Person-affecting utilitarianism

Ralf M. Bader
Merton College, University of Oxford

> "The mere notion of *amount* lets philosophers introduce a surrogate for the proper notion of utility – it gives them utilities which are not *someone's*, in the form of quanta of happiness which nobody has but which someone could have. As well as deploring the situation where a person lacks happiness, these philosophers also deplore the situation where some happiness lacks a person."
>
> (Bennett: 1978, pp. 63-64)

1 Introduction

Total utilitarianism is meant to be one of the most straightforward moral theories. It assesses the goodness of states of affairs in terms of how much utility they contain. One state of affairs is judged to be better than another iff the former contains a greater amount of utility (= happiness) than the latter. The betterness ordering is strongly separable across persons such that the contribution that the utility of an individual (or a plurality of individuals) makes to the goodness of the state of affairs is independent of the utility of other people. This ensures that the betterness ordering has an additive representation, such that goodness can be aggregated additively. The overall value of a distribution is the sum of the values of the members of the distribution.

$$V(D) \geq V(D') \text{ iff } \sum_{x \in D} V(x) \geq \sum_{x \in D'} V(x)$$

The consequentialist connection between goodness and rightness (which consists in a maximising function), then ensures that an action is right iff it brings about at least as much utility (or, in the context of uncertainty, expected utility) as all other available alternatives, whereby action ϕ is at least as good as ψ iff the state of affairs that ϕ-ing brings about is as at least as good as that which results from ψ-ing, i.e. $D_\phi \geq D_\psi$.

$$C(X) = \{\phi : \phi \in X \wedge \forall \psi \in X(\phi \geq \psi)\}$$

Total utilitarianism has a very simple value function, namely $V(D) = \Sigma_{x \in D} V(x)$. This function, however, is in an important sense schematic and underspecified. The problem is that it does not specify what type of value is at issue. That the value of a distribution is the sum of the values of the members of the distribution only specifies the form of the value function, namely that it is an additive function. It does not specify the type of value to which the function applies. We need to be told what kind of value is aggregated. What is it that is being added? Is it the personal goodness of the well-being of different individuals, or is it the impersonal goodness of their well-being? Whether it is personal good or impersonal good determines whether one is dealing with person-affecting utilitarianism or with impersonal utilitarianism.

Total utilitarianism has been much criticised for being an impersonal theory. Most notably, it has been objected that it treats persons as mere containers of utility that are replaceable and do not matter in their own right. For this reason, it has frequently been contrasted with a person-affecting approach that is concerned not with impersonal aggregates but with how persons are affected. Once we distinguish impersonal from person-affecting construals of utilitarianism, we can see that these criticisms do not apply to total utilitarianism per se but only to impersonal versions.

The first part of this chapter argues that impersonal versions of utilitarianism are objectionably impersonal. The problem is that they do not take personal good seriously. They do not attribute ethical significance to personal good, but instead (at best) only consider it to be ethically relevant. Although they satisfy a weak (information-theoretic) form of welfarism, impersonal versions do not satisfy a stronger form of welfarism according to which what is good for persons is of ethical significance. This means that they are not operating with the 'proper notion of utility', but with a spurious surrogate that leads us to 'deplore the situation where some happiness lacks a person'. As a result, they end up sub-ordinating and sacrificing personal good for the sake of impersonal good and thereby treat persons as mere containers of impersonal good, which gives rise to troubling implications in the case of variable-population comparisons.

By contrast, person-affecting versions are not objectionable in this way. They operate with the proper notion of utility and assign intrinsic ethical significance to personal good. Instead of being concerned with impersonal aggregates, they are concerned with how persons are affected. The second part evaluates the prospects for person-affecting versions of utilitarianism. It will be argued that person-affecting total utilitarianism presupposes comparativism, i.e. that existence and non-existence are comparable in terms of personal good, and as such involves problematic metaphysical commitments. Accordingly, same-number person-affecting utilitarianism turns out to be the only version of utilitarianism that neither involves an objectionable axiology nor requires problematic metaphysical commitments.

2 Impersonal utilitarianism

Impersonal utilitarianism is concerned with the impersonal value of the well-being of the various members of the distribution.

$$V_{\text{impersonal}}(D) = \sum_{x \in D} V_{\text{impersonal}}(x)$$

This approach comes in two forms, depending on whether only impersonal good is countenanced (monistic theories) or whether personal good is also included in the axiology (hybrid theories).

2.1 Monistic theories

The simplest (and most objectionable) version of impersonal utilitarianism is committed to a monistic theory of the good. Defenders of this approach take impersonal good to be basic and try to account for everything else in terms of it. Instead of recognising an independent notion of personal good, they merely countenance empirical (non-evaluative) claims about happiness. They accept the claim that one life is a happier life than another life and that the former contributes more to the goodness of the world than the latter. However, they reject the claim that the former is better for the person living that life than the latter.

Monistic views either reject the notion of personal good altogether, or they try to construct a substitute for this notion out of facts about impersonal good. For instance, it has been suggested that personal good can be understood in terms of impersonal good that is located in a particular life. On such a view, personal good is derivative and is explained in terms of the location of impersonal good, i.e. a state of affairs is good for a person x iff it is (i) impersonally good and (ii) located in x's life. In this way, personal good is nothing but an empirical relativisation of impersonal good. Such a conciliatory approach, however, runs into difficulties in trying to construct a notion of personal good out of impersonal good. For instance, there does not seem to be any systematic way of cashing out the 'occurring in a life' locution that is consonant with our intuitive judgements (cf. Rosati: 2008, pp. 332-345). Accordingly, it would be better for monistic theorists to opt for the strict Moorean line and reject the notion of personal good altogether and do without it.[1]

2.2 Utility containers

The monistic impersonal utilitarian wants there to be happiness. This, however, is not based in a concern for the persons who will experience the happiness in

[1]As Regan notes: "Moore has no theory of these concepts. He has no use for them. And in the end I agree with Moore about this too" (Regan: 2004, p. 211 footnote 21).

question. The state of affairs in which there is more happiness is not favoured for their sake. After all, the monist does not recognise any notion of personal good. Consequently, he cannot want them to be happy on the grounds that being happy is good for them. Instead, he wants them to be happy because this is better from the point of view of the world, because this outcome contains more impersonal good.[2] The impersonal utilitarian is concerned only with how much impersonal good there is in a state of affairs, not with whether and how good that state of affairs is for persons. On this approach persons are treated as mere containers of impersonal good that are dispensable and replaceable. This form of utilitarianism cares only about an abstract and impersonal form of goodness, not about persons and what is good for them, and consequently does not show adequate concern for persons.[3]

2.3 Hybrid theories

Given the problematic nature of monistic impersonal utilitarianism, one might opt for a hybrid theory that supplements impersonal with personal good. Unlike monistic theories, which construe well-being entirely in non-evaluative terms (i.e. as mere amounts of happiness), hybrid theories accept an evaluative construal of well-being. In addition to recognising impersonal good, they also recognise an independent notion of personal good that is not reduced to or analysed in terms of impersonal good. This makes them dualistic theories that countenance two independent and non-derivative types of good. Both personal good and impersonal good are accepted as basic goods – neither of them is constructed out of or reducible to the other. (We will first consider hybrid theories that identify impersonal good with moral good and take personal good to be a form of prudential good. In section 2.6 we will consider hybrid versions that treat both of them as forms of moral good.)

Hybrid theories allow for two different formulations of the total principle. In terms of impersonal good the hybrid theorist agrees with the monistic impersonal theorist.

$$\mathrm{V}_{\text{impersonal}}(\mathrm{D}) = \sum_{\mathrm{x}\in\mathrm{D}} \mathrm{V}_{\text{impersonal}}(\mathrm{x})$$

[2] If there could be disembodied states of happiness, then this would be just as good according to the impersonal utilitarian.

[3] The objection that impersonal utilitarianism does not take personal good seriously needs to be distinguished from two other objections. First, utilitarianism is not sensitive to deontic boundaries between persons and hence does not respect the separateness of persons. Since we are only focusing on axiological issues, we will set aside these troublesome deontic commitments. Second, total utilitarianism is concerned only with the total quantity of utility and is, accordingly, not sensitive to how utility is distributed. As long as the total quantity is the same, it is indifferent between the different ways of distributing this quantity. We will set aside this pattern-insensitivity since it can be addressed within an impersonal framework by means of suitable weighting functions as well as non-separable value functions.

In addition, the hybrid theorist can formulate the total principle in terms of personal good.

$$V_{\text{impersonal}}(D) = \sum_{x \in D} f(V_{\text{personal}}(x))$$

To connect the impersonal goodness of distributions to the personal good of the members of the distribution, one needs a conversion function f that connects these types of good. Since different types of good involve different units, we cannot end up with an impersonal evaluation by aggregating personal good. In order to arrive at an impersonal evaluation, one first needs to convert personal good into impersonal good. What aggregation does is to get us from restricted or local evaluations to a (relatively) unrestricted or global evaluation, i.e. it gets us from evaluations of the components taken separately to an evaluation of them taken together. However, it cannot get us from one type of good to another. That is what conversion functions do. The function f converts personal into impersonal good, i.e. $V_{\text{impersonal}}(x) = f(V_{\text{personal}}(x))$. The sum of the impersonal goodness of the well-being of the members of the distribution is thus equivalent to the sum of the converted personal goodness of their well-being. Hybrid theories, in this way, bring out especially clearly the need to be sensitive to units of goodness, as well as the need for conversion functions that specify how units of one type of good can be converted into units of a different type of good.[4]

It is the conversion function that differentiates prioritarianism from hybrid utilitarianism.[5] These theories accept the same value function at the level of impersonal good. However, they disagree about the conversion function between personal and impersonal good. As long as one focuses on the level of impersonal good, utilitarianism and prioritarianism will be indistinguishable. The difference

[4]Once two types of good are recognised as being basic, the question arises how they determine what the agent ought to do. There will then be two different types of goodness with respect to which alternatives can be evaluated, which raises the question how they can be integrated into an overall evaluation. Problems arise when there are clashes between morality and prudence, since the agent will then be subject to conflicting and irreconcilable demands. When evaluations in terms of personal good and impersonal good diverge, there will not be a univocal ordering but conflicting orderings corresponding to the different types of good. Accordingly, one will have to overcome the Sidgwickian dualism of practical reason. The fact that these different types of good are independent and basic, however, implies that there cannot be a further standpoint that subsumes them both and with respect to which conflicts can be adjudicated. Consequently, they cannot be weighed up against each other. (Here it is important to note that the conversion function is not an ethical trade-off ratio, i.e. it does not tell us how to weigh up the different types of good, but instead merely represents a descriptive functional relationship between them.) Nor can they stand in a lexical priority relation or be ordered in some other way. As a result, they give rise to incommensurable requirements that cannot be integrated into a coherent ordering (cf. Bader: 2015).

[5]Critical-level utilitarianism (cf. Blackorby, Bossert and Donaldson: 1997, Broome: 2004) and classical versions of hybrid utilitarianism are, likewise, differentiated in terms of their conversion functions. The latter but not the former function aligns the neutral level of personal good with the neutral level of impersonal good.

between them only becomes apparent when also considering personal good, in particular when examining the conversion function connecting these two types of good. Whereas the hybrid utilitarian considers this to be a linear function, the prioritarian deems the function to be strictly concave, such that personal good has diminishing marginal impersonal goodness.[6]

Both prioritarianism and hybrid versions of utilitarianism are dualistic axiologies that are committed to both personal and impersonal good. In each case, the different types of good are connected by a conversion function that specifies the ratios between the different units (without either being reducible to the other). Both theories agree that there are two distinct types of good that stand in a functional relationship, which means that both need to distinguish between how good something is for the person and how much it contributes to impersonal good. The difference between prioritarianism and utilitarianism is simply a matter of what exactly this function is. For the utilitarian they are linearly related. For the prioritarian, by contrast, they are not related by a linear but instead by a strictly concave function.

Critiques of prioritarianism that reject or problematise the distinction between how good something is for the person and how much it contributes to the (impersonal) goodness of the state of affairs (cf. Broome: 1991, McCarthy: 2006) thus turn out to be equally applicable to hybrid versions of utilitarianism.[7] Since hybrid views are dualistic theories, they also have to distinguish between how good something is for the person and how good it is in terms of impersonal good. Once these two separate types of good are recognised, one needs to distinguish evaluations corresponding to these two points of view. Hybrid utilitarianism and prioritarianism thus have analogous commitments.[8] It is only monistic versions of utilitarianism (whether impersonal or person-affecting) that do not need to draw such distinctions.

[6] The conversion function plays a crucial role even when dealing with a 1:1 ratio, since the inputs and outputs of the function involve different units. In the same way that a 1:1 conversion ratio between two currencies does not imply that one can use the one currency in the other country, so a 1:1 ratio between units of goodness does not mean that one is dealing with types of good that can be treated as being interchangeable.

[7] These objections do not apply to 'empirical' versions of prioritarianism that do not consider personal good but instead empirical amounts of happiness to have diminishing marginal impersonal value. Likewise for deontic versions that consider the conversion function between values and reasons to be strictly concave.

[8] At most, they differ in terms of their respective information requirements. Whereas utilitarianism requires a cardinal scale in fixed-population settings and a ratio scale in variable-population contexts, prioritarianism needs a ratio scale in fixed-population comparisons and a unit scale in variable-population contexts. Yet, even this difference disappears when applying the transformations under which the representation is meant to be invariant to both the individual utilities and the prioritarian weighting function (cf. Rabinowicz: 2002, p. 9 footnote 7).

2.4 Relevance v. significance

The hybrid theory has the advantage over the monistic impersonal approach that it does recognise personal good as an independent type of good. This, however, is not enough. Making room for a notion of personal good that is merely prudential in nature does not suffice for avoiding the deficiencies of the monistic impersonalist approach. Personal good, though not being rejected altogether, is not construed in the right manner by this type of hybrid theory. In particular, it is considered to be merely ethically relevant but not ethically significant. Rather than mattering in its own right, it is merely connected via the conversion function to what is taken to matter, namely impersonal good. For hybrid views, the ethical significance of happiness does not consist in the fact that it is good for persons, but in the fact that it makes the world a better place. Happiness is valued because it is good from the viewpoint of the universe, not because it is good for the person and makes his or her life better. As a result, this approach does not show adequate concern for personal good but leads to an unacceptable sub-ordination of personal good.

Intuitively, it seems that we have to take persons as well as personal good more seriously. Persons are not mere containers of impersonal good that matter only in that they are locations where impersonal good can be instantiated. Instead, they are beings that matter and that are ethically significant. We have reasons to help others because it makes their lives go well. Promoting their happiness is something that we have reason to do because their happiness is good for them, not because there being more happiness makes the world a better place. The fact that their lives go well is of direct ethical significance. (This is especially clear in the context of agent-relative theories.) This means that what is good for them is not just ethically relevant because it is connected via a conversion function to something that is ethically significant, insofar as a life that is good for the person is also a life that makes the world a better place. Rather, personal good itself matters and is morally good.[9,10]

---

[9]The situation is thus somewhat analogous to someone who recognises aesthetic value alongside monetary value (and who considers the two to be connected insofar as things that are more beautiful have greater monetary worth) but then appreciates a beautiful painting not because of its aesthetic merits but because of its monetary value.

[10]It is sometimes suggested that happiness is to be considered to be impersonally good because it is good for the person and that this renders the container objection ineffective. (This proposal cannot be adopted by critical-level theorists who allow that something that is good for the person, namely a worthwhile life that is below the critical level, is impersonally bad.) There are two ways of reading this suggestion, both of which are problematic. On the one hand, we should reject the idea that being good for someone is just a way of being good, so that whatever is 'good for' is also 'good simpliciter'. Whereas this kind of entailment holds in the case of the conciliatory monistic view that treats 'good for' as an empirical relativisation, it does not apply in the case of hybrid theories that consider personal good and impersonal good to be two distinct types of good. On the other hand, the idea of grounding impersonal goodness in personal goodness runs into difficulties. First, if one considers good-making features to be properties of the value-bearer,

Like the monistic approach, the hybrid theory cares only about impersonal good (when concerned with moral rather than prudential evaluations). Accordingly, it also wants there to be happiness for the wrong reasons. Whilst happiness is recognised to be something that is good for persons, it is not for this reason that the hybrid theorist wants there to be happiness. The utility containers objection thus applies equally to hybrid views. This can be brought out particularly clearly in variable-population cases, since personal and impersonal good can come apart in different-number cases. In such cases, personal and impersonal good diverge and the subordination of the former to the latter leads to personal good being sacrificed for the sake of impersonal good.

The addition of a happy person contributes to the impersonal goodness of the world and makes the world a better place. However, it is not in itself better for anyone. It does not by itself contribute to anyone's happiness and does not make anyone better off. It is for this reason that impersonal versions of utilitarianism lead to repugnant conclusions in variable-population cases. Total utilitarianism requires adding persons, as long as doing so has a positive net effect on impersonal good, even if existing persons are made miserable as a result. A world consisting of an extremely large number of lives that are barely worth living will be deemed to be better than one in which a large number of very happy lives are lived. All that matters is the total quantity of goodness independently of the number of containers amongst which it is distributed. As Parfit notes: "The greatest mass of milk might be found in a heap of bottles each containing only a single drop" (Parfit: 1984, p. 388).

This recommendation to expand the distribution is neither based on a concern for existing persons, nor on a concern for the non-existing persons to be added. It is neither better for those who exist independently (in fact, it can be considerably worse for them), nor better for those who will exist as a result of expanding the distribution (since existence is not comparable with non-existence for them). Instead, it is based on a concern for impersonal goodness. The hybrid utilitarian considers happiness to matter because it makes the world a better

---

and if one consider the state of affairs [x's being happy] to be good for x, then this proposal would imply that the good-maker of impersonal goodness will involve the property of being good for x, which means that [x's being happy] will not be the bearer of impersonal good but instead the state of affairs [x's being happy being good for x]. We do not end up with one state of affairs that has two types of values, but instead with different states of affairs having different values. Whereas 'being happy' will be the good-maker of personal goodness, 'being good for' will be the good-maker of impersonal goodness. Second, variable-population cases show that one situation can be impersonally better than another without being better for anyone, which implies that personal and impersonal good come apart. Accordingly, there is no general principle to the effect that something's being impersonally good is always grounded in being good for someone. For this reason, the suggestion does nothing to address the objectionable implications of impersonal theories in variable-population settings. Third, the suggested grounding connection does not address the objection that being concerned with impersonal good is to be concerned with the wrong thing, since persons are the proper objects of concern.

place, not because it is good for persons and makes their lives go better. In this manner, a concern for impersonal goodness becomes detached from a concern for persons and what is good for them. As a result, what should be done out of a concern for what is good for persons can radically diverge from what is required by a concern for impersonal good.[11]

One of the features that renders total utilitarianism susceptible to the repugnant conclusion is that it is not sensitive to numbers. All that matters is the aggregate quantity of utility, independently of the number of locations across which it is distributed. Alternative theories, such as critical-level utilitarianism, have been proposed to avoid this problem. These theories are impersonal yet are number-sensitive. Although it is possible to avoid the repugnant conclusion by suitably modifying the value function, this does not address the underlying problem with impersonal approaches. Though being number-sensitive, these theories do not care about persons and what is good for them. Instead of being concerned with personal good, they merely build number-sensitivity into the impersonal evaluation. As a result, they are still operating with utility containers. The milk bottle analogy still holds. All that they care about is the total quantity of milk – it simply happens that they try to reduce the number of containers amongst which a given total quantity of milk is distributed as much as possible, since the bottles leak a bit (where the amount of leakage per bottle corresponds to the critical-level parameter).[12]

The fact that a conversion function only ensures ethical relevance is particularly apparent in cases in which the connection between personal and impersonal good is not invariant. This happens for instance when dealing with average value functions. In such cases, personal good and impersonal good can come apart and it becomes clear which of the two types of good is of significance and which one does not matter in its own right but is relevant only as a result of being connected to something that has significance. The ethical significance of an additional life does not correspond to its level of personal good, but is instead determined by the impact that it has on the average. A worthwhile life can make things (impersonally) worse if it brings down the average and a miserable life can make things (impersonally) better by improving the average. How good well-being is for the individual then comes apart from how much it matters morally. In these cases

[11] The case of average utilitarianism provides a helpful illustration. Whether a person is to be added depends not on how well that person's life goes, but on how the average is affected. As a result, cases can arise where a concern for goodness requires adding miserable lives that have a positive effect on the average. Such a course of action clearly cannot be recommended out of a concern for the existing persons that are unaffected, nor out of a concern for the persons being added whose lives are not worth living, but only out of a concern for an impersonal ideal that is detached from what really matters.

[12] Since the critical level penalty applies at the level of impersonal good it might be more accurate to think not in terms of leakage but in terms of being paid for the total quantity of milk where one has to pay a fixed amount for each bottle.

the personal goodness of happiness is intrinsic and invariant. Its contribution to impersonal good, however, is extrinsic and derivative, since it is a function of the relation between the well-being of the individual and that of all the others. What is significant and what determines what is to be done according to such a value function is not whether the life is worth living but how the average is affected. It is the impersonal goodness of the distribution (which is determined by the average principle) that matters and that determines what is to be done, while personal good is merely relevant insofar as it is functionally related to impersonal good.

2.5 Welfarism

The distinction between relevance and significance allows us to distinguish two versions of welfarism. Although all versions of utilitarianism are welfarist theories, in that they are concerned with well-being, there are differences in terms of the significance that they assign to well-being.

Welfarism, understood along the lines proposed by Sen: 1979, amounts to an 'informational constraint'. It requires that information about well-being suffices for evaluating and ordering states of affairs. (As such, it conflicts with impersonal theories that countenance non-person-involving good-makers, such as biodiversity.) Satisfying this informational constraint by insisting that the betterness ordering is purely based on welfarist facts does not adequately address the concerns that have been raised. That well-being plays a role in the theory does not ensure that it enters in the right way. One can be concerned about well-being without caring about persons and what is good for them. In particular, theories that 'deplore the situation where some happiness lacks a person' take well-being into consideration but radically misconstrue its significance.

This should be readily apparent, given that the informational constraint can be satisfied by an impersonal monist who rejects the notion of personal good altogether but considers impersonal good to be a function of empirical (non-evaluative) facts about well-being. Even if welfare is understood evaluatively, such that it corresponds to personal good, the welfarist restriction is too weak. This is because it is compatible with a merely functional relation between personal good and what ultimately matters according to hybrid theories, namely impersonal good. Accordingly, it does not ensure that personal good is ethically significant and matters in its own right.

In order to ensure that personal good is not only ethically relevant but ethically significant, a more robust construal, such as the one put forward by Moore and Crisp: 1996, is needed. It has to go beyond the information-theoretic characterisation and has to build in ethical significance. According to this more demanding characterisation of welfarism, impersonal versions of utilitarianism do not classify as welfarist. Theories that are welfarist in this more robust sense consider personal good to be a form of moral good. Personal good then matters in its own right, i.e. ethical significance attaches to it and not only to something to which it is functionally connected.

2.6 Non-subordination hybrids

So far we have considered a hybrid theory that accepts both personal and impersonal good as basic notions that are not reducible to each other, but then subordinates personal to impersonal good by construing the former as being merely prudential. This ensures that evaluations of moral goodness will be restricted to impersonal good and that personal good is only at issue in prudential evaluations.

Once it is recognised that personal good has to be ethically significant, the possibility of a non-subordination hybrid theory becomes salient. Such a theory assigns ethical significance to both impersonal and personal good.[13] Although this approach does not suffer from the problem of subordinating personal to impersonal good, it is nevertheless untenable. As a result, one should reject rather than supplement impersonal good.

To begin with, this approach does not generate a version of total utilitarianism. Rather than summing the impersonal value of well-being, both impersonal and personal value will contribute to the overall value of a state of affairs.

$$V(D) = g(V_{\text{impersonal}}(D), V_{\text{personal}}(D))$$

The fact that two kinds of good are countenanced in this way is likely to lead to over-counting. If x's happiness is both good for x and good impersonally, and if both of these types of good are considered to have ethical significance and to give rise to (or at least be connected to) reasons for action, then there will be vast over-counting both in terms of axiological evaluations and in terms of assessments of reasons. While concerns about over-counting can generally be addressed by appealing to the notion of basic value, thereby ensuring that non-basic value is not counted in addition to the basic value from which it derives, this solution is not applicable in the case of hybrid models since both types of value will be basic.

The resulting over-counting gives rise to biases in cases in which personal and impersonal value come apart, i.e. when something has impersonal value but no personal value. For instance, adding happy lives counts as an improvement from the point of view of impersonal good, but not from the point of view of personal good (given that one rejects comparativism, cf. section 3.3). Accordingly, over-counting will privilege improving the life of an existing person by a certain amount over bringing into existence a new person with a level of well-being corresponding to that amount. If $D_0 = (2, \Omega)$ can be turned into $D_1 = (4, \Omega)$ or $D_2 = (2, 2)$, then the former turns out to be preferable since there is a +2 gain in terms of impersonal good as well as a +2 gain in terms of personal good, whereas

[13] Whilst personal good is frequently identified with non-moral or merely prudential good, it is worth noting that this identification is theory-driven.

in the latter case there is only the gain in terms of impersonal good. This ensures that these hybrid axiologies will be strongly biased against non-existent persons.[14]

Additionally, this approach runs into difficulties since one is dealing with two distinct types of values that will be incommensurable. Since there is no common unit of measurement, these values cannot be combined into an overall evaluation. One can only appeal to dominance principles applied over multi-dimensional value vectors that specify that a distribution D is better than another, D′, if D is at least as good as D′ with respect to each type of value and strictly better with respect to at least one type. Yet, there will not be any way to trade off the two values against each other and make sense of what is best on balance.

Moreover, since distributions of different sizes are not comparable in terms of personal good (due to the fact that existence is non-comparable with non-existence), dominance principles are not applicable in different-number cases. This is because dominance principles require weak betterness with respect to each dimension of the multi-dimensional value vector, which is not satisfied if the alternatives turn out to be non-comparable along one of the dimensions. Accordingly, the non-comparability with respect to personal good will render the overall evaluation incomplete. The resulting quasi-ordering induced by these dominance principles thus does not yield any determinate verdict when concerned with different-number comparisons.

The inadequacy of hybrid models shows that one cannot address the utility containers objection by supplementing an impersonal theory. Making room for personal good (when this is construed as being ethically significant) is not simply a matter of making a local adjustment to one's axiology. Instead, it requires a radical readjustment that involves switching from an impersonal theory to a person-affecting view that is entirely based on personal good.

3 Person-affecting utilitarianism

Person-affecting approaches reject impersonal good and instead only recognise personal good. The betterness ordering of distributions accordingly has to consist in betterness facts for the members of the distributions, which means that the evaluation of distributions has to be reducible to facts about personal good.[15] The total principle in this context requires summing up the personal goodness of

[14] If there were uniform over-counting that applied to all value-bearers, then the value function could be normalised so that no practical problems would result. Instead, only theoretical and intuitive concerns would be applicable.

[15] This reducibility constraint is significantly stronger than the 'person-affecting restriction', which only imposes a necessary condition on the betterness ordering of distributions, namely that in order for one distribution to be better than another it must be better for someone, and which is not sufficient for ruling out impersonal hybrid theories.

the well-being of all the individuals.

$$V_{general}(D) = \sum_{x \in D} V_{personal}(x)$$

This approach can assign ethical significance to personal good and does not succumb to the objection of impersonalism.[16]

In order to end up with a person-affecting total view, one needs: 1. aggregation in order to underwrite same-persons comparisons, 2. impartiality to get same-number comparisons, and 3. comparativism to make sense of different-number comparisons.[17] Whilst 1. and 2. are fine, 3. is objectionable. As a result, person-affecting total utilitarianism is to be rejected. The only defensible version of person-affecting utilitarianism is same-number utilitarianism.[18]

3.1 Aggregation

Person-affecting utilitarianism requires one to aggregate the personal good of different persons. However, it is frequently claimed that it is not possible to make sense of trade-offs of personal good across different persons and that interpersonal aggregation presupposes a commitment to an impersonal point of view.

> "when we judge that one person's claim outweighs the claim of someone else we are assuming an impersonal point of view. We are not just looking at things from one person's point of view and registering a loss, and then looking at things from another person's point

[16] On this approach, prudential and moral evaluations do not differ in kind but in scope. Prudential evaluations are restricted to the agent in question, whereas moral evaluations are unrestricted.

[17] More precisely, same-persons comparisons already involve a weak form of impartiality that requires the betterness relation to be invariant under permutations of the actual members of the domains (i.e. it does not matter who in the distribution is assigned which value). Same-number comparisons, by contrast, require a strong form of impartiality, according to which the betterness relation has to be invariant under permutations of possible members of the domains (i.e. it does not even matter who is in the distribution).

WEAK IMPARTIALITY: $D_1 \geq D_2$ iff any permutation of $D_1 \geq$ any permutation of D_2.

STRONG IMPARTIALITY: $D_1 \geq D_2$ iff the restriction to D_1 of any permutation of the universal domain $\geq$ the restriction to D_2 of any permutation of the universal domain.

[18] Parfit has put forward a wide total person-affecting principle as well as wide average person-affecting principle (cf. Parfit: 1984, pp. 396-401). Moving away from a narrow to a wide principle that allows for impartiality is a relatively straightforward step that is crucial for developing a satisfactory person-affecting approach. However, the question of aggregation is an altogether different matter that does not readily admit of an extension to different-number cases. As we will see, one can only make sense of a total principle if one accepts comparativism. By contrast, an average principle is incompatible with a person-affecting approach since it violates the reducibility constraint.

of view and registering a gain. We make a comparison that includes both people and their points of view. We judge that it is more important to help the first person, and this judgement is not made from any individual's point of view" (McKerlie: 1988 p. 222).[19]

The challenge is thus to make sense of aggregating and trading off personal good. To begin with, it should be noted that the notion of overall or aggregate goodness is not to be confused with impersonal goodness. This should be clear, given that one can aggregate the goodness to be found in different temporal locations from the point of view of personal good to arrive at an overall evaluation that is not restricted to particular times but considers the life as a whole. Likewise, one can aggregate probabilistically discounted personal good to be found in different states of natures to arrive at the expected personal goodness of a lottery. There is hence no problem in general when it comes to aggregating personal good.

What is difficult, though, is making sense of aggregation across different persons. Given that the gains and losses accrue to different persons, there is no person from whose point of view they can be evaluated. The balance of gains and losses is a relation amongst the different personal points of view, but does not occur within any such point of view. As such, it seems to transcend the perspective of individuals, which would suggest that a person-affecting view can only make sense of what is better or worse for particular individuals and that trade-offs across persons would consequently be ruled out by a person-affecting approach.

There is, however, no need to invoke an impersonal point of view. One can instead bring in the notion of a plural point of view and argue that to speak of the value of a distribution or world is to speak plurally of the values of the members of the distribution. Facts about general good, on this view, just are plural facts about personal good, and the general betterness relation ends up being a plural comparative. General good is thus not a distinct kind of good. In this way, one stays at the level of personal good when aggregating across persons. One does not bring in impersonal good but simply moves from a singular to a plural evaluation, allowing one to balance gains and losses by comparing them from the point of view of a plurality of individuals.[20]

3.2 Impartiality

Same-number comparisons that involve different persons require impartiality. This is unproblematic since a person-affecting view can be impartial. Impartiality is not to be confused with impersonality. Instead, it is to be identified with permutation-invariance. This means that permuting the identities of the members of the distribution, whilst holding fixed the structure or value profile of the distribution, does not affect the betterness ordering. Though impersonal

[19] Also cf. Regan: 2004, p. 213 footnote 25 and Arneson: 2010, p. 735 footnote 2

[20] Cf. "Aggregating, balancing, and the separateness of persons" (Bader: manuscript).

goodness is impartial, it is not the case that impartiality requires impersonality.[21] While an impersonal assessment is one that tries to transcend the perspectival nature of evaluation, one that tries to assess a situation independently of how it affects anyone and thus independently of whether it is good or bad for anyone, no such perspective-transcendence is required to achieve an impartial assessment. An impartial perspective need not be impersonal and the impartial point of view is not to be equated with the impersonal point of view.

Instead, it can be construed as the permutation-invariant point of view. One does not have to adopt a view from nowhere in order to achieve impartiality. Rather than adopting a view from nowhere, one needs to adopt the viewpoint of no one in particular. One needs to detach from one's own perspective by setting aside and abstracting from the particularities of the perspective that one occupies. An assessment that detaches from what is peculiar to particular perspectives will be invariant across perspectives, thereby making the resulting assessment permutation invariant. What holds in such an assessment is not something that transcends and holds independently of perspectival evaluations, but rather something that holds no matter which perspective is adopted.[22]

Since the impartial point of view is not the perspective-transcending impersonal point of view but the perspective-invariant point of view, a rejection of an impersonal betterness relation does not imply a rejection of an impartial betterness relation. An impartial betterness relation does not require a commitment to impersonality. This means that personal goodness can give rise to an impartial ordering that is permutation-invariant.[23]

3.3 Comparativism

Impersonal versions can straightforwardly make different-number comparisons. It can be impersonally better that a happy person exists than that the person does not exist, on the basis that this person's existence makes the world a better place.

[21] The fact that impersonal implies impartial has the consequence that one can only make sense of agent-relativity (= partiality) within the context of personal good (cf. "Personal, general, and impersonal good" Bader: manuscript).

[22] The distinction between understanding impartiality in terms of perspective-invariance and in terms of perspective-independence is analogous to that between understanding objectivity in terms of invariance across observers and in terms of observer-independence (cf. Eddington: 1920, p. 31 & p. 82). The distinction between the invariant and the independent also plays an important role in Fine's distinction between a proposition's being true whatever the circumstances and being true regardless of the circumstances (cf. Fine: 2005, ch. 9).

[23] Heyd holds that a commitment to impersonalism follows from "allowing comparisons between the welfare of two different potential beings, which is logically exactly the same as comparing the existence of a potential being with its non-existence" (Heyd: 1988, p. 161; also cf. Heyd: 1992, p. 105). This claim confuses impersonalism and impartiality. All that is required for ordering different lives is that the betterness relation be impartial. Since impartiality can be separated from impersonalism, no commitment to the latter view follows.

The situation in which the person does exist is comparable from the point of view of impersonal good with the situation in which the person does not exist.

Difficulties arise, however, when operating with a person-affecting theory. It is far from clear that the two situations are comparable from the point of view of personal good. Comparability only makes sense if one accepts the claim that it is better for a person to live a happy life than to not exist. Unless one accepts comparativism, a person-affecting approach will end up with a failure of completeness and will not be able to make different-number comparisons. In particular, if comparativism is rejected, one cannot compare different-number cases in terms of personal good, i.e. $V(D_1)$ will not be comparable with $V(D_2)$ if $|D_1| \neq |D_2|$. This means that a total principle based on personal good presupposes the contentious assumption that comparativism holds for the personal betterness relation. Existence and non-existence have to be comparable from the point of view of personal good. Living a life has to be better/neutral/worse for the person in question than not having lived at all.[24,25]

Comparativism not only allows one to make different-number comparisons within a person-affecting theory. It also mitigates the objection that the total principle implies the repugnant conclusion. This is because comparativism considers coming into existence to be a benefit for a person with a life that is worth living. Since the benefits accruing to those that are brought into existence outweigh the costs of those already in existence, there is no net sacrifice of personal good in moving from a quality to a quantity distribution. Given that the quantity distribution involves a very large number of persons that are slightly benefited, this does not involve any repugnant sub-ordination or sacrifice of personal good.

Comparativism, however, faces serious metaphysical problems. First, it needs to be explained how the betterness relation can hold if one of the relata is missing. The personal betterness relation is a dyadic relation that has lives as its relata. This means that we need two lives that are being compared in order for a personal bet-

[24] In particular, what is required is an unrestricted form of comparativism. For instance, it is not sufficient to adopt the version defended by Arrhenius and Rabinowicz: 2015, which involves a failure of the accessibility principle that if D_1 is better for x than D_2, then D_1 would be better for x than D_2 if D_1 obtained.

[25] It might be objected that comparativism is unnecessary since it is enough that existence is good for the person. On this view one should bring a happy life into existence, not because living the happy life is better for x than non-existence, but because living the happy life is good for x and because the person is (non-comparatively) benefitted by being brought into existence (cf. Bykvist: 2015). However, although a happy life has value and is good for the person living this life, this goodness is of no relevance relative to things that lack goodness, such as non-existence. Goodness operates by making things better, by providing more reasons than the alternative. In order for the goodness of option ϕ to favour ϕ over alternative ψ, the goodness of ψ must also be defined. There is only a stronger reason if there is more goodness and this requires comparability of the options. Accordingly, in order to favour choosing one rather than the other, it must be the case that the one is better than the other. Being good is not enough. What is needed is betterness. Goodness and betterness are thus inseparable when it comes to favouring actions. In short, goodness only operates within the field of betterness.

terness relation to hold. Yet, there is no life in the case of non-existence. Second, the comparativist needs to provide a supervenience base upon which betterness facts can supervene. This is rather problematic given that a non-existent life does not instantiate any non-evaluative properties that could determine what evaluative properties this life has and how it is to be ordered with respect to various other lives. Comparativism is thus not viable since there cannot be a betterness relation without relata, nor can there be goodness without good-making features.[26]

If comparativism is rejected, a person-affecting approach precludes ordering distributions by means of the total principle. Instead, it only allows us to generate a quasi-ordering (to which one can apply an optimising but not a maximising function) that yields the same results as total utilitarianism in same-number cases but that cannot be extended to different-number cases.[27]

$$D_1 \geq D_2 \text{ iff } |D_1| = |D_2| \wedge \sum_{x \in D_1} V_p(x) \geq \sum_{x \in D_2} V_p(x)$$

3.4 Empty distributions

Total utilitarianism is acceptable within a person-affecting approach if one accepts comparativism. In fact, the total principle that a comparativist can underwrite might be stronger than the one that an impersonal theorist can defend. The (potential) difference between the two total principles comes out when considering empty distributions.

For the comparativist, the empty distribution in which no one exists is as good as a distribution in which there are persons all of whom are living neutral lives. The empty distribution can be assigned a value and is straightforwardly comparable to non-empty distributions. By contrast, the empty distribution might not have value at all for the impersonalist and hence might not be comparable with non-empty distributions.

Distributions in the impersonal approach are the analogues of lives in the person-affecting approach, which makes the empty distribution the analogue of non-existence. In the same way that it is a substantive question whether existence is comparable with non-existence with respect to personal good, so it is a substantive question whether non-empty distributions are comparable with the empty distribution with respect to impersonal good. Adding a time-slice makes the life better without being better from the perspective of that time-slice. Likewise adding a person makes the distribution better without being better for that person. An empty life (i.e. one without any time-slices) is not a life and hence, according to the non-comparativist, is not comparable with non-empty lives. One

[26] Cf. "The neutrality of existence" (Bader: manuscript).

[27] This kind of theory might be thought to be unsatisfactory owing to its commitment to non-comparability. Cf. "Person-affecting population ethics" (Bader: manuscript) for arguments to the effect that a satisfactory population ethics can (and indeed must) recognise the non-comparability of different-number cases.

only has a life in one of the cases but not in the other and hence does not have the two relata that are connected by the personal betterness relation. Analogously, an empty distribution (i.e. one without any members) is not a distribution and hence, according to the non-comparativist, is not comparable with non-empty distributions. One needs two (non-empty) distributions as relata of the impersonal betterness relation. This means that there is a substantive question as to whether we have comparability in terms of impersonal good. As a result, the impersonal approach might differ extensionally from the person-affecting total view when it comes to the empty distribution.

4 Conclusion

Total utilitarianism has deeply problematic axiological commitments. By adopting an impersonalist approach, one ends up with an objectionable theory that either eschews any independent notion of personal good altogether, or that recognises personal good but is willing to sacrifice and sub-ordinate personal good for the sake of impersonal good. Such theories can at best assign ethical relevance but not ethical significance to personal good and thus operate with the spurious notion of 'utility' that Bennett warned us about. They do not show adequate concern for personal good but consider persons as mere containers of impersonal good. This type of theory is committed to various repugnant conclusions and does not take seriously the idea that what matters is how well peoples' lives go.

Alternatively, a person-affecting version of total utilitarianism that is not subject to the charge of subordinating and sacrificing personal good requires a commitment to comparativism. Such a theory operates with the proper notion of 'utility', but is nevertheless to be rejected, since the claim that existence is comparable with non-existence from the point of view of personal good runs into insurmountable metaphysical difficulties. Moreover, this type of theory can underwrite neither the neutrality intuition nor the asymmetry.

In order to avoid this predicament, one has to reject comparativism along with impersonalism and accept same-number person-affecting utilitarianism (considered as an axiological theory). This is an impartial theory that completely orders populations of the same size in terms of their total utility, but that renders populations of different sizes non-comparable. Such a theory neither has the problematic metaphysical commitments nor the objectionable ethical implications from which the various versions of total utilitarianism suffer.[28]

[28] For helpful comments, I would like to thank Roger Crisp, Teru Thomas and especially Tim Campbell. I am also grateful to the participants of the Oxford Population Ethics work in progress seminar as well as audiences at Umeå Universitet and the International Society for Utilitarian Studies conference at Yokohama.

References

[1] Arneson, R. Good, period. *Analysis 70*, 4 (2010), 731–744.

[2] Arrhenius, G., and Rabinowicz, W. The value of existence. In *The Oxford Handbook of Value Theory*, I. Hirose and J. Olson, Eds. Oxford University Press, 2015, pp. 424–443.

[3] Bader, R. M. Kantian Axiology and the Dualism of Practical Reason. In *The Oxford Handbook of Value Theory*, I. Hirose and J. Olson, Eds. Oxford University Press, 2015.

[4] Bennett, J. On Maximizing Happiness. In *Obligations to Future Generations*, R. I. Sikora and B. Barry, Eds. Temple University Press, 1978, pp. 61–73.

[5] Blackorby, C., Bossert, W., and Donaldson, D. Critical-level utilitarianism and the population ethics dilemma. *Economics and Philosophy 13* (1997), 197–230.

[6] Broome, J. *Weighing Goods*. Blackwell Publishers, 1991.

[7] Broome, J. *Weighing Lives*. Oxford University Press, 2004.

[8] Bykvist, K. Being and Wellbeing. In *Weighing & Reasoning: Themes from the Philosophy of John Broome*. Oxford University Press, 2015, pp. 87–94.

[9] Eddington, A. *Space, Time and Gravitation – An outline of the general relativity theory*. Cambridge University Press, 1920.

[10] Fine, K. *Modality and Tense: Philosophical Papers*. Oxford University Press, 2005.

[11] Heyd, D. Procreation and Value – Can ethics deal with futurity problems? *Philosophia 18* (1988), 151–170.

[12] Heyd, D. *Genethics – Moral Issues in the Creation of People*. University of California Press, 1992.

[13] McCarthy, D. Utilitarianism and prioritarianism I. *Economics and Philosophy 22* (2006), 1–29.

[14] McKerlie, D. Egalitarianism and the Separateness of Persons. *Canadian Journal of Philosophy 18*, 2 (1988), 205–226.

[15] Moore, A., and Crisp, R. Welfarism and moral theory. *Australasian Journal of Philosophy 74*, 4 (1996), 598–613.

[16] Parfit, D. *Reasons and Persons*. Oxford University Press, 1984.

[17] Rabinowicz, W. Prioritarianism for prospects. *Utilitas 14*, 1 (2002), 2–21.

[18] Regan, D. Why am I my brother's keeper? In *Reason and Value – Themes from the moral philosophy of Joseph Raz*, R. J. Wallace, P. Pettit, S. Scheffler, and M. Smith, Eds. Oxford University Press, 2004, pp. 202–230.

[19] Rosati, C. Objectivism and relational good. *Social Philosophy & Policy 25* (2008), 314–349.

[20] Sen, A. Utilitarianism and welfarism. *The Journal of Philosophy 76*, 9 (1979), 463–489.