simply no transcultural species-being which can be said to constitute or be definitive of moral personality.'[15]

The role of psychology in ethics, then, as Flanagan sees it, is threefold. First, psychology provides a general picture of how, in rudimentary terms, persons are put together. Second, by so doing, it sets constraints on our conception of what sorts of persons are possible, and a picture of the degree of difficulty likely to be involved in realizing any particular kind of moral personality from among these possibilities. Finally, psychology provides a way of drawing the (impure, but useful) distinction between natural psychological traits and the (typically narrower) psychological traits acquired through socialization.

[15] Flanagan, *Varieties*, p. 51.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

The former are more empirically unyielding *ab initio*: the latter can be extremely unyielding once in place.

Minimal psychological realism and utilitarianism

Psychological realism of type A is based on empirical constraints, and – at least officially – metaethically neutral. It is worth noting, though, that the 'so much the worse for us' option indicated by Scheffler means that even if this gets us off the hook, the hook may leave exit wounds. We may concede, with Kant, that out of the crooked timber of humanity nothing straight was ever made, but that leaves possible a 'too bad about the timber' response: the straight aims of morality could have succeeded, if only the timber had been better. What appeals to psychological realism advocate, instead, keeping to the metaphor for the moment, is that morality be guided more by the nature of the timber, even if Flanagan's version of such accounts is flexible as to what this could mean.

This does not imply that PMPR, when wielded by Flanagan, does not cut down considerably on which moral theories are considered viable. According to Flanagan, PMPR rules out at least direct utilitarianism and act-utilitarianism (as well as Kantian deontology and any form of virtue ethics that leans towards perfectionism or claims about the unity of the virtues): we simply do not have the capacities, traits or motivational structures that these would require. The identity-criteria for action opportunities are too vague: we cannot clearly tell what will promote overall utility best in an ongoing way as we go about our ordinary lives. Also – especially in the case of act-utilitarianism – Flanagan argues that the computations involved are too demanding, and require impossible levels of constant attention, a point which compounds the problem of vague identity criteria. Flanagan lets utilitarianism, and consequentialism more generally, off one hook though: he does not hold that they need require agents to be right about actual consequences, and hence to operate at impossible levels of demand on our capacities for prescience.

Indirect utilitarianism and rule-utilitarianism, by contrast, Flanagan thinks will pass the PMPR-test; these theories are not necessarily in the business of providing a decision-making criterion, but perhaps only providing a (theoretically independent) right-making criterion. Decision-making, on these accounts, may be done by criteria that are more psychologically practically realistic – though this may depend on the specific content of rules as well as on our more general capacity for rule-following, indirect and otherwise. As a general point, though, Flanagan argues, the motivational structure required of agents by a moral theory cannot simply be read off its criterion of

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

rightness. If we recast act-utilitarianism as a higher-level theory, and allow a rule-utilitarian approach at the immediate level of decision-making, Flanagan thinks this may provide a basis for a defense of act-utilitarianism against PMPR: but this already concedes act-utilitarianism's initial implausibility under PMPR to be a telling point.

It should be noted that, would-be minimalist as this version of the appeal to psychological realism is, even PMPR will tend to be notably difficult to pin down – appealing to 'ought implies can' may be tricky insofar as assumptions about 'cannot' are generally extrapolations from 'is not' and 'does not' – and the accuracy of any such extrapolations will be further cast into doubt by Flanagan's own points about the variability and plasticity of human nature. Also – and this takes us on to the next version of psychological realism – the fact that an agent or group of agents is capable of realizing a moral approach need not mean that they will like doing so, will thrive doing so, or will feel strongly enough, let alone overridingly, motivated to do so. It just means that they *could*. And we might keep in mind that incentives can be provided from outside, e.g. in the form of sanctions, that can shape traits, motivational structures and behaviours in ways nature never intended, both for particular individuals and for larger social groups. PMPR is compatible with the possibility that moral theories not currently viable on grounds of psychological realism can become so later.

(B) Appeals to psychological cost and human flourishing

Where PMPR sets the bar for moral theories to be psychologically realistic relatively low, perhaps the most famous lines of objection to utilitarianism from considerations of psychological realism, those of Williams and Stocker, do not rely primarily on appeals to the psychological *impossibility* of acting on and living by utilitarian rules. Arguments from psychological realism, type B, rather appeal to considerations of psychological cost and the impact on human flourishing of implementing a particular moral theory. So, for instance, Stocker's rejection of 'schizophrenic' theories – ones that set up a too-wide gap between considerations of motivation and considerations of justification – does not rely on a strong case that it is psychologically *impossible* for agents to live by schizophrenic theories, but emphasizes instead the damage living by a schizophrenic theory would do.

Moreover, the appeal to reduced flourishing applies even if the agent himself or herself is unperturbed, so the appeal is not simply to a loss of subjective well-being. Rather, the argument is that bad theories are bad either way: if the agent tries to live by them and fails to, he will fail to flourish, wasting his capacity for agency on a futile and worthless

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

goal. If he tries to live by bad theories and succeeds, but at a high cost in distress to himself or others (e.g. his nearest and dearest, sacrificing his relationships with them for the sake of general utility), that is a bad outcome too. And if he succeeds not only in living by them, but also in convincing himself that doing so is living well, that too is a bad outcome, maybe even worse than the former two options, since they at least allow the agent to perceive that the theory in question has flaws. Being insensible to, or even comfortable in, one's own schizophrenic alienation is not a good outcome.

Thus, for instance, in Stocker's hospital visitor example his argument is that the utilitarian approach is alienating for interpersonal relations and imposes a schizophrenic split between what motivates a psychologically healthy agent and what morally justifies actions. To subject an action like visiting a friend in need to criteria of general utility, whether directly or indirectly, is to lose sight of the friend and the friendship as intrinsically valuable, and to become alienated from them both. But Stocker's argument clearly implies that the damage would not be undone if the visitor – or the patient – came to see the utilitarian view of her situation as the right one, and to conceive of themselves and their relationship in utilitarian terms, direct or indirect.

While Stocker tends to argue that utilitarian considerations are not capable of becoming motivations, and that this is a telling point against utilitarianism, his revulsion at the prospect of their succeeding in doing so strongly suggests that plain psychological impossibility à la Flanagan's PMPR is not the sole issue here. Trying to implement utilitarianism at the level of motivational psychology need not, on this account, be straightforwardly doomed to fail, but more importantly, it would be bad, even wrong, for the agent to try, *and* bad, even wrong, for him to succeed.

Similarly, for Williams, utilitarian demands alienate agents from their own agency, which it instrumentalizes to the point of potentially undermining it altogether, both at the level of action and at the level of moral feeling, both of which are reduced by the utilitarian to the value of their conduciveness to overall utility.

If anyone should still be unfamiliar with them, here are Williams' key examples.[16] Jim, on entering a village, finds a group of Indians lined up in front of a firing squad, and is given the choice, by the head of the firing squad, of killing one of them himself or having all of them shot by the firing squad. George is an unemployed chemist with a young family, and like Jim a pacifist: the only job available to him is for an arms manufacturer, and if he does not take the job, someone

[16] Williams, 'A Critique of Utilitarianism', pp. 97–100.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

more enthusiastic about arms manufacturing will. To these dilemmas, Williams argues, utilitarianism holds that the right answers – and the obvious right answers, at that – are that Jim should kill one person and George should take the job.

The argument is not however simply that Jim and George, as they are, with their pacifist principles, cannot thrive, morally or personally, if they are forced to shoot innocent people as the only way to save other innocents (Jim) or use their skills to produce weaponry to which they are morally opposed (George). For if Jim and George were to become resigned to their lot, or even embrace it, this would hardly appear on Williams' view as a happy outcome.

Nor, more generally, if Jim and George were to become resigned to, or happy with, the more general problem that their personal projects and principles could at any time become secondary to considerations of the overall utility of outcomes, would this make the problem go away. It is not simply that Williams thinks it would be bad for agents to *feel* alienated from their own agency in this way: it would be bad for them to *be* so alienated, to be just the instruments of utility-maximization, even if this did not make them feel alienated, or made them leap about with glad little cries at the peak of their personal felicity-calculus ranges. The kind of projects a genuinely utilitarian agent could have are not ones that it would be good for a human being to have, whether or not the human being in question was psychologically *capable* of having these projects or not.

Psychological realism B and utilitarianism

The argument here is in part, then, that utilitarian considerations go against the grain of human nature, but the going against the grain is not primarily cashed out in 'ought implies can' terms. Rather, assumptions about human nature primarily come in at the level of appealing to what kinds of traits, behaviours, motivational structures, and ways of living and relating, human beings are capable of having and *flourishing* as human beings. These limitations on what can count as flourishing will of course in their turn depend on underlying limits of what human nature is and is capable of, of the kind appealed to by Flanagan, but the 'ought implies can' considerations adduced by Flanagan operate here as the underpinning of the central concerns rather than as the central concerns.

The kind of features of human nature to which Williams appeals – primarily, the separateness and particularity of persons – also underpin his resistance to the 'morality system'[17] as an approach to ethics: the

[17] Bernard Williams, 'Morality, the Peculiar Institution', from *Ethics and the Limits of Philosophy* (London, 1985), pp. 174–96.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

paradigmatic 'morality' approach for Williams is Kant's duty ethic, but he notes that utilitarianism, while only 'marginally moral' insofar as it subordinates all issues to utility-considerations, nevertheless tends in the direction of creating duties out of these utility-considerations, and that convinced utilitarians often exhibit a level of guilt and conscientiousness about their own moral performance that belies the overt marginality of their commitment to the notion of moral duties. Williams favours instead a notion of ethics (which he sees as broader than, and inclusive of, morality) where duties, including those relating to utility-promotion, are merely one consideration among others.[18]

The morality-approach to duties, Williams argues, also fails to take into consideration agent-regret, a moral emotion that might still be felt by someone who has fulfilled his moral obligations: moral obligations, by implication, can both go beyond and fall short of telling us what it is good for us to do. Williams also considers the 'no emigration' approach of morality systems to be questionable – i.e. that morality systems tend to operate as if their obligations hold on people even involuntarily, and even if they do not share the values in question. So in the conflict between for instance George's commitment to his pacifist principles and a utilitarian assessment of his situation, the utilitarian approach would exhibit the questionable virtues of the morality approach to ethics by indicating that it is obvious, from the point of view of utility, that George should take a job that will compromise his personal values and integrity.

Stocker's argument, while in some respects similar to Williams', emphasizes slightly different features of human nature in making his case against utilitarianism (and deontology). Like Williams, he emphasizes the separateness of persons as against the utilitarian approach of impartially adding across persons: humans are separate individuals, and good is only good-for-someone, not independent of whose good it is.

In addition, as already indicated, Stocker emphasizes a further aspect of the separateness of persons as carrying normative weight, holding that there is a psychologically and ethically foundational distinction between self and other: what is forbidden, permitted and required of us can depend fundamentally on whether the patient is self or other. This is not just a matter of allowing what Slote[19] calls *agent-favouring permissions* (putting one's own interests first because they are one's own): Stocker also wants to allow what Slote refers to as *agent-sacrificing omissions*. Suicide and self-harm are not, Stocker argues, morally comparable to murder and abuse of others; nor is it wrong to

[18] See especially Williams, *Morality*.
[19] Slote, 'Morality'.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

sacrifice my own interest for that of another even when doing so will be detrimental to overall utility. Using one of Stocker's own less dramatic examples,[20] if my friend would enjoy an ice-cream less than I would, it is still not wrong for me to buy him an ice-cream and not buy one for myself. An impartial adding across persons, however, would suggest that I should buy the ice-cream for the person who will get most utility from it. That my friend is unlikely to be impressed with my hogging all the ice-cream for myself out of concern for overall utility will only feature in the utilitarian's considerations, Stocker thinks, insofar as it might somewhat reduce the utility of my hogging all the ice-cream.

Stocker's arguments from the separateness of persons, then, are somewhat distinct from Williams', which tend to emphasize the agent-favouring reasons to deviate from impartial adding across persons. Stocker also has a stronger emphasis on particular relationships, which again he thinks will be detrimentally affected by utilitarian approaches. Psychologically fulfilling interpersonal relations, Stocker argues (both on the agent and patient sides), are particular and concrete, not abstract and universalistic. And the value of friendship, as well as the value of the friend, best embodies human flourishing when it is not instrumental or conditional on utility, but treats both the friend and the friendship as having intrinsic value. This need not mean *indefeasible* value: there could be reasons to terminate a friendship, on Stocker's account; the point is that entering, sustaining or ending the relationship will not be conditional on utility, if the friendship is really one.

Stocker's broader argument is this: both in their relationships with others and in pursuing their own projects, humans thrive when their motivations and what they perceive to be the justifications of their actions are in harmony, and disharmony, if sufficiently strong and pervasive, may even make the relevant forms of action psychologically impossible or unhealthy. Experiencing being, or unconsciously being, 'schizophrenic' (not being motivated by what one values, or not valuing what one is motivated by) does not promote human thriving.

Where does this leave utilitarianism? An important point to note here is that someone holding views such as Singer's or Kagan's could concede all these points without abandoning utilitarianism, albeit at the cost of some bullet-biting, as long as the costs implied by these (putative) facts do not outweigh the benefits of utilitarian approaches to moral issues enough to make it the less favoured option on utilitarian grounds. When it comes to purely descriptive claims, there might be little difference between the utilitarian and his critics. The utilitarian may even, insofar

[20] 'Agent and Other'.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

as this works for, or at least not against, the promotion of general utility, wish to make concessions to human nature on these lines – albeit, importantly, not indefeasible ones: personal projects, individuation, and particular relationships are still ultimately subject to considerations of utility. But this need not depend on a radically different assessment of what human nature *is* from that held by the anti-utilitarian, so much as on a very different view of the normative significance of the psychological facts.

I shall have more to say about the implications of this way of attacking utilitarianism shortly: in kind, though, these implications are sufficiently similar to those of the next (and last) type of appeal to psychological realism to be considered that the two will be discussed together. For now, it is worth noting the increasing implicit reliance of objections to utilitarianism from psychological realism on giving normative weight to (putative) facts of human nature. Even Flanagan's would-be minimalist approach at least favours an 'ought implies can' view rather than 'so much the worse for us if we can't; morality demands what it demands': Stocker and Williams rely on a somewhat thicker notion of what we are psychologically capable of doing *and* still flourishing.

(C) Thick psychological realism

The third and final version of appeals to psychological realism to be considered here is what I shall call 'thick' psychological realism, or version C. Here, human nature does not simply provide constraints on moral theory, but rather forms the substantive basis on which any theory of value must be built.

A paradigmatic version of such thick psychological realism would be Elizabeth Anscombe's view, notably put forward in 'Modern Moral Philosophy', that an adequate moral theory must be based on a notion of the (human) good derived from an adequate philosophical psychology, which, she argues, we do not currently possess (though her own near-concurrent work, *Intention*, is presumably at least in part meant to help correct this deficiency). Basically, this form of psychological realism requires some form of full-fledged psychological naturalism of ethical theory. Anscombe, notoriously, advocates the jettisoning of concepts of *moral* obligation, *moral* duty, *moral* right/wrong, *moral* sense of ought, 'insofar as this is psychologically possible', seeing these as survivors or derivatives of an earlier, divine command-based notion of ethics which is no longer viable, and without which the derivatives are only harmful – misguided attempts to put some more human-centred concerns (however Kantian) in the place previously occupied by the divine. Anscombe waxes particularly scathing on the Kantian notion of

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

self-legislation, but her opinion of utilitarianism, and consequentialism (a term of her coinage) more generally, is hardly much higher.

She rejects both the notion that what is right depends on the outcome, and the notion that the rightness of the outcome can be judged by some standard based on pleasure; she argues, furthermore, that the notion of utility as a unitary standard against which different kinds of right and wrong actions can be judged obscures the significant differences between the particular virtues and vices embodied in the acts. Anscombe sees consequentialist ethics, in being consequentialist, as a real departure both from a divine command ethic and from its continuation in later deontology, but she also sees it as another dead end for ethics.

Rather, Anscombe argues, we need to look for a basis for norms in human virtues, taking as our starting point the species (not its individual members) regarded not purely biologically but from the point of view of the activity of thought or choice in regard to the various departments in life – as having such and such virtues (courage, compassion, greatness of soul). This notion of 'man' with the complete set of virtues would function as the 'norm' in much the same way as, biologically, 'man' with for instance a complete set of teeth is a norm. The notion of (moral) norm here is more an Aristotelian notion than a *quasi*-law like one, and in notable contrast to Flanagan, it is implicitly *not* a multiply realizable one: to quote Aristotle himself, it seems that for Anscombe men go wrong in many ways but right only in one.

On this kind of account, the main problem with utilitarianism is in a sense that it starts at the wrong end, with only a very thin notion of human nature, and then goes on to apply principles not derived from an account of human nature to this thin account. The paradigm case for this would presumably be something like Bentham's approach: a decision-procedure derived from mathematical and economical considerations on a large institutional scale such as national legislation, applied to a very thin characterization of human nature fleshed out primarily through psychological hedonism. In other words: take as your basic model of human nature that people want more pleasure, happiness, well-being, and put these together, in a quantifying manner, under the umbrella of utility; also take it that people want less pain and suffering, and quantify these as having negative utility: then add up the numbers to weigh the likely outcomes of available options for action, and decide. By contrast, Anscombe advocates an Aristotelian approach: start from a thick notion of what human beings are (rational social animals, with a range of spheres of operation, each sphere having distinctive virtues and vices), and derive value theories, both at the personal and the larger social level, from this thick notion of human nature.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

PSYCHOLOGICAL REALISM: OVERALL IMPLICATIONS FOR UTILITARIANISM

Options B and C are not ones where utilitarianism is likely to fall at the first hurdle from simple psychological impossibility; rather, the claim is that utilitarian demands in the longer run are self-defeating, for empirical reasons, or (in a more secondary sense) for formal reasons, and damaging even where they are not self-defeating. Options B and C however are not straightforwardly about the constraints empirical facts of psychology put on moral demands: rather, they give normative weight to (putative) facts about human nature which a utilitarian might concede empirically but give little or no normative weight. Kagan, for instance, refers to the demonstrably limited actual ability of utility-maximization to motivate agents as 'mere hindrance'.[21]

B and C see facts of human nature as constitutive of the content and form of (adequate) moral theory, not just as setting limiting conditions on it. In this they may also be partly de-emphasizing (though they may both concede, and Anscombe would tend to strongly share) Flanagan's point that natural human inclinations are normatively speaking 'something of a mixed bag', and lack clear normative implications until they have a social space, even if possibly just an implicit one, in which they can help or harm.

Flanagan's account, as stated, is at least officially metaethically neutral, beyond holding that morality must not demand what is psychologically impossible. And there are some fairly clear limits on how far arguments against utilitarianism from psychological realism can be made in purely descriptive terms. So, when options of type B and C appeal to considerations of human flourishing, this approach may itself be open to empirical question. First, how high and how pervasive the psychological costs of utilitarianism in terms of flourishing themselves are, is in itself largely an empirical matter. Granted, there is always the possibility in any particular instance that considerations of utility can trump considerations of flourishing, especially individual flourishing. But conversely, utilitarianism, if it is to function for humans, must also make at least some concessions to what could count as flourishing: the utilitarian notion of well-being could only with difficulty be entirely disconnected from this and still pass some minimal test such as PMPR. Furthermore, if we concede TMR (contra Anscombe) *and* assume that human nature is to a considerable degree plastic, the problems may be contravened to some extent by this route: human nature could perhaps be made more adaptable to utilitarianism, as well as vice versa, in which case appealing to human nature 'as is' in

[21] 'Defending Options'.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

order to undermine utilitarianism's credibility would not be as viable a route. Given plasticity and variability, any normative appeals to human nature would have to set out, and fully justify, why any particular form human nature could take (e.g. Anscombe's theoretical norm) should be given privileged consideration in moral theory.

Beyond the purely empirical, then, the descriptive features of human nature that underpin the concerns will have to be given normative weight, so as to block any 'so much the worse for us' moves, a point that also applies if we want to use 'ought implies can' to get us off the hook with only minimal exit wounds. The main point, here, though, is this: utilitarians can share any (combination of) the descriptive assumptions adduced by the critics. (That these assumptions are many and varied, and not necessarily mutually compatible between different schools of anti-utilitarians, may also serve to undermine the force of appealing to any particular claim about human nature – or at least suggest that not *all* the critics can be quite right.) Utilitarians need not deny the separateness of persons, or hold that general utility can be experienced by individual agents. In fact, it would be hard to think of a utilitarian who held such a bizarre view. Moreover, utilitarians in general need not even be wed to some of the assumptions that they can more legitimately be accused of holding, e.g. a somewhat shallow hedonism: as far back at least as Mill's differences with Bentham, there has been room for both philosophers and pigs in utilitarian notions of thriving. The utilitarian 'church' as a whole is a broad one, though some of its members may use narrower doctrinal approaches than others. What utilitarians will be committed to, rather, is a particular view of the normative implications of the facts of human nature adduced by anti-utilitarians, namely, that in terms of working out what is right, these considerations are secondary to the pursuit of overall utility, and defeasible by it. Whether one feels easy with this is of course another matter.

Conversely, though, in any arguing against hardline impartialism, we would need to face the extent to which demands such as those of utilitarianism are there in large part to address the biases and distortions inherent in the very kind of empirical psychological features that for instance Stocker and Williams think morality should make more central room for. In their view, if morality cannot make this adjustment (and they both raise this as a live option), then morality, at least in its current form, should be replaced by some other approach. But it is not clear that a morality that made easier room for individual projects to take precedence over the general good, or for people to favour those to whom they are in close relationships, would necessarily be an improvement over one that favours greater impartiality. It would have advantages impartialist approaches would not, for certain, but it would also have disadvantages of its own, and if bullets need to be bitten, it

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press

is not clear that biting the 'who needs impartiality' bullet is always the better option. Perhaps ideally some middle ground could be found that was not simply a muddled hybrid – but that goes well beyond the scope of the present discussion.[22]

There is certainly value to bringing considerations of human nature into ethical theory, and the question of what normative weight they should be given is a crucial and interesting one. Given the variety of notions of human nature adduced (descriptively and normatively) by advocates of such approaches though, it also seems some caution would be called for, and a clearer idea both of what the facts of human nature are (insofar as this can be ascertained) and of why they matter.[23]

srietti@uottawa.ca

[22] A commonly raised possibility, also considered by Stocker ('Schizophrenia', 'Emotional Identification'), is that impartialist ethics that abstract from the 'thicker' psychology may be more suited to institutional contexts and decision-making for large populations, while more thickly psychologically realistic approaches might be suited to a level of personal interaction and individual life-choices. But this in turn raises a problem about how cleanly the levels can be kept distinct in practice.

[23] This article builds on a presentation originally made to a panel on consequentialism for the Canadian Philosophical Association. Thanks are due to the panel members, the organizer and to the audience members for feedback and comments.

https://doi.org/10.1017/S0953820809990094 Published online by Cambridge University Press