



The Oxford Handbook of Well-Being and Public Policy

Matthew D. Adler (ed.), Marc Fleurbaey (ed.)

<https://doi.org/10.1093/oxfordhb/9780199325818.001.0001>

Published: 2016

Online ISBN: 9780199349715

Print ISBN: 9780199325818

CHAPTER

30 The Well-Being of Future Generations

John Broome

<https://doi.org/10.1093/oxfordhb/9780199325818.013.29> Pages 901–928

Published: 02 November 2016

Keywords: [discounting](#), [prioritarianism](#), [utilitarianism](#), [population ethics](#), [value of population](#), [intuition of neutrality](#), [neutral level](#), [nonidentity](#), [Pareto principle](#)

Subject: [Health, Education, and Welfare](#), [Economics](#)

Series: [Oxford Handbooks](#)

Collection: [Oxford Handbooks Online](#)

30.1 Introduction

GOVERNMENTS should promote the well-being of existing people and of people who do not yet exist. The well-being of those who do not yet exist raises a number of difficult questions that do not arise with existing people. This chapter surveys some of them.

I treat this as an exercise in valuation: in judging the relative value, or goodness, of different ways the world might go. I shall not consider how people or states or governments should act in response to valuations. The goodness of consequences is an important consideration in determining how agents should act, but it may not be the only one. For example, it may be wrong to impose draconian population policies even if they would be beneficial, because of the damage they do to people's freedoms. I set those matters aside.

I shall focus on people and how well their lives go—on what I shall call their lifetime well-beings. I take account of the fact that each person is born at some time, lives for a while, and then dies. An alternative approach is to focus on the quality of life at particular times: how good life is at each time—on people's temporal well-beings, as I call them, at each time. The latter approach to valuation is common in economics, and is briefly considered in section 30.4. But there I explain that it is workable only under very implausible assumptions, and I do not use it elsewhere.

In evaluating alternative ways the world might go, we often need to recognize that different alternatives may lead to the existence of different populations of people. Actions we take can easily affect the world's future population. They can affect the number of people who live, and even when they do not affect the number, they can affect the identities of the people who live—which particular people live. Nevertheless, the valuations that inform practical decision-making almost universally ignore changes in population.

p. 902 section 30.5 explains that this is a mistake: these changes can make an  important difference to value, and

should definitely be taken into account. section 30.6 goes on to investigate how the numbers of people should be taken into account. sections 30.7 and 30.8 investigate issues that arise when our actions affect the identities of future people.

Before coming to matters of population, sections 30.2, 30.3, and 30.4 examine the discounting of future well-being. Those sections abstract away from changes in population.

This whole chapter (but see Mongin and Pivato, chapter 24, this Handbook) also abstracts away from uncertainty. I assume that the way to treat uncertainty in decision-making is, in principle, to start by evaluating each of the possible results of each alternative decision. This chapter is concerned with how these separate evaluations should be made. The separate evaluations must then be put together to determine an overall valuation of each alternative in a way that takes proper account of the uncertainty. It is often possible to shortcut this process in practice, and in some special cases uncertainty about the future can be treated as a temporal discount factor. But I do not consider discounting for uncertainty in this chapter.

30.2 Discounting Lives

Should the well-being of future people be *discounted* relative to the well-being of present people? In general should the well-being of later people be discounted relative to the well-being of earlier people? section 30.3 considers this question. But to express it clearly I need to do some preliminary work and make a number of special assumptions. That is the business of this section. By the end of it, the question itself will have been amended.

I assume the question arises in a context where we are evaluating alternative ways the world might go, and aiming to make a judgment of their overall, or general, goodness. We need to determine which alternatives are better than which. However, we do not need to determine the absolute goodness of the alternatives; only their ordering by betterness.

I assume this ordering depends on how well off individual people are and perhaps on other factors. If there are I people, general goodness can be ordinally represented by a *value function*:

$$U = U(w_1, w_2, \dots, w_I, z).$$

Until section 30.5, I assume that the number and identities of the people who exist are constant; I assume we are not trying to evaluate policies that affect which people exist. That is a matter for later sections. However, I do not assume that all the people live at the same time; they may be born and die at different times in history. w_i stands for how well the i th person's life goes as a whole; it is i 's *lifetime well-being*. It is to be distinguished ↵ from the well-being that i enjoys at each particular time in her life, which I call her *temporal well-being*. In this chapter, "well-being" on its own always refers to lifetime well-being. I take well-being to be quantitative to the extent that it is measured on a cardinal scale, and I take this scale to be comparable between people.

The argument z in $U()$ stands for other factors that contribute to determining general goodness, besides people's well-beings. It is doubtless a vector, containing several components. It no doubt includes the well-being of animals and perhaps other things. I shall assume that people's well-beings, taken together, can be evaluated independently of the factors in z . That is to say, the function $U()$ takes this form:

$$U = U(w_1, w_2, \dots, w_I, z) = U(V(w_1, w_2, \dots, w_I), z).$$

Technically, this is to assume that well-beings are *separable* from other factors. Given that, we can make sense of a value function of human well-being alone:

$$V = V(w_1, w_2, \dots, w_I). \quad (30.1)$$

This is the function I shall work with; this chapter deals with the value of human well-being only. I call V simply “value” or “general value.”

Equation (30.1) incorporates an assumption of another sort of separability: separability (technically, weak separability) between the lives of different people. This means that general value depends only on how each person’s life goes taken as a whole, and independently of how other people’s lives go. It is a substantive assumption that may be questioned (Broome 2004, ch. 8). For example, it rules out a particular view about the badness of inequality. Imagine that everyone’s life is equally good and has the same pattern: each life is hard at the beginning, but ends up with a golden old age. At each time there are some badly off young people and some well-off old people, so there is inequality at every time. But this inequality does not appear in formula (30.1) because there is no inequality between lifetime well-beings and V is a function of lifetime well-beings only. Still, you might think there is something bad about the inequality that exists at every time (McKerlie 1989). If you do, you must reject (30.1) as a value function and deny its implicit assumption of separability between lives, or at least you must do some analytic maneuvering to accommodate your view to the formula (Broome 2004, 123–6).

Nevertheless, separability between lives is defensible (Broome 2004, sec. 8.2), and I assume it throughout this chapter. It focuses attention on the people who make up future generations, and the quality of their lives as a whole, rather than on the people’s temporal well-beings at particular dates in the future. section 30.4, however, does consider the discounting of temporal well-beings.

The separability of lives raises the question of what determines each person’s lifetime well-being. No doubt this depends—though possibly not exclusively—on the person’s temporal well-being at each time in her life. One issue is how the temporal well-being a person enjoys at one time weighs against her temporal well-being at another time in determining her lifetime well-being. It is sometimes claimed that the end of life is ^{p. 904} more important than the beginning, so that temporal well-being at the end counts for more than temporal well-being at the beginning (Velleman 1991). The opposite is often assumed implicitly in economics: that well-being at the beginning of life weighs more than well-being at the end—later well-being is discounted within a life, that is to say (Dasgupta, forthcoming). This chapter leaves aside the question of what determines lifetime well-being (but see Ponthiere, chapter 29, this Handbook). In particular, except implicitly in section 30.4, it does not consider discounting within a life; it is concerned with discounting between lives rather than within lives.

In (30.1) take the first derivatives of value with respect to each person’s well-being: $\partial V/\partial w_1$, $\partial V/\partial w_2$ and so on. These derivatives measure the marginal contributions that each person’s well-being makes to value. If the marginal contribution of the well-being of people who live later is less than the marginal contribution of the well-being of those who live earlier, that means later people’s well-being is discounted. More precisely, it is discounted at the margin. Discounting at the margin like this could arise from contingent features of the distribution of well-beings in a way that is described below. It is first more important to know whether discounting is embedded in the structure of the value function itself. Let us turn to that question.

If discounting is embedded in the value function, then the function is *partial* rather than *impartial* between people. Mathematically, partiality is a matter of whether or not different people’s well-beings appear symmetrically in the function $V()$. If well-beings were permuted among the people, would that make any

difference to value as given by $V()$? If not, the function is symmetric, which is to say impartial; each person's well-being contributes to general value in the same way. On the other hand, if permuting well-beings can make a difference to value, the function is asymmetric and partial.

Discounting is a particular sort of partiality. But in order to describe it properly, I need some more assumptions. One is that the value function $V()$ takes an *additively separable* form. That is:

$$V = v_1(w_1) + v_2(w_2) + \dots + v_I(w_I). \quad (30.2)$$

$v_i(w_i)$ can be thought of as the value of the i th person's well-being; it is the contribution her well-being makes to general value. I call the functions $v_i()$ *personal value functions*.

We have already assumed separability between lives. It is debatable whether this assumption should be strengthened to additive separability. For one thing, additive separability rules out a particular view, known as strict egalitarianism, about the value of equality (Temkin 1993). Nevertheless, a good case can be made for additive separability.¹ I shall not rehearse the arguments, but simply assume for the sake of moving forward that (30.2) is the correct formula.

p. 905 I also make the further assumption that $v_i()$ takes the specific form $d_i v()$, so that (30.2) becomes

$$V = d_1 v(w_1) + d_2 v(w_2) + \dots + d_I v(w_I). \quad (30.3)$$

The assumption is that the value of the i th person's well-being receives a weight d_i , but is otherwise given by the same function $v()$ as everyone else's. d_i may be called the *discount factor* of the i th person. Each person's well-being is valued in the same way apart from this factor. It is hard to see why this assumption might be doubted.

If the d_i s are not all the same, (30.3) is partial rather than impartial between people. If the discount factors of people who live later are less than those of people who live earlier, (30.3) discounts the well-being of later people. This case when later people have smaller discount factors than earlier ones is called *pure discounting*.

As a matter of terminology, I call $v(w_i)$ the "undiscounted value" of i 's well-being, and $d_i v(w_i)$ its "discounted value." So discount factors are applied to the undiscounted values of well-being. Pure discounting is the discounting of these undiscounted values. It is not the discounting of well-being exactly, but of the value of well-being. This is the sort of discounting that may be embedded in the value function.

Even if there is no pure discounting, well-being may still be discounted at the margin. When there is no pure discounting, all the discount factors are the same. This means they may all be treated as 1. Equation (30.3) becomes

$$V = v(w_1) + v(w_2) + \dots + v(w_I). \quad (30.4)$$

This is an impartial value function, since permuting the well-beings w_i between people does not affect V according to this function. Nevertheless, the marginal contributions of different people's well-beings will

not normally be equal.

For example, suppose the personal value function $v()$ is strictly concave. Then (30.4) represents a view known as *prioritarianism*.² The prioritarian slogan is “priority to the worse off.” Prioritarians think it more valuable to add well-being to worse-off than to better-off people. So they think the value of well-being diverges from well-being itself. Indeed, they think well-being has “diminishing marginal value”: the more well-being a person has, the less the value of her acquiring still more. Now suppose that later people are better off than earlier people—perhaps there is beneficial economic growth. Then according to prioritarianism, an increase in well-being that comes to a later, better-off person is less valuable than an increase of the same amount that comes to an earlier, less well-off person. This means well-being is discounted at the margin. It is a consequence of the diminishing marginal value of well-being.

p. 906 But suppose we now make the further assumption that $v()$ is a linear function. In that case we may treat it as the identity function. We need make no distinction between \hookrightarrow well-being and the value of well-being. Equation (30.3) becomes

$$V = d_1 w_1 + d_2 w_2 + \cdots + d_I w_I. \quad (30.5)$$

This formula has discount factors. The linear case without discounting is *utilitarianism*. It is the theory that general goodness can be represented by a value function that is simply the total of well-being:

$$V = w_1 + w_2 + \cdots + w_I. \quad (30.6)$$

We might call the theory represented by (30.5) *discounted utilitarianism*. This term has an oxymoronic flavor, however, since a commitment to impartiality between people is generally considered an essential part of utilitarianism.

Pure discounting is not generally equivalent to discounting well-being, but in the one special case of discounted utilitarianism, it is equivalent. Economists do not generally distinguish the value of well-being from well-being itself, so they are generally able to ignore the distinction between pure discounting and discounting of well-being. Nevertheless, in a discussion of discounting it is important to be clear what sort of discounting is in question and in particular what it is discounting of. “Discounting” is a general term for attaching less value to something in some circumstances than in others. We are concerned particularly with temporal discounting, which means giving less value to something at a later time than to the same thing at an earlier time. The “something” may be well-being, the value of well-being, or something else. Pure discounting is discounting the value of well-being, and only in one special linear case is it discounting well-being itself.

In many practical contexts, the “something” is a quantity of a particular commodity. We may discount rice, which means that we attach less value to rice at a later time than at an earlier time. There is generally a good reason to apply temporal discounting to commodities such as rice. Economic growth generally causes people to be better off at later times than at earlier times, and commodities that are consumed by better-off people are generally less valuable than commodities that are consumed by worse-off people. For one thing, we can assume that commodities have “diminishing marginal benefit,” which means that the more of a commodity a person consumes, the less an extra amount of it augments her well-being. Second, if we accept prioritarianism, well-being itself has diminishing marginal value.

Discounting future commodities is not the issue in this chapter. Nor is discounting aggregate consumption of commodities. I mention it here for two reasons. The first is that it is absolutely essential to distinguish it from pure discounting, which is the issue in this chapter. Sections 30.3 and 30.4 are about pure discounting. Even if it turns out that pure discounting is impermissible, it may well be permissible to discount many particular commodities. Economists do it constantly, and I explained in the previous paragraph why they are very often right to do so.

p. 907 The second reason is to draw attention to an inconsistency in the common practice of cost-benefit analysis. Future commodities are normally discounted in cost-benefit analysis, as they normally should be for the reason I have given: a commodity consumed by a better-off person has less value than the same commodity consumed by a worse-off person. This reason explains why future commodities should be discounted, and it also applies among contemporaries: a commodity consumed by a better-off person has less value than the same commodity consumed by a worse-off person, whether or not both people live at the same time. Yet in cost-benefit analysis commodities are typically valued according to their price in money, and when they are contemporaries, both better-off and worse-off people pay this same price. So a commodity at any particular time is assigned the same value, whoever consumes it. Best practice in cost-benefit analysis is to correct this error by making adjustments to the values of commodities (Drèze and Stern 1987; Boadway, chapter 3, this Handbook), but with some exceptions (e.g., Azar and Sterner, 1996) actual practice rarely follows this best practice. So the inconsistency remains: adjustments to values are made intertemporally in the form of discounting, but they are rarely made intratemporally.

30.3 Arguments for and against Discounting Lives

A strong consensus is opposed to pure discounting (e.g., Parfit 1984, appendix F; Pigou 1932, 29–30; Ramsey 1928, 261; Solow 1974, 9). The consensus is backed largely by a conviction that morality must be impartial between people. We may rhetorically ask, “How could it possibly be true that some people count for less than others, just because they live later in time?” This question takes it for granted that impartiality is the default position, so that a deviation from impartiality needs justification. It assumes that impartiality itself does not need justification, and not much justification is generally provided.

Still, some real argument can be mustered against pure discounting. One is this (Cowen 1992). Compare two possible worlds. Exactly the same people live in each, and each person has the same lifetime well-being in one world as in the other. Also, each person lives at the same time in one world as in the other, with one exception: one person lives at different times in two worlds. This is possible; it is possible that a particular person might live at one time or alternatively at a different time. It is now possible even in practice, since embryos can be frozen and kept. If the well-being of later people is discounted in evaluating the worlds, the world in which this person lives at a later time is less good than the one in which she lives at an earlier time. This contradicts the intuitively appealing principle that, if two worlds contain the same people and are equally good for each person, then the worlds are equally good. This is a version of the Pareto principle.

p. 908 This is not a very powerful argument against discounting. It is an arcane application of the Pareto principle. Anyone who is inclined to discount the well-being of later people is likely to be willing to make an exception to the Pareto principle in this application of it.

A stronger argument starts by recognizing an argument in favor of discounting. There is an effective objection to the intuitive assumption of impartiality (Arrow 1999). The intuition behind impartiality is not as strong as I suggested initially. In some cases intuition actually seems opposed to impartiality. Intuitively, it is at least morally permissible, and perhaps morally required, for a parent to give more weight to her children’s well-being than to other people’s. More generally, intuition suggests we may morally give more weight to the well-being of those who are near us rather than to distant strangers’ well-being. Given that,

intuition may not be opposed to discounting, which gives more weight to the well-being of people who are near in time than to the well-being of those who are distant in time.

But there is a cost to making the value function partial in this way: the theory of value that supports it will almost certainly have to be *relativist* as well as partial. A relativist theory takes the value function to be relative to a particular standpoint, so that each standpoint has a value function of its own. For example, one theory takes value to be relative to people, so that each person has her own function. Another takes value to be relative to times, so that there are different value functions for different times. A value theory that supports parents in giving more weight to their children's well-being than to other people's is relativist. It gives each parent a value function that attaches extra weight to her own children's well-being, so different parents will have different functions. By contrast, a value theory is *neutral* rather than relativist if its value function is independent of any particular point of view, or if it adopts "the point of view of the universe" (Sidgwick 1907, 420).

A partial value function is not necessarily relativist. Take a theory of value that gives more weight to the emperor's well-being than to other people's. Its value function is partial. But the theory might claim that everyone—the emperor and everyone else—should give more weight to the emperor's well-being. If so, this theory is neutral.

Furthermore, even a theory of value that incorporates discounting can be neutral. It may claim that the same value function applies at every time, but that this function gives less value to later lives than to earlier ones. Whatever time is the present, this means that past lives have more value than present lives, and present lives have more value than future ones. The further in the past a life is, the more valuable it is. But this is an intuitively unattractive theory. The intuition that supports partiality gives more value to lives that are nearby in time, and less to those that are further away. This means giving less, not more, value to lives in the past than to present lives. I doubt anyone would accept a neutral discounting theory.

The consequence is that those authors who favor discounting also favor relativity. Specifically, they favor *temporal* relativity: they take the value function to be relative to the time of evaluation. Kenneth Arrow (1999) explicitly supports relativity; indeed he does not distinguish it from partiality, which he also supports. Arrow thinks future lives are less valuable than present one, but he does not think present lives are less valuable than past ones.

p. 909 But temporal relativity such as Arrow's is subject to a strong objection. It makes value relative to a point of view that is a time. As time passes, a single agent occupies a sequence of these points of view, and so has a sequence of different value functions. Inevitably, what is best according to the function at one time may not be best according to the function at another time.

For example, when a government plans ahead in 2020, according to a relativist theory it will rightly give more value to the well-being of people who are alive at that time than to the well-being of those who are not yet born. On that basis, it may rightly judge that a particular policy is the best. If we assume it ought to adopt the best policy, that is the one it ought to adopt. But if the same government re-evaluates the same policy in 2023, some people will have died and become past people, and others will have been born and become present rather than future people. According to the relativist theory, the value function will have changed by 2023. It will give less value to those who have died and more to those who have been born. Consequently, the policy that was best in 2020 may no longer be the best. From the perspective of 2023, it may be that the government ought not to have adopted the policy it did adopt in 2020. Moreover, it may be perfectly predictable in 2020 that this change will happen by 2023. So the relativist theory implies that a government may rightly select a policy in the full knowledge that in a few years' time it will rightly judge the policy to have been wrong. Indeed, at that time the policy will actually be wrong. This sort of incoherence in policymaking is not consistent with rational agency.

This is a powerful objection to temporal relativity, and consequently to discounting. (It is not an objection to relativity in general. For example, it is no objection to the relativist theory that a parent should give more value to her children's well-being than to other people's. A parent occupies only one point of view, not a sequence of points of view, so the possibility of incoherence does not arise.)

I have given two arguments against pure discounting. There are also some arguments in its favor. One is that, if we do not discount, great sacrifices will be demanded of us. If present people were to spend less on themselves, and instead invest more resources in making the world better in the future—perhaps through conventional investment in infrastructure or by emitting less greenhouse gas—the benefit to future people would persist for a very long time. When the present generation makes a sacrifice by investing, many generations receive the benefit. If the well-being of all those future generations counts equally with the present generation's, big sacrifices will be called for. When the figures are worked out, utilitarianism implies that we, the present generation, may be asked to sacrifice most of our income for the sake of future people (Dasgupta 2008). This can seem too demanding to be credible.

The problem of “demandingness” is well known in the philosophy of utilitarianism (e.g., Kagan 1989; Scheffler 1992). Because commodities generally have diminishing marginal benefit, transferring commodities from better-off to worse-off people generally increases the total of well-being. Since utilitarianism values the total of well-being, it favors this sort of transfer. This is very demanding on the world's rich, because it implies they should transfer most of their wealth to the poor. Prioritarianism is even
 p. 910 ↪ more demanding because it gives further priority to worse-off people. So is any theory that gives more value to equality than utilitarianism does.

The intertemporal problem of demandingness is different. People generally get better off as time passes. The conclusion that the present generation should sacrifice most of its income for the sake of the future implies that the relatively poor should sacrifice most of their income for the relatively rich. This makes the argument especially poignant in a way. But it also opens up a way of softening the conclusion without adopting pure discounting. We may instead adopt a different value theory from utilitarianism. Prioritarianism of a sufficient strength makes the conclusion much less demanding (Dasgupta 2008). Alternatively, we might give value to equality in some other way. So the demandingness problem does not force us to pure discounting.

It seems fair to say that the balance of argument, as well as the balance of opinion, is opposed to pure discounting of the value of people's lives.

30.4 Discounting Value at Times

The sort of discounting I have been discussing is not the sort that appears in the common practice of cost-benefit analysis. The common practice is to start by evaluating costs and benefits at particular times, and then aggregate these temporal valuations across times. Discounting then appears in the form of discounting values at times, rather than discounting the values of people's lives. In particular, the value of temporal well-being at later times may be discounted compared with the value of temporal well-being at earlier times. Is this sort of discounting justified?

Some assumptions need to be made in order to express this question accurately, just as assumptions were needed to express accurately the question of discounting lives. The chief assumption is that general value can be split into values at different times, which can then be aggregated across times. In a formula:

$$V = \bar{V}(\bar{v}^1(w_1^1, w_2^1, \dots, w_I^1), \bar{v}^2(w_1^2, w_2^2, \dots, w_I^2), \dots, \bar{v}^T(w_1^T, w_2^T, \dots, w_I^T)). \quad (30.7)$$

In this formula, times are indexed by superscripts. There are T times altogether. w_i^t is the temporal well-being of the i th person at the t th time. $\bar{v}^t()$ is the temporal value function at the t th time. It gives the general value of people's temporal well-being at that time. $\bar{V}()$ is the general value function, which is a function of the values at individual times.

To assume that general value takes this form is to assume separability between times: that values can be assigned to each time, on the basis of people's temporal well-beings at that time, in such a way that general value is determined by those values assigned to each time. This is a strong assumption with intuitively unattractive consequences.

p. 911 For one thing, it makes it impossible to give any value to lifetime inequalities between people. In section 30.2, I gave the example of a society where everyone is badly off when young and well off when old. There is inequality at any time, but people are equally well off in their lives as a whole. The formula (30.7) can recognize only the inequality at each particular time; it cannot recognize the lifetime equality. Compare a different society where there is lifetime inequality: half the people are badly off throughout their lives and the other half are well off throughout their lives. Formula (30.7) has to give the same value to both these societies.

For another thing, (30.7) gives no value to extending people's lives. More precisely, it gives the same value to extending a person's life, so that she enjoys a sequence of temporal well-beings at subsequent times, as it gives to creating a new person who will enjoy that same sequence of temporal well-beings.³

The common practice of cost-benefit analysis is committed to these unattractive consequences, and we need to accept them temporarily if we are to formulate accurately the question of discounting of well-being at times.

Next let us add further assumptions. First, let us assume that times are *additively* separable, so that

$$V = \bar{v}^1(w_1^1, w_2^1, \dots, w_I^1) + \bar{v}^2(w_1^2, w_2^2, \dots, w_I^2) + \dots + \bar{v}^T(w_1^T, w_2^T, \dots, w_I^T). \quad (30.8)$$

This is not really a further assumption; a theorem of Gorman's (1968; see also Broome 1991, ch. 4) shows that it is implied by separability between times together with separability between lives, which I assume throughout this chapter. Then let us assume that all the functions $\bar{v}^t()$ have the same form except for, possibly, a discount factor \bar{d}^t :

$$V = \bar{d}^1 \bar{v}(w_1^1, w_2^1, \dots, w_I^1) + \bar{d}^2 \bar{v}(w_1^2, w_2^2, \dots, w_I^2) + \dots + \bar{d}^T \bar{v}(w_1^T, w_2^T, \dots, w_I^T). \quad (30.9)$$

Granted the assumptions we have accepted so far, this seems very plausible.

These assumptions allow me to formulate the question of discounting: are the discount factors \bar{d}^t the same for all times t ? If they are different, and later times have lower discount factors, this would be a sort of pure discounting.

It would be different from the pure discounting of the value of lifetime well-being, which is examined in section 30.3. However, if there are different discount factors for different times, there will inevitably be different discount factors for different people, because people are born at different times and die at different times. I have already said the balance of argument is opposed to pure discounting of lifetimes well-beings. It is therefore also opposed to pure discounting of value at times.

p. 912

30.5 Population and the Intuition of Neutrality

I now turn to the question of how to set a value on the size of the earth's population. I mean its *timeless* population: all the people who will live at any time. A great many policies influence the number of future people there will be. For example, changes in taxation or in housing policy will affect people's decisions about when to have children, or even whether to have children at all. Some policies are deliberately targeted at population; China's one-child policy is a conspicuous example.

Other policies have a very direct effect on population, even if that is not their aim. An example is the provision of fertility treatment by health services. Another is the constant drive of many governments to improve their people's safety. An improvement in safety, at least if it benefits young people, nearly always increases the world's timeless population. Many young people whose lives are saved will later have children, who would otherwise never have existed. Indeed they may have a long line of descendants.

Climate change will undoubtedly affect the world's future population just because it is a huge, global event that will change so many people's lives. It also creates a small chance of a dramatic collapse of the population. There is a small chance the accumulation of greenhouse gas in the atmosphere will lead to extreme climate change, say of six, eight, or ten degrees or even more. That would be catastrophic. It would very much diminish the earth's capacity to support a human population. Our population would collapse, even possibly to extinction. If we are to make proper judgments of climate policy, the chance of catastrophe must be taken into account (Weitzman 2009). That means setting a value on a collapse of population, and on our extinction.

When a policy affects the population, either by causing more people to be created or by preventing people from being created, properly evaluating the policy must include evaluating this effect. Some judgment about the value of population is at least implicit. China's policy is clearly based on the view that it is bad to add to the country's population. But little coherent analysis of the value of population is to be found in the public discussion of policy.

Indeed, the common practice is to ignore the effects of a policy on population. For example, it is perfectly predictable that saving the lives of young people will lead to more births. Yet, although safety measures are regularly assessed by means of cost-benefit analysis, their effect on population is always ignored. Many countries include in their analysis a value for extending the lives of the people who are saved by a policy, but none includes a positive or negative value for lives that the policy adds to the population. Implicitly this means assigning those lives a value of zero.

p. 913

The UK National Institute for Health and Care Excellence (NICE) provides another example. NICE's job is to evaluate various sorts of healthcare. To measure the benefit of extending the life of an existing person, it uses "QALYs"—quality-adjusted life years. But when NICE turned its attention to IVF—in-vitro fertilization—it gave no positive or negative value to the people who are created as a result of the treatment. It did at least recognize a need to explain why not. It says (National Institute for Health and Clinical Excellence 2013, 54; internal quotations are from Devlin and Parkin 2003 and Collins 2002):

It is not logical to try to derive a quality-adjusted life year (QALY) measure from live births arising from IVF. It has been argued that:

—QALYs are intended to capture improvements in health among patients. They are not appropriate for placing a value on additional lives. Additional lives are not improvements in health; preventing someone’s death is not the same as creating their life and it is not possible to improve the quality of life of someone who has not been conceived by conceiving them.

Another review stated that:

—Cost-utility analysis has little relevance to the management of infertility where lives are produced and not saved.

This is a valid argument, so QALYs cannot be reported in the context of assisted reproduction unless they are related only to the couple seeking treatment.

This passage constitutes a poor defense of NICE’s method. It argues by means of quotations from other authors. The second quotation is not an argument at all, although NICE calls it one; it is a mere assertion. The first quotation does contain some argument, but it is not clear what conclusion it is supposed to lead to. The explicit conclusion is that QALYs are not appropriate for placing a value on additional lives because they are intended to measure improvements in health only. Even if that were true, it would be no reason not to place a value of some other sort on additional lives. Yet NICE draws the conclusion that it should not attach value to additional lives, which means it attaches zero value to them. The remark that “it is not possible to improve the quality of life of someone who has not been conceived by conceiving them” could form part of a real argument for NICE’s conclusion, but it is far short of a complete argument.

So NICE has no good argument. I believe it is in the grip of an intuition that very many of us share, which I call “the intuition of neutrality.” We think intuitively that adding a person to the population of the world is not in itself a good thing or a bad thing: it is neither better nor worse than not adding her. We recognize that her presence may do good or harm to other people. She will make some demands on the earth’s resources, leaving less for other people, and she may also make beneficial contributions. But we do not think that her own existence, and whatever well-being she herself enjoys, counts either for or against her creation. In the nice words of Jan Narveson (1976 [1973], 73), “We are in favour of making people happy, but neutral about making happy people.”

The intuition of neutrality is attractive whatever the well-being of the person who is created, within limits. Whether the person will be well off or less well off, we think her existence has neutral value in itself. But most of us recognize some limits. Most of us think intuitively that it would be bad to bring someone into existence if she will have a life of suffering. Some of us think it would be good to bring someone into existence if her life would go outstandingly well. But nearly all of us think intuitively that there is a range of levels of well-being, which we may call the “neutral range,” such that adding to the population a person whose well-being is within that range is neither better nor worse than not adding her. That is the intuition of neutrality more accurately formulated.

Figure 30.1

 The problem with the strong intuition of neutrality

The problem with the strong intuition of neutrality

I call this, more precisely, the *weak intuition of neutrality*. The *strong intuition of neutrality* is the intuition that there is a neutral range of levels of well-being such that adding a person to the population whose well-being

is within that range is equally as good as not adding her. The weak intuition allows that adding the person may be incommensurate in value with not adding her, whereas the strong intuition insists it is equally as good as not adding her.

I think either the weak or the strong version of this intuition underlies the common practice of neglecting changes in population when evaluating policies. Policymakers assume that adding to or subtracting from the population is in itself neutral in value, so they ignore it. They evaluate policies on the basis of their effect on people who exist or will exist anyway, whatever the policy.

But the common practice cannot correctly be justified this way. The strong version of the intuition is false, and the weak version is inconsistent with the common practice, as I shall now explain.

Figure 30.1 illustrates why the strong version is false. The figure shows three alternative worlds. *A* contains four people; *B* and *C* contain the same four, and a fifth person as well—the same person in each. The vertical dimension in the diagram shows the people's well-being; the higher up a person is, the better off she is. Since the extra person in *B* is within the neutral range and everyone else is equally as well off in *B* as in *A*, *B* is equally as good as *A* according to the strong intuition of neutrality. For the same reason, *C* is equally as good as *A* according to the strong intuition. It follows that *B* and *C* are equally as good as each other. But this is plainly false. *B* and *C* are the same in all respects except that the added person is better off in *B* than in *C*. So *B* is better than *C*. The strong intuition of neutrality implies a falsehood. It is therefore false itself.

p. 915

Figure 30.2

The problem with the weak intuition of neutrality

The problem with the weak intuition of neutrality

The weak intuition of neutrality cannot be shown in the same way to be false. However, it is inconsistent with the common practice of ignoring a policy's effect on population when evaluating the policy. The argument for this point is illustrated in figure 30.2.

Imagine *A* in figure 30.2 is the status quo, and we are considering a policy that takes us to *C*. This policy has the effect of diminishing the well-being of one person (the fourth), and adding an extra person. Think of it as a stripped-down version of a policy that causes the population to grow at the expense of existing people. Perhaps the fourth person suffers because the added fifth person uses up some resources that in *A* are available to the fourth person.

If we evaluate this policy according to the common practice of ignoring the added person, we shall conclude that *A* is better than *C*. It is better for one of the existing people, and we ignore the addition of a further person. But that conclusion is inconsistent with the weak intuition of neutrality. To see why, start by comparing *C* with *B*. *C* is plainly better than *B*. It contains all the same people as *B*, but in *C* those people have more well-being in total than in *B*. Moreover, that well-being is more equally distributed in *C*. Any reasonable theory of value would agree that *C* is better than *B*. If *A* is better than *C*, as the common practice implies, it follows that *A* is better than *B*. But the weak intuition of neutrality says that *A* is neither better nor worse than *B*. In particular, *A* is not better than *B*. So the common practice is inconsistent with the weak intuition.

Yet it is this very intuition that sustains the common practice. Nothing else could justify it. We have to conclude that the common practice is mistaken. In evaluating policies, added or subtracted people should not be ignored. The means that evaluations need to be based on some theory about the value of population. I now turn to the available theories.

p. 916

30.6 Valuing Population

There has been great deal of writing in moral philosophy about the value of population. It was initiated by Derek Parfit in Part IV of *Reasons and Persons*. The topic has proved intractable. Every theory turns out to conflict with some appealing intuition. This unfortunate state of affairs has been crystallized in a series of impossibility theorems by Gustaf Arrhenius (forthcoming). Arrhenius shows that various sets of attractive intuitions are inconsistent with each other. So, whatever theory we adopt, we have to brace ourselves for some intuitively unattractive consequences. Nevertheless, in choosing among theories, we have no alternative to accepting the guidance of our intuitions to some extent. We have to look for a theory that is internally consistent and conflicts with intuition only to an extent we can tolerate.

I shall outline some theories. They are not the only ones available, but they are the ones I believe to be best supported. I shall get to them by making assumptions that cut through difficulties. I do not pretend the assumptions are unquestionable.

To start with, notice something more about the example in figure 30.2. I used that example to show that the common practice in evaluation is inconsistent with the weak intuition of neutrality, but the example also exposes a difficulty with the intuition itself. Given that *C* is better than *B* in the example, the intuition implies that *C* is not worse than *A*. This is because it implies that *B* is not worse than *A*. Yet anyone attracted by the intuition of neutrality should find this conclusion unattractive. If we imagine moving from *A* to *C*, two things happen. First, the fourth person is made worse off by the move; this is a bad change. Second, an extra person is added with a well-being within the neutral range; this is a neutral change. A bad change and a neutral change should together amount to a bad change, so the result should be that *C* is worse than *A*. Yet the conclusion is that *C* is not worse than *A*. The change from *A* to *C* is neutral overall.

The problem is that the addition of the extra person, though supposedly neutral, is able to swallow up the badness of diminishing one person's well-being. This sort of neutrality is "greedy," as I put it. Greedy neutrality is not intuitively neutral at all. Intuitively a neutral change should not be able to neutralize a bad change.

This is why the weak intuition is inconsistent with the common practice of ignoring changes in the population. Even though a change in population may be neutral in a sense, according to the weak intuition it may cancel out other good or bad changes. For example (Broome 2004, sec. 14.2), suppose we need to evaluate a policy aimed at slowing global warming. This will benefit people, a definitely good thing. It will save the lives of very many people who would otherwise be killed by the effects of climate change. But suppose it will also cause a change in the population, either by either increasing or decreasing it. This change is supposedly neutral, but it may be enough to swallow up the good done to people. So without evaluating the change in population, we cannot know whether the policy is a good one, even though it saves people's lives and does nothing bad according to the intuition. This is disturbing.

I conclude that the weak intuition of neutrality is not really an intuition of neutrality as we naturally think of it. Consequently, it is not as intuitively attractive as it seemed at first. It may nevertheless be true. It allows for incommensurateness between worlds that have different populations, and it may well be that a correct theory of the value of population incorporates incommensurateness. I shall come back to this question at the end of this section.

But first, for the sake of making progress, I shall develop a theory on the assumption that there is no incommensurateness. At the end of the section I shall explain how the resulting theory may be generalized to incorporate incommensurateness once again. So for the moment I shall assume, of any pair of worlds, that if neither is better than the other, they are equally good.

This means that there is an ordering of worlds by their goodness. There is therefore a value function that represents this ordering. It has the form

$$V = V(w_1, w_2, \dots, w_K). \quad (30.10)$$

This looks like (30.1), which is a value function for a fixed population. But K is the number of people who might or might not exist at any time. Many of them will not actually exist in each particular world. When the person i exists, w_i is her well-being. When she does not exist, I give w_i an arbitrary value that indicates her nonexistence. I write it Ω .

Let us assume that, when a particular number of people exist, it does not matter which particular people they are. Let us also assume that value is impartial between people in other respects too. Then the value of $V()$ is unaffected by permuting its arguments—well-beings and Ω s—among people. This is an extension of the assumption of impartiality I made in section 30.2. One consequence is to rule out discounting the well-being of later lives compared with earlier ones. Since I argued against this sort of discounting in section 30.3, I shall not allow for it here.

Take a world in which there are I people. Because of impartiality, we may take $V()$, delete all the arguments that have the value Ω , and arrive at the value function:

$$V = V(w_1, w_2, \dots, w_I). \quad (30.1)$$

This function serves to evaluate all worlds with the same population I . It is just (30.1) from section 30.2. So when the population is held fixed, (30.10) implies (30.1). Turning around, this means we can see our task as starting from (30.1) and extending that formula to cover different populations.

We can build on the work of section 30.2, therefore. Let us take advantage of another assumption made in section 30.2: that the value function is additively separable. Together with impartiality, this gives us

$$V = v(w_1) + v(w_2) + \dots + v(w_I)$$

as our formula for the value of worlds with a given population I .

Next, we can learn something more from figure 30.1. There cannot be a neutral range of well-being such that adding a person to the population within that range is equally as good as not adding her. If there were, it would lead to the conclusion that B and C in figure 30.1 are equally good, which they are not. There must be just one single neutral level of well-being such that adding a person at that level is equally good as not adding her. Adding a person at a higher level is better than not adding her. Adding a person at a lower level is worse than not adding her. (I ignore two other possibilities: that adding a person is always worse than not adding her [Fehige 1998], or always better than not adding her, whatever her well-being.)

This conclusion does not depend on any particular feature of figure 30.1; it is quite general. Whatever world we start from—however many people it contains at whatever levels of well-being—there is always just one neutral level. However, this level may vary according to the state of the initial world.

For example, the theory known as “average utilitarianism” implies that the neutral level varies. According to average utilitarianism, the value of a world is equal to the average well-being of the people it contains.

For any population I , its value function is

$$V = (w_1 + w_2 + \dots + w_I) / I. \quad (30.11)$$

This is consistent with (30.4). It implies that the neutral level for adding a person to a world is the average well-being of the people who already exist in that world. Adding a person with a well-being above this average improves the world by pulling up its average; adding a person with a well-being below the average worsens the world by pulling down its average.

Average utilitarianism is implausible, however. Remember we are dealing with the well-beings of all the people who live at any time. The average of their well-beings therefore depends on the well-being of people who lived long ago and of those who will live far in the future. So average utilitarianism implies that the value of adding a person to the world in the 21st century depends on what the standard of living was in the Stone Age and what it will be in future ages. This is implausible.

Indeed, there are good arguments why in general we should take the neutral level to be constant, and not dependent on the initial state (Broome 2004, ch. 13). From now on I shall assume that the neutral level of well-being is constant.

Because (30.4) has an additively separable structure, the personal value function $v()$ in (30.4) is only cardinally significant. That is to say, if $v()$ is rescaled and its zero reset, (30.4) will still represent the same ordering of worlds with population I . So we are free to set the value of the neutral level of well-being to zero. Doing so is very convenient, because it means that (30.4) then serves as a general value function for any size of population. To evaluate any world, take the well-beings of everyone in it, transform \hookrightarrow their well-beings by the function $v()$, making sure the zero of the function is set to the neutral level. Then add up. The sum gives the value of the world. In this way, (30.4) represents a general theory of value for all values of I . Let us call it “the additive theory.”

If the personal value function $v()$ is strictly concave, the additive theory represented in (30.4) is a prioritarian formula. If $v()$ is linear, it is utilitarian. It implies that, of two worlds that have the same population, the one with the greater total of well-being is better. It may be written

$$V = (w_1 - u) + (w_2 - u) + \dots + (w_I - u), \quad (30.12)$$

where u is the neutral level of well-being. It says the value of a world is the total of the amounts by which people’s well-being exceeds the neutral level. Charles Blackorby, David Donaldson, and Walter Bossert (2005) call this formula “critical-level utilitarianism”; “critical level” is their term for the neutral level. They support it by an axiomatic derivation. This utilitarian theory is well grounded, but support could also be mustered for the prioritarian alternative given by a strictly concave personal value function $v()$.

I assumed in section 30.2 that the quantity of well-being is cardinally significant, but so far I have left the zero of well-being unspecified. One option is to set it at the neutral level; this is one available normalization for the scale of well-being. Then u would be zero, and (30.12) would say simply that general value is the arithmetic total of individual’s well-being. This is the theory known as “total utilitarianism.” Total utilitarianism implies that adding a person to the world is equally as good as not adding her if and only if her well-being is zero. That is to say, it implies that the neutral level is zero.

Total utilitarianism and critical-level utilitarianism are sometimes treated as different theories, but they are really only different presentations of the same theory. In total utilitarianism, the scale of well-being is normalized to make the neutral level zero. In critical level utilitarianism it is not; a different normalization may be chosen instead. There is no real disagreement between critical-level utilitarianism and total utilitarianism.

However, there can be real disagreement about what the neutral level is. Whether we accept only the additive theory of value in general or the special utilitarian version of it, we still have to settle the neutral level. What is the level of well-being such that adding a person to the world at this level is equally as good as not adding her?

One version of hedonism offers an answer to this question. Hedonism is the idea that the only good things are good experiences and the only bad things are bad experiences (Haybron, chapter 12, this Handbook). One version of hedonism applies this idea to evaluating people's lives; it claims that the lifetime well-being of a person is the excess of good over bad experiences that the person enjoys in her life. I call this "personal hedonism." Another version applies the idea to evaluating the world as a whole; it claims that the value of the world is the excess of good over bad experiences it contains. I call this "general hedonism." If a person has no good or bad experiences in her life (for example, if she lives her whole life in a coma), adding this person to the population has no effect on the excess of good over bad experiences contained in the world. General hedonism therefore implies that adding this person is equally as good as not adding her. That is to say, the neutral level of lifetime well-being is the level of a life that has no good or bad experiences. Let us call this a "blank life."

That is the implication of general hedonism. But hedonism as a whole is open to serious objections (e.g., Nagel 1970) and general hedonism to more (Broome 2004, 208–9). Suppose someone would, if she lived, have no good or bad experiences except for one fleeting good one. Would it be better that this person lives than that she does not? General hedonism implies that it would, but it is easy to doubt this conclusion.

Moreover, there is a widely accepted objection to general hedonism's claim that the neutral level is the level of a blank life. Take a world that contains many people, all with very good lives. The additive theory implies that there is a better world containing many more people all with lives just above the neutral level. But if the neutral level is the level of a blank life, a life just above this level would be very meager. So, given the additive theory, general hedonism implies that a world containing many people all with very good lives is worse than a world containing many more people all with very meager lives. Derek Parfit calls this conclusion "repugnant" (1984, ch. 17), and his view is widely shared. Few philosophers are willing to accept a theory that has this conclusion, so (if they accept the additive theory) they will not accept general hedonism.

So the additive theory of the value of population leaves a crucial parameter—the neutral level—unsettled. It is hard to settle it. Indeed, there might be no determinate answer to the question of what it is. The difficulty of identifying a single neutral level gives us some reason to think there might not be one. Instead, there might be a neutral range rather than a single level. This takes us back to the intuition of neutrality that I set aside near the beginning of this section. We might think there is a range of levels of well-being such that adding to the population a person at that level is neither better nor worse than not adding her. We have a new reason for thinking this might be so, which is the difficulty of identifying a single neutral level.

The additive theory can easily be extended to cover the possibility of a neutral range. We can use the method of supervaluation (Blackorby, Bossert, and Donaldson 1996; Rabinowicz 2009; Broome 2004, sec. 14.2). If there is a neutral range, take each level of well-being within the range in turn, treat it as the neutral level, and evaluate worlds on that basis. One world is better than another if and only if it is better when evaluated

according to every level within the neutral range. A consequence will be that some pairs of worlds will be incommensurate with each other: neither will be better than the other.

So a neutral range is formally easy to incorporate in the theory. But remember that, as I explained at the beginning of this section, when there is a neutral range as opposed to a single neutral level, the sort of neutrality that emerges is not intuitively truly neutral. It is “greedy,” as I put it. I said that a theory of the value of population is bound to conflict with intuition somewhere, and this is an example.

Still, in the additive theory we have the basis of a workable and defensible general theory of value, which can take into account changes in population.

30.7 The Nonidentity Effect and Justice

p. 921

A different range of issues revolve around the identities rather than the numbers of people in future generations. The identity of a person depends on the particular pair of sperm and egg she originated from. Someone who originated from a different sperm or a different egg would not be the same person. This means that a person’s identity is sensitive to conditions at the time of her conception. Had conditions been slightly different, that person would not have existed. A different person might have existed instead. So even a slight change in people’s way of life will mean that different people will be created.

Government policies make a difference to people’s lives. For example, a change in the tax system or in environmental regulations changes people’s behavior. People may travel less or more, take different jobs or do less overtime, and generally participate differently in society. People will form different partnerships, perhaps have babies with different partners, and have babies at different times. This means that different babies will be born. Any policy that makes a significant difference to people’s lives, such as a carbon tax or another policy aimed at slowing climate change, will mean that the next generation will have different people in it. Within just a few generations, the entire population of the world will consist of different people from those who would have lived had the policy not been adopted.

This is the “nonidentity effect.”⁴ It means that, when we think about policies that have a long-term effect, we have to be cautious about applying moral principles that depend on people’s identity.

One group of these principles are principles of justice. Very roughly, our moral duties can be divided into duties of beneficence—promoting good—and duties of justice. (There may be duties of other sorts too, but these are the ones I need to take note of here.) The need for this division can be illustrated by an example that comes from Judith Jarvis Thomson (1985, 1396). A surgeon has in her hospital five patients, each in need of a different organ for transplant. One needs a new heart, one a new liver, and so on. Each will die unless she gets a new organ. The surgeon kills an innocent visitor to the hospital and distributes her organs to her patients. Thereby the surgeon saves five lives at the expense of one. She successfully promotes good. Yet she clearly acts wrongly.

What could explain that? It must be that there is some principle of morality that can conflict with the duty to promote good, and is sometimes important enough to override it. In this particular case, it is evidently a duty not to inflict harm on someone, even for the sake of greater general goodness. I take this to be a duty of justice. The characteristic of duties of justice is that each duty is owed to a particular person. If you act unjustly, there is always someone to whom you have done an injustice. We can say this person has a right that you have infringed; rights are correlative to justice. In the example, the innocent visitor has a right not to be harmed even for the sake of the greater general good.

p. 922

In general we each owe a duty of justice not to harm other people. Governments have the same duty not to harm individuals even for the greater good. Individuals have the right not to be harmed by governments or

by other people. There are limits to this right. For example, it is sometimes permissible for a government to harm an individual if it is a proper punishment for wrongdoing, or the only way to avoid great harm to others, or to achieve a great good. We do not have to investigate the limits here.

It is commonly thought that future generations possess a right not to be harmed by present people or their governments, and that a policy of permitting climate change to continue unabated infringes this right. If so, a duty of justice rests on present people and their governments to do what they can to control climate change.

The nonidentity effect casts doubt on this common thought. Suppose the present generation continues to emit greenhouse gas profligately, so that life 150 years from now is not as good as it would have been had we been less profligate. Take a person living at that date. Could we say that this person has a right not to be harmed by us, which we infringe? Had we acted differently and reduced our emissions, the nonidentity effect means that this person would not even have existed. Nor would her right have existed, therefore. So this is a right that cannot be satisfied; doing what is required by the right puts the right out of existence. It is hard to believe there could be a self-destroying right like this. If there is not, then the present generation does not owe a duty of justice to future generations.

This is not a conclusive argument. Various responses might be made. Perhaps there are indeed self-destroying rights. Or the present generation might owe a duty, not to particular people living in the future, but to future generations, considered as collective entities of a particular sort. A generation would maintain its identity in different possible worlds, even though in different worlds it would be made up of different people.

The responses are not conclusive either. There remains a good reason to doubt that our moral duties to future generations can be seen as duties of justice. If not, they must be duties of beneficence. We have a duty to promote the good of future generations, by which I mean a duty to make future people better off than future people might have been. This duty is not affected by the nonidentity effect. It is not a duty of justice, but this does not mean it is not a stringent duty.

Even if we owe no duty of justice to future generations, it does not follow that no considerations of justice need be taken into account in policymaking for the future. Even if the nonidentity effect means that members of future generations cannot suffer injustices from the present generation, they may still suffer injustices from their contemporaries. Moreover, present policy may have an effect on the amount of injustice that future people do suffer from their contemporaries.

Economic inequality provides an example. It is sometimes said to be distributively unjust if equally meritorious and deserving people do not receive equal benefits from life. As it happens, I think this is strictly an unfairness rather than an injustice (Broome 1991, ch. 9), but fairness is allied to justice, and it shares with justice the feature \hookrightarrow that its duties are owed to particular people. It is safe to ignore the difference here, and use the widely accepted term “distributive justice.” Distributive justice is owed by contemporaneous people to each other.

p. 923

Present policies may have an effect on future distributive justice. For example, climate policy may be conducted in a way that helps to reduce global inequality, or in a way that does not. Clearly this difference should be taken into account in evaluating the policies. But if policy toward future generations should be governed by the duty of beneficence, aimed at promoting goodness rather than at justice, how can it take account of distributive justice?

By treating injustice as itself a bad thing. Future injustice should be set against other future goods, such as people's well-being. This is commonly done: economic analysis of policy commonly treats equality as a good and inequality as a bad, to be accounted alongside the total of people's well-being as another good

(e.g., Atkinson 1970). The term “value of equality” is used in economics as an alternative term to “distributive justice,” with the same meaning.

This does not contradict what I have been saying. Promoting future justice is promoting a sort of good. It falls under beneficence. It is not itself a duty of justice. It is something the present generation should do as part of its general duty to promote good. But the nonidentity effect means it is not a duty owed by the present generation to future generations.

In this chapter, we have been considering how to take account of the well-being of future people in policymaking. This is a matter of judging the goodness of different options for policy. This section has argued that this is indeed the appropriate aim of long-term policymaking; it should be regulated by the duty to promote goodness rather than the duty of justice. Nevertheless, avoiding future injustices, including future inequality, should be included as one of the goods to be promoted.

30.8 The Nonidentity Effect and the Pareto Principle

The nonidentity effect disrupts justice, and it also disrupts the application of the Pareto principle. When there is a nonidentity effect, the Pareto principle is false in some of its formulations, and inapplicable in others.

One formulation is the following: Necessarily, A is better than B if no one in B has more well-being in B than she has in A and one person in B has less well-being in B than she has in A . (30.13)

p. 924

When there is a nonidentity effect, this is plainly false. Suppose nearly everyone in B possesses a high level of well-being, but none of those well-off people exists in A . Nearly everyone in A has a low level of well-being. But each of the few people who exist in both A and B has less well-being in B than in A . Then the condition in (30.13) is satisfied, so (30.13) implies that A is better than B . But that may not be so, since nearly everyone in B is well off and nearly everyone in A is badly off.

Another formulation is the following: Necessarily, A is better than B if each person in A has at least as much well-being in A as she has in B and one person in A has more well-being in A than she has in B . (30.14)

This may be true. But the condition in this formulation can be satisfied only if everyone who exists in A also exists in B . When there is a nonidentity effect between A and B , this condition cannot be satisfied. So the Pareto principle formulated this way cannot be applied.

The Pareto principle depends on there being an identical population in the two alternatives being compared. It inevitably fails in the face of the nonidentity effect.

This means that, when thinking about future generations, we cannot rely on all the standard conclusions of welfare economics. For example, it is a standard conclusion that (except in some very special circumstances) an externality leads to Pareto inefficiency. That is to say, it leads to a state of the world such that some other state is better according to the Pareto principle. But if the externality has a nonidentity effect, this standard conclusion may be false. Greenhouse gas is a powerful externality, and it is natural to expect it to lead to Pareto inefficiency.⁵ Yet it may not. Take a world B in which there are emissions of greenhouse gas that damage the quality of life of people living in the future. The nonidentity effect entails that any world A with less emissions than B will be one where some of the future people in A do not exist in B . So it will not be true that each person who lives in A has at least as much well-being in A as she has in B . The condition in (30.14) will not be satisfied.

This is damaging to the welfare economics of future generations. It would be useful if we could extend the conclusions of welfare economics to cases that involve a nonidentity effect. To do so successfully, we shall need to replace Pareto efficiency with some other notion of efficiency.

A natural first idea (Suppes, 1966) is to adopt this extended Pareto principle:

Necessarily, *A* is better than *B* if the same number of people live in *A* as in *B*, and if there is one-one correspondence between the people in *A* and the people in *B* such that each person in *A* has at least much well-being as the corresponding person in *B*, and one person in *A* has more well-being than the corresponding person in *B*. (30.15)

p. 925

This seems very plausible, but it is too demanding to yield standard conclusions of welfare economics when there is a nonidentity effect. For example, suppose that \hookrightarrow reducing greenhouse gas emissions causes the creation of some people who are congenitally grumpy. Let *B* be a world with emissions. There may be no world *A* with reduced emissions that is better according to (30.15). It might be that the grumpy people in *A* have less well-being than anyone in *B*. The externality is therefore not inefficient by this criterion. So (30.15) does not supply the notion of efficiency we need if we are to preserve the standard conclusions.

Examples like the one I have just given suggest it would be better to attend to the resources that are available to future people, rather than to the well-being future people derive from their resources. Well-being is too intimately connected with identity. Here is an alternative idea. Let us say that a state is “weakly inefficient” if it would be possible to improve the well-being of one person in the present generation, without reducing the well-being of anyone in the present generation, and also without diminishing any of the resources that are available to future generations.

Except in some very special circumstances, an externality leads to weak inefficiency, unless it affects no one in the present generation. But this weak inefficiency is too weak to be very interesting. It captures intragenerational inefficiency, but there is intuitively a sort of intergenerational inefficiency that it does not capture.

Here is an example. The present generation’s greenhouse gas emissions mean that future generations will be left with a dirty atmosphere, but they will also be left with some amount of conventional capital such as economic infrastructure. It would be possible for the present generation to switch its investment from building conventional capital to cleaning up the atmosphere instead. Since this is a switch of investment rather than consumption, it could be managed without any sacrifice of the present generation’s well-being; each present person could remain as well off as before. The switch could provide future generations with resources that make them materially better off in some sense. They would receive from the present generation less conventional capital but a cleaner atmosphere. The increase in one resource could more than compensate for the diminution of the other. The switch would then be intuitively an improvement in efficiency. But it is not an improvement in weak efficiency, since the resources left to future generations are not exactly the same; one resource has been substituted for another. We need a wider notion of efficiency.

I conclude that more work needs to be done in welfare economics to take proper account of the nonidentity effect.

30.9 Conclusion

I have surveyed a few of the issues that arise when we try to evaluate alternative ways the world might go, taking into account the well-being of future generations.

I considered whether the well-being of future people, or more exactly the value of their well-being, should be discounted, and argued it should not.

p. 926 I then turned to issues that surround the numbers of future people and their identities. These issues are commonly ignored in practical policymaking, but I argued they should not be ignored.

I proposed in outline a theory of value that takes into account changes in the numbers of future people.

Finally I considered problems raised by the “nonidentity effect”—the fact that present actions can alter the identities of future people. I argued that it means we should not see our duties to future generations primarily in terms of justice, though we need to take account of intragenerational injustices as a sort of bad. I also argued that the nonidentity effect raises a difficulty for welfare economics over applying the Pareto principle to future generations. This difficulty is not yet solved.

Acknowledgments

Research for this chapter was supported by the Australian Research Council Discovery grant number DP140102468.

References

Adams, Robert. 1979. "Existence, Self-Interest, and the Problem of Evil." *Nous* 13: 53–65.

[Google Scholar](#) [WorldCat](#)

Adler, Matthew. 2012. *Well-Being and Fair Distribution*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Arrhenius, Gustaf. Forthcoming. *Population Ethics: The Challenge of Future Generations*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Arrow, Kenneth. 1999. "Discounting, Morality, and Gaming." P. R. Portney and J. P. Weyant, eds., *Discounting and Intergenerational Equity*, 13–21. Washington, DC: Resources for the Future.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2: 244–63.

[Google Scholar](#) [WorldCat](#)

Azar, C., and T. Sterner. 1996. "Discounting and Distributional Considerations in the Context of Global Warming." *Ecological Economics* 19: 169–84.

[Google Scholar](#) [WorldCat](#)

Blackorby, Charles, Walter Bossert, and David Donaldson. 1996. "Quasi-Orderings and Population Ethics." *Social Choice and Welfare* 13: 129–50.

[Google Scholar](#) [WorldCat](#)

Blackorby, Charles, Walter Bossert, and David Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics and Ethics*. New York: Cambridge University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Broome, John. 2004. *Weighing Lives*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Broome, John. 2012. *Climate Matters: Ethics in a Warming World*. New York: Norton.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Collins, J. 2002. "An International Survey of the Health Economics of IVF and ICSI." *Human Reproduction Update* 8: 265–77.

[Google Scholar](#) [WorldCat](#)

Cowen, Tyler. 1992. "Consequentialism Implies a Zero Rate of Intergenerational Discount." Peter Laslett and James S. Fishkin, eds., *Justice between Age Groups and Generations*, 162–8. New Haven: Yale University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Dasgupta, Partha. 2008. "Discounting Climate Change." *Journal of Risk and Uncertainty* 37: 141–69.

[Google Scholar](#) [WorldCat](#)

p. 927 Dasgupta, Partha. Forthcoming. "Time and the Generations."

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Devlin, N., and D. Parkin. 2003. "Funding Fertility: Issues in the Allocation and Distribution of Resources to Assisted Reproductive Technologies." *Human Fertility* 6: S2–S6.

[Google Scholar](#) [WorldCat](#)

Drèze, Jean, and Nicholas Stern. 1987. "The Theory of Cost-Benefit Analysis." Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, vol. 2, 909–89. Amsterdam: North-Holland.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Fehige, Christoph. 1998. "A Pareto Principle for Possible People." Christoph Fehige and Ulla Wessels, eds., *Preferences*, 508–43. Berlin; New York: de Gruyter.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Foley, Duncan. 2009. "The Economic Fundamentals of Global Warming." Jonathan M. Harris and Neva R. Goodwin, eds., *Twenty-First Century Macroeconomics: Responding to the Climate Challenge*, 115–26. Northampton, MA: Edward Elgar.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Gorman, W. M. 1968. "The Structure of Utility Functions." *Review of Economic Studies* 35: 367–90.

[Google Scholar](#) [WorldCat](#)

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63: 309–21.

[Google Scholar](#) [WorldCat](#)

Kagan, Shelly. 1989. *The Limits of Morality*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

McKerlie, Dennis. 1989. "Equality and Time." *Ethics* 99: 475–91.

[Google Scholar](#) [WorldCat](#)

Nagel, Thomas. 1970. "Death." *Nous* 4: 73–80. Reprinted in his *Mortal Questions*, 1–10. New York: Cambridge University Press, 1979.

[Google Scholar](#) [WorldCat](#)

Narveson, Jan. 1973. "Moral Problems of Population." *Monist* 57: 62–86. Reprinted in Michael D. Bayles, ed., *Ethics and Population*, 59–80. Cambridge, MA: Schenkman.

[Google Scholar](#) [WorldCat](#)

National Institute for Health and Clinical Excellence. 2013. *Fertility: Assessment and Treatment for People with Fertility Problems*.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Parfit, Derek. 1976. "On Doing the Best for Our Children." Michael D. Bayles, ed., *Ethics and Population*, 100–115. Cambridge, MA: Schenkman.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Parfit, Derek. 1984. *Reasons and Persons*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Parfit, Derek. 1991. *Equality or Priority?* The Lindley Lecture, University of Kansas. Reprinted in Matthew Clayton and Andrew Williams, eds., *The Ideal of Equality*, 81–125. New York: Macmillan, 2000.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Pigou, A. C. 1932. *The Economics of Welfare*. 4th ed. New York: Macmillan.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Rabinowicz, Wlodek. 2009. "Broome and the Intuition of Neutrality." *Philosophical Issues* 19 389–411.

[Google Scholar](#) [WorldCat](#)

Ramsey, Frank. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38: 543–59. Reprinted in his *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. D. H. Mellor, 261–81. London: Routledge and Kegan Paul.

[Google Scholar](#) [WorldCat](#)

Rawls, John. 1972. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Sen, Amartya. 1973. *On Economic Inequality*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Scheffler, Samuel. 1992. *Human Morality*. Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Sidgwick, Henry. 1907. *The Methods of Ethics*. 7th ed. New York: Macmillan.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Solow, Robert. 1974. "The Economics of Resources or the Resources of Economics." *American Economic Review Papers and Proceedings* 64: 1–14.

[Google Scholar](#) [WorldCat](#)

Suppes, Patrick. 1966. "Some formal models of grading principles." *Synthese* 16: 284–306.

[Google Scholar](#) [WorldCat](#)

Temkin, Larry S. 1993. *Inequality*. New York: Oxford University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Thomson, Judith Jarvis. 1985. "The Trolley Problem." *Yale Law Journal* 94: 1395–415.

[Google Scholar](#) [WorldCat](#)

Velleman, David. 1991. "Well-Being and Time." *Pacific Philosophical Quarterly* 72: 48–77.

[Google Scholar](#) [WorldCat](#)

Weitzman, Martin. 2009. "On Modeling and Interpreting the Economics of Catastrophic Climate Change." *Review of Economics and Statistics* 91: 1–19. ↴

[Google Scholar](#) [WorldCat](#)

p. 928

Notes

- 1 Sen 1973, 39–41, contains a discussion of this question. It has subsequently been much debated. Broome 1991 contains an argument for additive separability that is derived from Harsanyi 1955.
- 2 This term came into use after Parfit (1991). For instance, see Adler 2012, ch. 5. However, prioritarianism was formulated by Sen (1973, 39–41) if not earlier.
- 3 Ponthiere (chapter 29, this Handbook) points out that this problem could be overcome by "dispersing" the value of longevity across times in a person's life. In Broome 2004, sec. 7.4, I expressed doubts about this strategy.
- 4 Early discussions of its ethical significance are in Adams 1979 and Parfit 1976.
- 5 I drew this conclusion in Broome 2012, ch. 3, following Foley (2009). Foley is basically right, but his conclusion needs some reformulation.