

## Non-consequentialist utilitarianism

Nir Eyal • Harvard University • Draft paper for ISUS X • Please do not cite • [nir\\_eyal@hms.harvard.edu](mailto:nir_eyal@hms.harvard.edu)

### Abstract

Leading contemporary philosophers are confident that utilitarianism is a version of consequentialism. But it is not a version of consequentialism. *Utilitarianism* says that an act is morally right if and only if it maximises total utility. *Consequentialism* says that an act is morally right if and only if it maximises good. Total utility need not be thought to represent a good, or the only good. You may consistently hold that a certain act is right because it maximises utility, but that that act does not maximise good. If you do, then you may be a non-consequentialist utilitarian. Thus, non-consequentialist utilitarianism is a consistent philosophical stance, and utilitarianism is not simply a version of consequentialism. I make this point, address two replies, show that philosophers who recently referred to ‘non-consequentialist utilitarianism’ made different points, and note a paramount moral theory that may be utilitarian and non-consequentialist.

### Introduction<sup>1</sup>

Philosophers often say that utilitarianism is a version or a form of consequentialism. Shelly Kagan, for example, discusses the ‘consequentialist component of utilitarianism’ and says that utilitarianism is ‘the best known consequentialist theory.’ The statement ‘The paradigm

case of consequentialism is utilitarianism’ opens Walter Sinnott-Armstrong’s encyclopaedic entry on consequentialism. Sam Scheffler speculates that ‘utilitarianism refuses to fade from the scene in large part because...it is the most familiar consequentialist theory.’ For Bernard Williams, ‘Any kind of utilitarianism is by definition consequentialist.’<sup>2</sup>

Such statements often seek to clarify the relatively recent notion of consequentialism—the theory that we should maximise the goodness of consequences. Different versions of consequentialism, the clarification goes, differ on what constitutes a good consequence. One version takes utility to be the only good, and tells us to maximise utility. That version is utilitarianism. Other, less familiar versions of consequentialism take the good to coincide with things other than utility.

This article defends the surprising and seemingly-absurd view that, under conventional definitions, utilitarianism is not a version of consequentialism. Although some utilitarian outlooks are consequentialist, others aren’t.

Section I builds an initial case for this thesis. Sections II and III defend it from incompatible definitions of utilitarianism and consequentialism. Section IV shows that this thesis was not predated by three writers’ allusion to what they call ‘non-consequentialist utilitarianism.’ Section V exposes the non-consequentialist utilitarian commitments of a leading moral theory.

## **I. Why utilitarianism is not a version of consequentialism**

Consider Henry Sidgwick’s definition of utilitarianism and Sam Scheffler’s definition of consequentialism:

By Utilitarianism is here meant the ethical theory, that the conduct which,

under any given circumstances, is objectively right, is that which will produce the greatest amount of happiness on the whole.

Consequentialism provides a very simple *theory of the right*: an act is morally right (or morally permissible) if and only if it produces the best consequences.<sup>3</sup>

The following formulations seem in line with these conventional definitions:

*Utilitarianism:*

An act is morally right if and only if it maximises total utility.

*Consequentialism:*

An act is morally right if and only if it maximises good consequences.

(In these formulations, we could substitute ‘maximize’ by ‘satisfice’ or ‘promote’. We could substitute ‘action’ by ‘rule,’ ‘motive,’ ‘disposition,’ or ‘action, rule, motive, or disposition.’ We could define utility and good consequences indexically: relative to an agent or a moment (I assumed that they are agent-neutral and moment neutral). But such alterations in *both* definitions would not affect the argument that follows. For simplicity’s sake, I focus on establishing that utilitarianism, understood as maximising nonindexical act utilitarianism, is not a version of consequentialism, understood as maximising nonindexical act consequentialism. Parallel arguments would show that satisficing act utilitarianism is not a version of satisficing act consequentialism, and so forth.)

My argument to the effect that utilitarianism is not a version of consequentialism is simple. Utilitarianism tells you to maximise utility; consequentialism, to maximise good. You

may think that a high level of utility is not always good; that it is only part of the good, which sometimes comes at the expense of other important parts: beauty, excellence, rights fulfilment, desert, equality, species diversity...; or that while in actuality total utility coincides completely with good, there is a possible world in which they diverge. If you also think that, when they diverge, the right thing to do is to maximise utility, and not the good, then you are a non-consequentialist utilitarian.

Granted, the view that utility necessarily coincides with good is also consistent.

Utilitarians can be consequentialists. In fact, one argument for utilitarianism is that we should maximise good and that good and utility coincide, so we should maximise utility. But other arguments for utilitarianism are possible: that we should maximise utility out of equal respect and concern second-personally owed; that we should maximise utility out of personal virtues like compassion, care, and solidarity; or that utilitarianism generates intuitive judgments in the shallow pond case, in the simple trolley case, in catastrophe cases, and in enough other cases. Thus, not only is a utilitarian non-consequentialist stance coherent; some arguments for utilitarianism do not assume consequentialism.

Let me illustrate a possible non-consequentialist utilitarian stance. Imagine that justly punishing a certain evil assassin would decrease total utility. The punishment would harm the assassin, without benefiting anyone else. The assassin is broadly perceived as innocent. Her punishment would be seen as a fluke and lack deterrent or consoling effects. Furthermore, the assassin is now a quadriplegic, whose punishment would lack a preventive effect. A judge realising all that could at the same time acknowledge the good consequences of punishing the assassin in terms of increased proportion to desert in the world. An evil assassin would suffer in jail, and that would make the world more just, and in that respect, a little better than a similar world in which the assassin goes free.<sup>4</sup> The judge may further believe that punishing the assassin would generate more good consequences (in terms of proportion to desert) than

bad consequences (in terms of disutility). She does not, for example, believe that undeserved benefits constitute a grave harm to the beneficiary or to others.<sup>5</sup> Nevertheless, the judge may set the assassin free, out of utilitarianism: the judge believes that she should always maximise utility (perhaps as a matter of care and compassion) and that setting the assassin free would maximise utility. The judge thereby breaches consequentialist recommendations in the name of utilitarianism.

Or imagine a botanist who believes that preserving a certain obscure plant species would maximise diversity; that diversity is good; but that in this case maximising diversity would benefit no one; and that, because utilitarianism is true, it is not her duty to maximise diversity by preserving that plant species. The botanist could also be a non-consequentialist utilitarian.

The judge's and the botanist's assumptions that something (proportion to desert and species diversity) is good but that it does not provide them any moral reason for action (because it does not maximise utility) may seem odd. My goal is not to defend non-consequentialist utilitarianism, just to show that it is basically logically consistent and that an important philosopher may be committed to it. Some may argue, however, that this oddity is so extreme that, on judgment-internalism, no one can hold non-consequentialist utilitarianism: that no judge or botanist could hold all the beliefs that I imputed to them. However, some philosophers write that certain good prospects generate no moral reason for action. For instance, I take this to be the upshot of Frances Kamm's discussion of 'irrelevant utilities.'<sup>6</sup> Being honest and smart, Kamm presumably believes in what she writes. Since Kamm can believe that something is good without providing her any moral reason for action, so can the judge and the botanist.

Let me stress again that I am not defending non-consequentialist utilitarianism. It so happens that I am both a consequentialist and a non-utilitarian. My main thesis is that

utilitarianism is logically compatible with non-consequentialism. I also hope to show that a paramount political theory is utilitarian and non-consequentialist. Thus, non-consequentialist utilitarianism, a theory that I personally reject, is both consistent and important.

But first, let me address several objections to my claim that utilitarianism is logically compatible with non-consequentialism.

## **II. A different definition of utilitarianism?**

Some philosophers view utilitarianism as a version of consequentialism because they define utilitarianism differently than I proposed.

*Utilitarianism, according to a definition alternative to mine:*

1. An act is morally right if and only if it maximises total utility.
2. The reason for 1 is that
  - 2.1 An act is morally right if and only if it maximises good consequences.
  - 2.2 Good coincides with total utility.

Since 2.1 is identical to consequentialism, people who accept this alternative definition of utilitarianism take utilitarianism to be a version of consequentialism. According to their definition, utilitarianism conjoins consequentialism (2.1) with details that distinguish utilitarianism from other versions of consequentialism (1 and 2.2). Indeed, this seems to be Derek Parfit's reason for classifying utilitarianism under 'C,' which is his term for consequentialism:

To apply C, we must ask what makes outcomes better or worse. The simplest

answer is given by *Utilitarianism*. This theory combines C with the following claim: the best outcome is the one that gives to people the greatest net sum of benefits minus burdens... There are many other versions of C.<sup>7</sup>

However, this alternative definition is neither the traditional one nor the most helpful one. As philosophers traditionally use the term, utilitarianism contains only conjunct 1. Classical utilitarianism states that you should maximise utility; it does not state that you would thereby maximise good. To illustrate, note Bentham's, Mill's, and (again) Sidgwick's canonical statements of utilitarianism:

By the principle of utility is meant that principle which approves or disapproves of every action whatsoever, according to the tendency it appears to have to augment or diminish the happiness of the party whose interest is in question...

The creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness.

By Utilitarianism is here meant the ethical theory, that the conduct which, under any given circumstances, is objectively right, is that which will produce the greatest amount of happiness on the whole.<sup>8</sup>

These classical definitions identify right conduct with the promotion of happiness or utility. They do not identify it, additionally, with the promotion of good consequences. Those

contemporary writers who give that alternative definition break with the classical notion of utilitarianism. They often do so misleadingly, without acknowledging that they do so. The presumption is always in favour of keeping classical definitions in place, and to stipulate new meanings for familiar terms (openly and perhaps with the community's assent) only if the new definition represents a major improvement.

But incorporating a gloss (condition 2) into the familiar classical definition (condition 1) does not improve that definition. Usually it is best to keep definitions simple, and to introduce complexity only when a simpler definition is too crude to pick out a complex notion. This is not the case here, for typically the classical notion of utilitarianism has already been picked out when the complex definition is being offered. Presumably, the reason why contemporary writers offer the complex definition is that they thereby hope to situate utilitarianism in what they assume is its relation to other theories. But that compels us to accept their assumption on how utilitarianism relates to other theories if we are even to refer to utilitarianism. It is better to keep our definitions minimally committed, and to argue for determinate interrelations between notions separately. In our case, keeping the definition of utilitarianism minimally committed—maintaining the classical definition, as I do—is compatible with adding, if one so wishes, separately from the definition, that the best justification for utilitarianism is consequentialism.<sup>9</sup>

Perhaps as a result, some contemporary writers waver erratically between the alternative definition that they advance as though it were the classical one, and the quite different classical definition that still animates them. On the same page on which he expounds an alternative definition of utilitarianism, very much like Parfit's, Kagan says,

Historically, many utilitarians accepted hedonism [namely, the view 'that well-being consists solely in the presence of pleasure and the absence of



pain']<sup>10</sup> and so, utilitarianism is also known as the greatest happiness principle: the right act is the act that leads to the greatest total amount of happiness overall.<sup>11</sup>

If Kagan did not latently use the classical notion, then the conjunction of hedonism with utilitarianism would have yielded a definition with two necessary conditions: the greatest happiness principle *and* condition 2. That Kagan slips back to using the classical notion suggests that he too uses the classical definition of utilitarianism that I use, and not his own alternative definition.

### **III. A different definition of consequentialism?**

Others take utilitarianism to be a version of consequentialism because they define consequentialism differently than I proposed:

*Consequentialism, according to a definition alternative to mine:*

An act is morally right if and only if it maximises *a certain kind of thing* (but not necessarily good consequences).

*Consequentialism, according to a second alternative definition:*

An act is morally right if and only if it maximises *a certain kind of consequence* (but not necessarily a good one).

As Thomas Pogge pointed out to me, if such alternative definitions of consequentialism are accurate, then utilitarianism is a version of consequentialism, after all. There is no denial

that total utility is ‘*a certain kind of thing*,’ and that the level of total utility that an act generates is ‘*a certain kind of consequence*.’ Hence, these alternative definitions would make it silly to endorse utilitarianism while opposing consequentialism. They would justify dubbing utilitarianism a version of consequentialism.

The first alternative definition was proposed to me by Pogge. The second alternative definition is in the spirit of definitions by Jonathan Dancy (quoted below), by Philip Pettit<sup>12</sup> (who also gives other definitions),<sup>13</sup> and by Walter Sinnott-Armstrong.<sup>14</sup> However, as I move to argue, these two alternative definitions are inaccurate. Consequentialism tells you to maximise good consequences, not to maximise *whatever*, or *whatever consequence*. If a theory that tells you to maximise something, say, utility or equality, is consequentialist, then that theory must assume that maximising it maximises good consequences. Otherwise that theory is not consequentialist.

Consider Shelly Kagan’s definition of consequentialism. For Kagan, consequentialism is the theory that ‘an act is morally right (or morally permissible) if and only if it produces the best consequences.’ Note that Kagan focuses on the best consequences—neither on an empty place holder nor on a type of consequence that need not be good. Most contemporary authors focus similarly, and the alternative definitions are nearly idiosyncratic.<sup>15</sup>

Why adopt the majority’s definition of consequentialism and not the two alternative definitions? There are two indications that defining consequentialism in terms of *good consequences*, as Kagan, most contemporary philosophers and I do, dovetails with a shared latent notion of consequentialism better than these alternative definitions do.

One indication is that, on these alternative definitions of consequentialism, important ethical theories make too little sense. For example, John Rawls observes that consequentialism (or ‘teleology’) subjects the right to the good. If consequentialism subjected right conduct only to maximising *something*, or to bringing about *some* consequence, then it

clearly wouldn't necessarily subject the right to the *good*. For *something* and *consequences* can designate things other than *good* consequences. Rawls's casual observation would be out rightly false. (It might be thought that Rawls in this discussion uses 'good' simply for *something* that you can maximize. But he does not.)<sup>16</sup>

A second indication is that even the few authors who use the two alternative definitions of consequentialism inadvertently slip back to a definition like Kagan's and mine. For example, Dancy's definition of consequentialism makes no mention of the goodness or the value of consequences: 'Consequentialism claims that we assess the moral worth of an action by appeal to its consequences—to the difference it makes to the world that the action was done.'<sup>17</sup> It resembles the second alternative definition just mentioned. Immediately, however, Dancy elaborates that consequentialism is 'flexible about what is to count as a 'consequence'...So a consequentialist need not deny the existence of *value* in an action that is an expression of a deeply felt personal commitment...a consequentialist may adopt a theory of *value* under which the world is a *better* place for having such expressions of personal commitment going on in it.'<sup>18</sup> Contrary to first appearances, then, Dancy presupposes that, for consequentialists, the consequences to be maximized are only those laden with 'value,' those under which the world is a 'better' place. Dancy may toy with a new definition of consequentialism (in the first excerpt), but our shared notion of consequentialism surfaces in his wording (in the second excerpt).

If you remain unconvinced that the notion of consequentialism focuses on good consequences, you may reject my main thesis, that utilitarianism is not a version of consequentialism. Still, you can accept that *if* we define consequentialism with a focus on good consequences, as Kagan, Sen, Scheffler, Parfit, Darwall, and others do, *then* utilitarianism is not a version of consequentialism. This would also be an interesting conclusion, because these astute thinkers usually take utilitarianism to be a version of

consequentialism.

#### **IV. Alleged non-consequentialist utilitarians**

Daniel Jacobson, Judith Thomson and Will Kymlicka have labelled certain moralities ‘non-consequentialist utilitarian’ or ‘deontological utilitarian.’ However, these moralities are not non-consequentialist utilitarianism in *my* sense.

According to Daniel Jacobson, John Stuart Mill displays ‘utilitarianism without consequentialism.’ In a 2003 paper, Jacobson says that Mill’s book *Utilitarianism* is ‘ecumenical’ and extremely non-committal on the precise nature of the utilitarianism that it espouses. At points, Mill is so vague that he could be read as non-consequentialist.<sup>19</sup> Jacobson shows that Mill’s masterpiece can be vague and inconsistent, but little in Jacobson’s early argument forces us to ascribe to Mill a determinate position that is both non-consequentialist and utilitarian. By contrast, I argue that there is a possible position that is both utilitarian and non-consequentialist. Therefore, if utilitarianism excluded non-consequentialism, in the way that greenness excludes redness, then the early Jacobson would be saying that a certain apple is red and green in different parts (so to speak); I would be saying that apples can be red and green all over. In that way, my argument is more ambitious.

Jacobson’s 2008 article is also more ambitious: ‘there is no paradox involved in claiming that there is logical space for a utilitarian theory that rejects consequentialism, and there is considerable evidence for ascribing such a view to that most renowned, though not most orthodox, utilitarian, John Stuart Mill.’ Jacobson’s interpretation is now that ‘In Mill’s view, morality does not treat everyone’s happiness in exactly the same way (as deontic impartiality [essential to consequentialism—*AUTHOR’S INITIALS*] demands) even though everyone’s happiness is of equal value (as axiological impartiality [sufficient for

utilitarianism—*AUTHOR'S INITIALS*] requires).’ Specifically, ‘Mill expressly rejects deontic impartiality by claiming that self-regarding but harmful acts are not amenable to moral disapprobation and that we cannot be compelled for our own good.’<sup>20</sup> My own definition of utilitarianism differs from Jacobson’s. For me, utilitarianism is a theory of the right. It is not a theory of the good, of what has value and how much value it has. My definition strikes me as both more standard and closer to Mill’s (for whom ‘actions are *right* in proportion as they tend to promote happiness’—my italics). But if I am wrong on this exegetical point, it still remains the case that my thesis differs considerably from Jacobson’s. The logical space I invoke for what I call ‘non-consequentialist utilitarianism’ does not revolve around impartiality and the possibility of agent-relative duties and options, central to Jacobson’s new thesis.

Judith Thomson defines:

NON-CONSEQUENTIALIST ACT UTILITARIANISM: X ought to do alpha if and only if the act-plus-consequence-set of X’s doing alpha would be better than the act-plus-consequence-sets of X’s doing any of the other things it is open to X to do instead... [Non-consequentialist act utilitarianism ‘tells us to attend [not only to the consequences but] also to the acts themselves—to X’s doing alpha itself [insofar as acts have good or bad intrinsic value].’<sup>21</sup>

My notion of non-consequentialist (act) utilitarianism is different. For me, the overall consequences of an act include among other things the fact that that act was performed. What Thomson calls non-consequentialist act utilitarianism pretty much corresponds to what I would call consequentialism. That something is non-consequentialist in Thomson’s sense does not make it non-consequentialist in my sense. As I said above, my definition of

consequentialism is the standard one. Unlike Thomson, I do not re-define consequentialism. I seek to point out a logical possibility within existing definitions. In that respect, my project is more ambitious.

Will Kymlicka expounds a utilitarian outlook that he calls ‘deontological:’

On one interpretation utilitarianism is...a moral theory because it purports to treat people as equal, with equal concern and respect. It does so by counting everyone for one, and no one for more than one...The problem, on this interpretation of utilitarianism, is how to treat distinct people fairly. The standard solution is to give each person’s interests equal weight. Each person’s life matters equally, from the moral point of view, and hence each person’s interests deserve equal consideration... If we decide how to act on this basis, then...[m]aximization occurs, but as a by-product of a decision-procedure that is intended to aggregate people’s preferences fairly.<sup>22</sup>

Kymlicka holds important utilitarian theories to be deontological in his sense:

...it is the concern with equal consideration that clearly underlies Bentham’s argument and is explicitly affirmed by recent utilitarians such as John Harsanyi and James Griffin. And while this is not his preferred method, R. M. Hare too claims that one could defend utilitarianism by reference to a foundation premise of equal consideration.<sup>23</sup>

It may seem as though, on Kymlicka’s interpretation, thinkers like Bentham, Harsanyi, Griffin and Hare are non-consequentialist utilitarians in my sense. But they are not. In order

for Kymlicka's interpretation to show that they are non-consequentialist, consequentialism would have to mean something different than it means to me, such as:

*Consequentialism, according to a third alternative definition:*

An act is morally right if and only if *and because* it maximises good consequences.

The third alternative definition resembles the one that I use above, except that it incorporates 'and because.' It assumes that, for consequentialists, it is the contribution to the best outcomes that makes acts right (not the other way around, say). As Kymlicka interprets them, these utilitarian thinkers could accept that right acts always maximize good; they could be consequentialists in my sense. They affirm deontology only in the sense of insisting that maximization of good 'occurs, but as a by-product:' neither as the necessary motivation nor as the justifying ground of the act. Maximization of good coincides with right action, but it is not what makes actions right.

The third alternative definition, which Kymlicka uses, is fairly intuitive,<sup>24</sup> and some recent definitions of consequentialism incorporate similar 'and because' clauses.<sup>25</sup> However, it is not the definition that we started out with; Kymlicka has not indicated any theory, historical or possible, that is non-consequentialist utilitarian in the sense given above. My own thesis is in that respect more ambitious.

Furthermore, Kymlicka's emphasis on the justifying ground would, consistently applied, make Bentham, Harsanyi, Hare, and so forth neither consequentialist nor utilitarian. Why? Because on Kymlicka's interpretation these thinkers consider acts right neither because these acts maximise good consequences nor because they maximise utility. Acts are right for them because, by maximising good consequences and utility, these acts display equal consideration for individuals. The ultimate commitment of these thinkers, on Kymlicka's

interpretation, is toward individuals, neither toward good consequences nor toward total utility or the social collectives that bear it; the result—increase both in good consequences and in collective utility—is not the source of rightness.

In other words, Kymlicka determines ethicists' positions by checking what these ethicists see as the justifying grounds of action. Therefore, a consistent Kymlicka would have to use:

*Utilitarianism, according to a second alternative definition:*

An act is morally right if and only if *and because* it maximises total utility.<sup>26</sup>

The problem for Kymlicka is that on *that* definition, Bentham, Hare, and so forth are not utilitarian either. For they deem acts rights not because these acts maximise (collectives') total utility, but because they treat each individual with equal consideration. Utilitarianism is not for these thinkers a justifying ground, not even as Kymlicka interprets those thinkers. Being non-utilitarian, these thinkers could not be non-consequentialist utilitarian.

## **V. A real non-consequentialist utilitarian?**

Surprisingly, perhaps the best example of a non-consequentialist utilitarian is John Rawls. Or so I wish to suggest.

That Rawls is a non-consequentialist is plain and generally accepted. For Rawls, basic political institutions ought to treat citizens with justice, as free and equal, and so to honour fair decisions: ones that hypothetical representatives, standing as equals behind a veil of ignorance, would make freely. The reason to honour their decisions is not that we should maximize good. Rather, political institutions must honour these decisions because justice is



the first virtue of political institutions, and justice demands compliance with these decisions. Nor does honouring the representatives' decisions always bring about the agent-neutrally best consequences. Rawls insists that political institutions must honour the representatives' decisions across the board, even when doing so does not bring about the best consequences, for example, in terms of improving animal welfare<sup>27</sup> and human excellence.<sup>28</sup>

But is Rawls really committed to utilitarianism on the state level, as I wish to argue? Rawls emphatically opposes his theory to utilitarianism! Nevertheless, and despite protests from Rawlsians, astute observers Harsanyi, Arrow, Hare and others make a compelling case that the commitments of Rawlsian theory for the state are utilitarian.<sup>29</sup>

As these observers show, the parties in a Rawlsian original position will vote for results that maximise expected social utility. Why? Because, ignorant of the determinate social position of the individual she represents, each representative must vote for the arrangement that maximises social utility. For this is the arrangement most likely to promote that individual's interests. Therefore, all parties vote for maximal social utility; honouring their decision coincides fully with honouring the utilitarian duty to promote social utility.

It is true that, in Rawls's system, the parties assign special importance to the 'higher-order' interests of the individuals they represent: their interests in preserving their two moral powers—the power to form, pursue, and revise plans and the power to treat others' pursuits with justice. But the parties prioritise these interests precisely because higher-order interests are especially strong *interests*, and fulfilling them is especially good for the individuals whom they represent on a 'full' conception of their good.<sup>30</sup> It is only because the 'overall aim [of the parties] is to fulfill their responsibility and to do the best they can to advance the determinate good of the persons they represent'<sup>31</sup> that the parties focus chiefly on these relatively strong interests. Thus, when the parties prioritise the higher-order interests over other interests, the result of their choice remains the maximization of social utility.

It may also seem as though Rawlsian parties vote against a system that ‘sacrifices’ the good of individuals on the altar of maximal social utility. Famously, the parties elect maximin, not maximum; and they protect individual liberties and rights as bulwarks against society. However, the parties do so primarily for risk-aversion or uncertainty-aversion (and not, say, because God forbade sacrificing any of Her children). As Michael Smith once put this point in a conversation, such aversions toward risk or uncertainty are *aversions* (read with Australian accent). Such second order *aversions* are apparently stronger than first order plans, in Rawls’s view, so his parties *maximise* individuals’ fulfilment of first-order plans. But they do so only because maximin on that superficial level will maximise expected utility, on a deeper level. In voting against the sacrifice of anyone’s plans, basic political liberties and basic income on the altar of social interests, the parties maximise properly understood social utility.<sup>32</sup>

Admittedly, another reason that Rawlsian parties espouse maximin and individual liberty is that these principles express respect for citizens as free and equal, and thus maintain the ‘social bases of self-respect.’ But even this Kantian-ringing reason is plausibly seen as utilitarian. In Rawls’s system, self-respect is protected because it is a primary good—only in virtue of its instrumental value for the effective pursuit of plans (and not, say, because protecting it is otherwise a duty toward oneself), a point made by several interpreters, myself included.<sup>33</sup> The promotion of the effective pursuit of plans may, of course, constitute a utilitarian goal.

A final objection is that Rawls advances the publicity condition as a counter to utilitarianism. Roughly, the publicity condition states that all justified collective policy could be publicized without loss in efficiency. Some policies that appear to maximise utility require deceit and misinformation, and the publicity condition would rule them out. It may seem as though Rawls’s championship of the publicity condition makes him a clear non-utilitarian.

Nevertheless, in Rawls's own system, the publicity condition, far from being anti-utilitarian, is the handmaiden of utilitarianism. It serves the deeper utilitarian project of maximising utility over time.<sup>34</sup> The parties value publicity precisely because it enhances long-term stability, which they value because it enables citizens to reap the fruits of a well-ordered political system over time. Since the fruits of that system are utilitarian—as I suggested above—publicity has clear utilitarian value. It maximises social utility over time. Admittedly, full publicity would undermine some policies that utilitarians would have recommended if they overlooked the utilitarian value of publicity. But that just shows that utilitarianism does not recommend all policies that it initially appears to recommend.

The decisions of the parties in the original position may at first appear non-utilitarian, but in sophisticated ways, their decisions maximise utility. If and when the derivation of Rawlsian principles for the state is sound, then sound utilitarianism would generate the same principles for the state, or so I would argue. Furthermore, it seems fair and charitable to interpret Rawls taking the core foundations of his framework—the original position and the veil of ignorance—as our fixed points. Sound Rawlsian state-level<sup>35</sup> theory is therefore both non-consequentialist and utilitarian.

## **Conclusion**

Although a utilitarian view can be consequentialist, it need not be consequentialist, not by definition. We can coherently hold that right acts maximise utility without holding that they maximize good. Despite what many authors take to be an obvious truth, utilitarianism is not a version of consequentialism. Indeed, the most influential political theory of our time has strong utilitarian and non-consequentialist commitments.

## References

- Anscombe, E. 1958: Modern Moral Philosophy. *Philosophy* 33: 1–19
- Arrow, K. J. 1973: Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice. *Journal of Philosophy* 70 (9): 245–63
- Bentham, J. 1996 [1781]: *An Introduction to the Principles of Morals and Legislation*. Ed. by J. H. Burns & H. L. A. Hart. Oxford: Oxford UP
- Broome, J. 1995: *Weighing Goods*. Oxford: Blackwell
- 1999: *Ethics out of Economics*. Cambridge: Cambridge UP
- Carruthers, P. 1992: *The Animals Issue: Morality in Practice*. Cambridge: Cambridge UP
- Cohen, G. A. forthcoming: *Saving Justice from Constructivism*
- Dancy, J. 1993: *Moral Reasons*. Oxford: Blackwell
- Darwall, S. 2003: Introduction. In Darwall (ed., 2003): 1–8
- Darwall, S. (ed.) 2003: *Consequentialism*. Oxford: Blackwell
- Feldman, F. 1997: *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*, Cambridge: Cambridge UP
- Frankena, W. 1973 [1963]: *Ethics*, second edition. Englewood Cliffs: Prentice Hall
- Freeman, S. 1994: Utilitarianism, Deontology, and the Priority of Right, *Philosophy and Public Affairs* 23 (4): 313–49
- Frey, R. G. 1984: Introduction: Utilitarianism and Persons. In Frey (ed., 1984): 1–19
- Frey, R. G. (ed.) 1984: *Utility and Rights*. Minneapolis: University of Minnesota Press
- Griffin, J. 1986: *Wellbeing*. Oxford: Oxford UP
- 1992: The Human Good and the Ambitions of Consequentialism. *Social philosophy and Policy* 9: 118–32
- Hare, R. M. 1973: A Review of Rawls' A Theory of Justice. *The Philosophical Quarterly* 23

- (91): 144–55
- 1984: Rights, Utility and Universalization: Reply to J. L. Mackey. In Frey (ed., 1984): 106–20
- Harsanyi, J. C. 1953: Cardinal Utility in Welfare Economics and in the Theory of Risk-taking. *The Journal of Political Economy* 61 (5): 434–5
- 1955: Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *The Journal of Political Economy* 63 (4): 309–21
- Hutcheson, F. 1971 [1725]: *Inquiry into the Origins of our Ideas of Beauty and Virtue*, in *Two Treatises*, second edition. New York: Garland Publishing
- Jacobson, D. 2003: J.S. Mill and the Diversity of Utilitarianism. *Philosophers' Imprint* 3 (2): 1-18. Available at [www.umich.edu/~philos/Imprint/frameset.html?browse](http://www.umich.edu/~philos/Imprint/frameset.html?browse)
- 2008: Utilitarianism without Consequentialism: The Case of John Stuart Mill. *Philosophical Review* 117 (2): 159-191
- Kagan, S. 1989: *The Limits of Morality*. Oxford: Clarendon Press
- 1998: *Normative Ethics*. Boulder, Colorado: Westview
- Kamm, F. M. 1993: *Morality, Mortality*, Vol. I. New York: Oxford UP
- Kymlicka, W. 1988: Rawls on Teleology and Deontology. *Philosophy and Public Affairs* 17 (3): 173–90
- Lyons, D. 1965: *Forms and Limits of Utilitarianism*. Oxford: Clarendon Press
- Mill, J. S. 1990: On Utilitarianism. In his *Utilitarianism/On Liberty/Essay on Bentham*. Ed. by M. Warnock. Glasgow: Fontana Press: 251–321
- Moore, G. E. 1971: *Principia Ethica*. Cambridge: Cambridge UP
- Nagel, T. 1972: *Philosophy and Public Affairs* 1 (2): 123–44
- Parfit, D. 1986: *Reasons and Persons*. Oxford: Clarendon Press
- 2003: Justifiability to Each Person. *Ratio* 16 (4): 368–90

- Pettit, P. 1991: Consequentialism. In P. Singer (ed.), *A Companion to Ethics*. Oxford: Blackwell: 230–40
- 1993: Introduction. In P. Pettit (ed.), *Consequentialism*. Aldershot: Dartmouth Publishing Company: viii–xix
- Pettit, P. & Smith, M. 2000: Global Consequentialism. In B. Hooker et al (eds.), *Morality, Rules, and Consequences*. Edinburgh: Edinburgh UP: 121–33
- Pogge, T. 1995: Three Problems with Contractualist-Consequentialist Ways of Assessing Social Institutions. *Social Philosophy and Policy* 12 (2): 241–66
- 2004: Equal Liberty for All? Trans. by D. Tulodziecki. *Midwest Studies in Philosophy* 28: 266–81
- Portmore, D. W. 2005: Combining Teleological Theories with Evaluator Relativism: A Promising Result. *Pacific Philosophical Quarterly* 86 (1): 95–113
- Rawls, J. 1999: *A Theory of Justice*, revised edition. Cambridge, MA: Harvard UP
- Savulescu, J. 2002: The Embryonic Stem Cell Lottery and the Cannibalization of Human Beings. *Bioethics* 16 (6): 508–29
- Scheffler, S. 1988: Introduction. In S. Scheffler (ed.), *Consequentialism and its Critics*. Oxford: Oxford UP: 1–13
- 1994: *The Rejection of Consequentialism*, revised edition. Oxford: Oxford UP
- Sen, A. K. 1993: Utilitarianism and Welfarism. Reprinted in Darwall (ed., 2003): 259–85
- Sidgwick, H. 1981 [1907]: *Methods of Ethics*, seventh edition. Indianapolis: Hackett
- Sinnott-Armstrong, W. 2003: Consequentialism. *Stanford Encyclopedia of Philosophy*. Available online at <http://plato.stanford.edu/entries/consequentialism/>
- Stark, C. A. 1998: Self-Respect. *Routledge Encyclopaedia of Philosophy*. Available online for subscribers
- Temkin, L. 1994: Weighing Goods: Some Questions and Comments, *Philosophy and Public*

*Affairs* 23 (4): 350–380

— 2003: Egalitarianism Defended. *Ethics* 113 (2003): 764–82

Thomson, J. 1990: *The Realm of Rights*. Cambridge, MA: Harvard UP

Williams, B. 1973: A Critique of Utilitarianism. In J. J. C. Smart & B. Williams (eds.),  
*Utilitarianism—For and Against*. Cambridge: Cambridge UP: 77–150

---

1 I wish to thank Iwao Hirose, Thomas Pogge and Alex Rajczi for helpful comments.

2 Kagan (1998: 62); Sinnott-Armstrong (2003); Scheffler (1994: 4); Williams (1973: 79). See also Scheffler (1988: 2); Rawls (1999: 26); Frankena (1973: 34); Parfit (1986: 26); Frey (1984: 4); Dancy (1993: 167); Darwall (2003: 5); Pettit (1993: viii).

3 Sidgwick (1981: 411); Scheffler (1988: 1).

4 Larry Temkin argues convincingly that desertarian justice and luck-egalitarian justice make states of affairs better than similar unjust ones (Temkin [1994: 353ff.]; Temkin [2003]).

5 Fred Feldman comes close to voicing such a view: ‘when persons do not deserve a certain good but get it anyway, then it is not very good for them to receive it...’ (1997: 11). Hegel believed that letting off convicted criminals—treating them better than they deserve—harms them by treating them as though they had not been responsible for their criminal actions. John Broome (1995: ch. 9, and 1999: p. 207) argues that receiving less than one’s fair equal share harms one. On the other hand, Temkin (Ibid.) is clear that the badness of disproportion to desert is not a matter of disutility for any individual.

6 Kamm (1993: ch. 8). For Kamm, in certain situations, we have no moral reason to produce certain utilities. Since she takes this point to undermine consequentialism, she seems to hold that producing these utilities would have maximized good.

7 Parfit (1986: 26. But see his recent concession in support of the view that I defend [2003: 370–1]). Similarly, for Amartya Sen, ‘A utilitarian structure consists of the central element of outcome utilitarianism combined with some consequentialist method of translating judgment of outcomes into judgments of actions’ (Sen [1993: 261]. See also pp. 263, 279); For Kagan, all versions of consequentialism ‘share the basic

---

consequentialist claim that an act is right if and only if it best promotes the good, but they differ insofar as they incorporate different theories of the good. The most famous consequentialist theory is utilitarianism, which is the result of combining consequentialist with welfarism' (Kagan [1998: 61]. See also p. 215). This view is also shared by Griffin (1992: 119–20), Darwall (2003: 5), Dancy (1993: 167), Frey (1984: 4), and Scheffler (1988: 2).

8 Bentham (1996: 11–2); Mill (1990: 257); Sidgwick, op. cit. See also Hutcheson (1971: 177–8): 'An action is best, which procures the greatest happiness for the greatest numbers; and that, worst, which, in like manner, occasions misery.'

9 Nagel (1972: 125), Frankena (1973: 34–5), and Lyons (1965: vii, 9) offer a different challenge to my definition of utilitarianism. They define utilitarianism in the way that I propose to define consequentialism. According to Nagel, for example, 'Utilitarianism says that one should try... to maximize good and minimize evil....' I take such different definitions to reflect merely philosophers' confusion of utilitarianism with consequentialism prior to the mid-Seventies, when the two concepts were separated sharply. Earlier, from the Sixties onwards, the need for a concept that subjects the right to the good became evident, but 'consequentialism' was not yet in wide circulation. During that period, philosophers often casually used 'utilitarianism' to fill that need. This is clear in the case of Frankena. He says that utilitarianism is only one kind of 'teleological' theory; but his definitions of utilitarianism (pp. 34–5) and of teleological theories in general (pp. 14–5) are very similar. Both theories tell us to bring about 'the greatest balance of good over evil.'

10 Kagan (1998: 31)

11 Kagan (1998: 62)

12 'Roughly speaking, consequentialism is the theory that the way to tell whether a particular choice is the right choice for an agent to have made is to look at the relevant consequences of the decision: to look at the relevant effects of the decision on the world' (Pettit [1993: viii]). Note, however, that Pettit's definition begins with the words 'Roughly speaking.'

13 Pettit sometimes defines consequentialism in terms of promoting valuable things (in general), not good consequences: 'Consequentialism is the view that whatever values an individual or institutional agent adopts, the proper response to those values is to promote them. The agent should honour the values only so far as honouring them is part of promoting them, or is necessary in order to promote them' (Pettit [1991: 231]. See



---

also Pettit & Smith [2000: 121]). This definition of consequentialism focuses on values not specifically on good consequences. Hence, Consequentialism, according to a fourth alternative definition (the third one is in the next section): An act is morally right if and only if it maximises value (but not necessarily good consequences). Utilitarianism is not a version of consequentialism on this definition either; for one may hold that utility must always be maximized although it is not always of value. In addition, it is not clear how literally Pettit would like us to take this definition. Pettit & Smith equate it with Parfit's (quoted above), which explicitly defines consequentialism in terms of good consequences. Thus, Pettit may tacitly assume that all valuable things are good consequences.

14 Sinnott-Armstrong (2003)

15 Kagan (1998: 61). See also Kagan (1989: xi, 8), Freeman (1994: 313), Scheffler (1994: 1–2), and the full quote from Scheffler (1988: 1): 'Consequentialism in its purest and simplest form is a moral doctrine which says that the right act in any given situation is the one that will produce the best overall outcome, as judged from an impersonal standpoint which gives equal weight to the interests of everyone...The idea is that 'whatever people ought to do is to minimize evil and maximize good, to try, in other words, to make the world as good a place as possible.' See also Sen (1993: 260): 'Act consequentialism: An action  $\alpha$  is right if and only if the state of affairs  $x$  resulting from  $\alpha$  is at least as good as each of the alternative states of affairs that would have resulted respectively from the alternative feasible acts.' Among historical thinkers, Anscombe (1958: 12) coins 'consequentialism' and presumably identifies it with the view that 'the right action is the action which produces the best possible outcomes' (p. 9). Earlier, Moore notes our duty to perform 'that action, which will cause more good to exist in the Universe than any possible alternative.' (Moore [1971: 148]).

16 As is evident when Rawls analyzes why teleology is initially attractive. On his account, we initially feel that 'rationality is maximizing something and...in morals it must be maximizing the good.' (Rawls (1999: 22)) Rawls cannot be stating that we feel that rationality is maximising something and in morals it must be maximising something.)

17 Dancy (1993: 167)

18 Ibid. My italics.

19 Jacobson (2003: esp. pp. 8, 14–5)

20 Jacobson (2008)

21 Thomson (1990: 129ff.)

---

22 Kymlicka (1988: 177–8).

23 Ibid, pp. 176–7. Kymlicka cites Bentham (1996: 26–7), Harsanyi (1955: 315–6), Griffin (1986: 167–70, 208–15, 239–42, 295–301), and Hare (1984: 106–12). Kymlicka himself does not endorse utilitarianism, not even on that ‘deontological’ interpretation.

24 For example, consider the philosophical position that right acts are right because they are virtuous and that, unbeknownst to agents, God rewards right action by making the entire world instantaneously better, something that God never does otherwise. Intuitively, that position could be non-consequentialist, although it entails that right acts, and only right acts, maximise good consequences. What seems to be missing is the implication that these acts are right because they maximise good, as opposed to maximising it merely as a side effect of being right.

25 See, e.g., ‘consequentialism is an agent-neutral theory that takes all actions to be permissible (or impermissible) purely in virtue of the value of their resultant states of affairs.’ (Portmore [2005: 95]. My italics). See also Kagan’s definition of foundational (as opposed to factorial) consequentialism. (Kagan [1998: 212ff.]), and Parfit’s definition of consequentialism as (fundamentally) the claim that ‘There is one ultimate moral aim: that outcomes be as good as possible.’ (Parfit 1986: 24). Conceivably, earlier definitions did not incorporate the ‘and because’ clause only since appreciation for the importance of such clauses is a recent development in ethics. As Iwao Hirose suggested to me, such appreciation became standard only during recent debates on whether Scanlonians, who hold that right acts are right because they are not reasonably rejectable, can hold the following view: that some acts are not reasonably rejectable because they are right.

26 It is also worth noting here Sam Freeman’s criticism of Kymlicka’s interpretation of Hare and these other utilitarian thinkers. Freeman (1994) says that these thinkers espouse impartiality between the utilities of distinct people merely as a principle of rational deliberation, and not at all as a matter of equal concern and respect for persons or some other moral principle of right action. My own argument does not rest on any exegetical commitments.

27 For that reason, justice as fairness ‘would seem to include only our relationships with other persons and to leave out of account how we are to conduct ourselves toward animals and the rest of nature.’ (Rawls [1999: 15]). Rawls concedes on the same page that these relationships are also important. They lie only outside the scope of his book on justice. However, the book famously starts with the claim that for political institutions justice is the first priority. Carruthers (1992: ch. 5) argues that in Rawls’s contractualist framework animals

---

must lack intrinsic moral status.

28 Rawls (1999: 289f.) accepts that human excellence may be intrinsically good, and that maximising it would be a different policy than the one dictated by his contractualist system. But he thinks that we should follow his contractualist system.

29 Harsanyi (1953); Harsanyi (1955: 316), Arrow (1973); Hare (1973); Pogge (1995); Pogge (2004); Savulescu (2002).

30 Rawls (1993: 106–7, 202–3); Freeman (1994: 346 )

31 Rawls (1993: 307). See also pp. 74–7, 105–6.

32 For related positions, see Cohen (in progress); {SELF-IDENTIFYING REFERENCE}.

33 For Rawls (1999: 386), without self-respect ‘...we cannot ... continue in our endeavors’ and that is what makes it ‘...clear why self-respect is a primary good...’ See also Rawls (1993: 318). A Rawlsian ‘political society is a good for citizens in that it secures for them the... social bases of self-respect.’ (Rawls [1993: 203]. My italics). For a similar interpretation see Stark (1998: 1) and my {SELF-IDENTIFYING REFERENCE}.

34 E.g. Rawls (1999: §69, §82)

35 Things become more complicated outside the state. I have not examined whether a consistent Rawls would be a non-consequentialist utilitarian on the global level.