10 December 2018

Birth and Death

Arrow Lecture

Partha Dasgupta

Contents

Table 1

Figures 1-8

People have children for many reasons. The mix of motivations depends on the customs and institutions we inherit, as well as on our character and circumstances. That children are valuable in themselves is emotionally so compelling that it may seem too obvious to require acknowledgement, but social anthropologists have shown that children are not just valuable to us because of the innate desire we have to bear and rear them, but also because they represent the fulfilment of tradition and religious dictates, and because they are the clearest avenue open to self-transcendence. A common refrain, that our children are priceless, is an expression of how innately valuable they are to us.[1]

In places where formal institutions are underdeveloped, children also substitute for other assets, and are thus valuable for the many benefits they bring to their parents. This is most apparent in the poorest regions of the world. Children serve as security in old-age in places that have neither pension schemes nor adequate capital and land markets. They are also a source of labour in households possessing few labour saving devices. Children mind their siblings, tend to domestic animals, pick berries and herbs, collect firewood, draw water, and help with cooking. Children in poor countries are valued by their parents also as capital and producer goods.[2]

1 Economic Demography

Those childhood activities are so unfamiliar today in the West that they direct us to study the motivations governing procreation by contrasting rich regions from poor regions. There are notable exceptions of course, but broadly speaking fertility and mortality rates are high and health status and education attainments are low in poor countries, whereas the corresponding statistics in rich countries read the other way. Table 1, which presents a snap shot (roughly, the period 2014-15), speaks to that by displaying data published by the World Bank, where countries are classified according GDP per capita. I have labelled the two categories "rich" (the World Bank labels them "high income countries") and "poor" (the World Bank labels them "low income countries"). Countries have been known to make a transition from **the latter category to the former** (that's what economic development is usually taken to mean); moreover, the bulk of the world's population and many of the world's poorest people live in neither rich nor poor countries, and international statistics say there are enormously rich people in poor countries. It nevertheless pays to study sharp contrasts, as in Table 1.

Reproductive decisions and our use of the natural environment have consequences for others, including our descendants, that are unaccounted for under prevailing institutions and social mores (markets, government policy, communitarian engagements, religious injunctions). Economists use the term *externalities* to denote those consequences of our decisions for others that are not accounted for.

---

1 One such injunction emanates from the cult of the ancestor, which takes religion to be the act of reproducing the lineage. See Fortes (1978). The accompanying essay (Dasgupta and Dasgupta, 2017), reports anthropological findings on the motivations that guide fertility behaviour.

2 In South Asia children have been observed to work in household production from age six.

The qualifier "not accounted for" means that the consequences in question follow without prior engagement with those who are, or who will be, affected. The required engagements don't have to be face-to-face. Many of our actions can be expected to have consequences for our descendants; but if the actions were taken with due care and concern (we take many actions - for example, saving for the future - with our descendants very much in our mind), they would not give rise to externalities. We begin to engage with future people when we deliberate whether current rates of carbon emissions into the atmosphere will place an unjust burden on our descendants. The presence of externalities explains why and how it can be that a people are settled on a pattern of reproductive behaviour and environmental-resource use they would all prefer to alter but do not because no one has the necessary motivation to change their behaviour unilaterally. Externalities raise deep ethical issues. Not only do they extend to contemporaries and can be expected to extend to future people, it is also that some people will be born in consequence of the decisions we take, while some who would have been born had we acted otherwise will not be born.[3]

Caldwell (1981, 1982) drew on an idea that is suggested by Table 1, that the intergenerational transfer of wealth is from parents to children in rich countries but from children to parents in poor societies. The suggestion has been easier to confirm in rich countries, where the rate of investment in children's education has been found to be as high as 6-7 per cent of GDP (Haveman and Wolfe, 1995). Because a vast range of activities in poor societies are undertaken outside the institution of markets, it is especially hard to identify the direction in which resources there flow across the generations. Nevertheless, the Caldwell-hypothesis has been questioned for poor societies. Studies have found that even there the direction is from the old to the young (Lee, 2000, 2007). Further investigations may find hidden transfers from the young to the old in poor societies that confirm Caldwell's thesis, but as of now it would seem that throughout the world intergenerational resource transfers are made by the old to the young.

Differences in the social statistics in Table 1 are striking. They are traceable to kinship structures, marriage practices, and rules of inheritance. The implied line of thinking says that over the long run it is differences in institutions, beliefs, and social norms of behaviour that lie behind differences in reproductive behaviour among peoples.[4] Theoretical models have been built on that premise. Causality isn't traced to differences in income or wealth. It is not that fertility and mortality rates are high and health status and education attainments are low in poor regions *because* people there are poor, it is that very low incomes go hand in hand with those features of life. The variables

---

3 The latter should be thought of as consequences to the potential parents. Decisions on consumption and production also create externalities. They are connected to reproductive externalities. Here I focus on decisions over reproduction. We uncover the pervasiveness of externalities in Section 5.7, Section 10, and Appendices 1 and 2.

4 We should expect institutions and the norms of behaviour that support them originate in some measure on geography: tropical regions vs. temperate regions; drylands vs. wetlands; sedentary vs. migratory societies; and so on.

are mutually determined over time.[5]

Table 1 is a snapshot. It says that in comparison to people in rich countries, people in poor countries receive less basic education, have more children, die younger, enjoy fewer political and civil liberties, and suffer from greater failure in governance. There is no suggestion that poor societies will remain poor, nor that rich countries may not find their place reversed in the long run. Regional differences in fertility, education, and output per capita were slight until the start of the Early Modern era (roughly, 1500 CE). Global aggregates of earlier eras look much the same as their regional aggregates.[6] Although regional aggregates have diverged since then, global aggregates (a weighted average of regional statistics) have shown a steady move toward and beyond "fertility transition," that is, transition from high fertility and mortality rates to low fertility and mortality rates.

Economists have offered a number of explanations for the historical experience.[7] What is common to them is a presumption that parental choices over fertility, consumption, and investment determine long run outcomes. The models trace the relative urgencies of parental needs, desires, and obligations to the constraints on choices faced by parents in each generation.[8] Some authors stress economic constraints, others pay attention to social and ecological constraints. A few (as in Dasgupta and Dasgupta, 2017, reproduced here) speak especially to the pervasiveness of reproductive and

---

5 For theoretical models that speak to the mutual determination, see Dasgupta (1993, 2010), Brander and Taylor (1998), Harford (1997, 1998), Dasgupta and Ehrlich (2013), and Bohn and Stewart (2015). Crist et al. (2017) offers a description of the mutual determination using global figures. Sub-Saharan Africa has long been regarded as special, even among poor regions (Goody, 1976; **Fortes, 1978**; Bledsoe, 1994; Guyer, 1994; Bongaarts and Casterline, 2013). In an early review of fertility intentions Cochrane and Farid (1989) noted that both the urban and rural, the educated and uneducated in sub-Saharan Africa have more, and want more, children than their counterparts in other less-developed regions. Even young women there expressed a desire for an average of 2.6 more children than women in the Middle East, 2.8 more than women in North Africa, and 3.6 to 3.7 more than women in Latin America and Asia. Updated versions of these figures are available, but it is worth considering the data from the mid-1980s because the income gap between Africa and the rest of the developing world was smaller at that time than it is now.

Evolutionary biologists distinguish reproductive strategies in the animal kingdom that correspond to the two extreme types of fertility and human capital outcomes displayed in Table 1. An *r*-strategy gives rise to many offspring combined with low parental investment, and each offspring has a low probability of survival. In contrast a *K*-strategy gives rise to few offspring over a longer lifespan and high parental investment. The motivation driving animals is inclusive fitness. Each of the strategies is an equilibrium, depending on the underlying evolutionary game.

6 The classic on this is Maddison (2001), who provided estimates of expectancy of life at birth, population size, and output from 1 CE until year 1998 in various regions of the world.

7 See the references in footnote 5, and Galor (2011).

8 "Parental choices" are only a contemporary way of alluding to choices that are often, perhaps even usually, influenced by the extended family, kinship, the couple's peer group, and the power-relationship between the couple. On this see Dasgupta (1993, 2010).

environmental externalities. Economic demographers have commonly avoided moral theories in their study of reproductive behaviour.

In contrast, philosophical discourses on population have been built on normative reasoning, directed at four questions: (1) What are the nature, ground, and limits of parental responsibility for existing children? (2) Does producing a child interfere with the rights of children the couple already have? (3) Do individuals have a duty not to have children whose lives are likely to be bad for them (negative well-being)? Do they have a duty to have children whose lives are likely to be good for them (positive well-being)? (4) How should one value possible populations so as to decide which would be best?

One way to contrast the two disciplinary approaches is to say that the economic demographer's task is to explain Table 1, while the aim in population ethics is to produce a normative theory that one could use to evaluate the behaviour patterns that give rise to Table 1 and prescribe better ones. Question (4), which stands in sharp contrast to those that demographers study, is at the heart of "population axiology," which attempts to uncover the ethics involved in what Parfit (1984: 356) called Different Numbers Choices. To me it remains a puzzle though that population axiologists haven't subjected their reasoning to a world facing socio-ecological constraints of the kind we have now come to know.

In Part I population axiology is studied with an eye on resource constraints (the biosphere's ability to supply goods and services). I try to integrate demographic and environmental concerns while accommodating contemporary sensibilities over birth, life, and death. In Part II the formulation is applied to quantitative data on humanity's reliance on the biosphere. My idea there is to obtain a sense of the numbers involved, nothing more.

2 Utilitarian Ethics

In his statement of Utilitarianism, Sidgwick (1907: 415-416) wrote:

"... if we take Utilitarianism to prescribe, as the ultimate end of action, happiness as a whole, and not any individual's happiness, unless considered as an element of the whole, it would follow that, if the additional population enjoy on the whole positive happiness, we ought to weigh the amount of happiness gained by the extra number against the amount lost by the remainder. So that, strictly conceived, the point up to which, on Utilitarian principles, population ought to be encouraged to increase, is not that at which average happiness is the greatest possible ... but that at which the product formed by multiplying the number of persons living into the amount of average happiness reaches its maximum."

The account is read by philosophers as saying that the basis for evaluation is not gains and losses to people, but gains and losses in total utility. An ethics grounded on the latter reflects the view that lives have an intrinsic value, and that the better the life is for the person as measured in terms of utility, the greater is the value. Sidgwick took utility to be a numerical measure of happiness. His Utilitarianism asks us to evaluate alternative states of affair in terms of the sum of personal utilities.

State of affairs *X* is judged to be superior to state of affairs *Y* if total utility in *X* exceeds total utility in *Y*: period.[9]

Sidgwick's theory, which Rawls (1972) called Classical Utilitarianism, involves two related notions: (i) Personal happiness, whose numerical measure is individual utility; (ii) summation as the required operation for combining individual utilities. Sidgwick (1907: 119-150) contains three chapters on empirical hedonism, where the sense in which "happiness" is used is a lot more considered than is suggested in the frequent criticism that Classical Utilitarianism views humans to be mere pleasure machines. Nevertheless, I am reluctant to give the impression that the ethical theory I am led to relies on Sidgwick's notion of the personal good. So I shall use the term "well-being". Griffin (1986) contains a measured, book-length analysis of the concept in its many guises, but he also develops his preferred interpretation. Briefly, he thinks of personal well-being as a measure of the extent to which one's informed desires are realized. He also discusses measurement problems.

The qualifier “informed” in “informed desires” is meant to bear ethical weight. We are to interpret the fulfilment of informed desires as a flourishing life. While composing this essay I have kept Griffin's conception in mind. But the mathematical structure of the ethics that is constructed here is not tied to this particular notion of well-being.

Personal well-being will be denoted by *U*. It is assumed to be a function of the standard of living. By living standard I mean an aggregate measure of all the factors that influence personal well-being. In the formal models that I study here, those factors are represented by an all-purpose consumption good.

The presumption that the sole factor in well-being is consumption may seem otiose. It is widely held that personal and social engagements are central to how one's life gets shaped. Chance of course plays a role. And there are factors such as genetic endowment, which at least for now are not subject to choice by the person. We may regard them as parameters of the *U*-function. Because all engagements require goods and services, we may think of engagements as production activities, in which goods and services are inputs. Partaking of a meal (alone or in the company of others) is an engagement, but food items are the inputs that make the meal possible. Friendship involves investment in time, which is a scarce resource. And so on. Goods and services come with various characteristics, and for each person there is (given the resources at his command) a best way of obtaining the commodities that are best suited for the activities that are from his point of view the best for him to engage in. Of course, the best way of obtaining goods and services with those characteristics involves further engagements (getting a job, making contacts), and so on; but the regress is circular, meaning that it is closed. The economist's presumption that the only things people care about are goods and services should be

9 I am grateful to Krister Bykvist for correspondence on the way Sidgwick should be interpreted. The ground of binding reason in Sidgwick's population ethics is total human happiness ("happiness as a whole"), not the algebraic sum of gains and losses to people.

seen as a filtered expression of their projects and aims. Commodities have instrumental value; they don't necessarily have intrinsic worth. However, irrespective of what a person values, he will be found to value commodities. The *U*-function postulated here should be interpreted in that light.

Part I of this essay studies Sidgwick's theory as amended when happiness is replaced by well-being, and then offers reasons why we should still move away from it. I depart from the amended theory in two ways: First, I alter the way individual well-beings are aggregated by the decision maker and propose an alternative that commends itself. I am drawn to the alternative because population axiology is not only about identifying desirable demographic states of affairs, but also about the reproductive choices people can justify to themselves. Answers to the latter problem do not follow from resolutions of the former exercise. Secondly, I am drawn for several reasons to a different reading of the idea of a neutral life, which is a central notion in population axiology. **We shall call the theory I am led to, Generation-Relative Utilitarianism.**

Differences between Total Utilitarianism and **Generation-Relative Utilitarianism** are exposed in Part II. There I apply both theories to quantitative data on humanity's reliance on the biosphere.

3 Ends and Means

Humanity's future will be shaped by the portfolio of assets we inherit and choose to pass on, and by the balance we strike between the portfolio and the size of our population. Assets are durable objects. Their durability enables us to save them for our own future, offer them as gifts to others, exchange them for other goods and services, and bequeath them to our children. Durability doesn't mean everlasting. Assets depreciate (machines suffer from wear and tear, trees die), but they are not fleeting. Perhaps because financial capital has figured prominently in economists' writings, the qualifier “capital” is sometimes added to assets, as in "capital assets."

3.1 Capital Goods

Durability distinguishes **assets** from services, such as live performances, which vanish as soon as they are produced. It is true that a memorable performance is remembered long after, but memory of a performance differs from the performance itself. Memory confers durability to the performance by creating a substitute product. Recording serves a similar, in some ways a more faithful, purpose. But even though we can enjoy the recording, it's a recording we are enjoying, not the actual performance.

Assets acquire their value from the services they provide over their remaining life. A refrigerator preserves food products by keeping them cool. It provides that service until it breaks down beyond repair. The refrigerator's worth is a measure of the benefits it provides over its **remaining** life. How that measure should be constructed is something we discuss below. The important thing to remember here is that an asset's future performance is built into it today.

In common language, to say an object is an "asset" is to convey the idea that it has positive worth; assets are taken to be goods. But one virtue of moving from the literary to the formal is being able to extend the use of concepts so as to bring disparate objects under a common intellectual framework. So we note that assets not only contribute directly to our well-being, but indirectly too, as sinks for pollution (contemporary carbon emissions into the atmosphere are a prominent example). One way to conceptualize pollution is to view it as the depreciation of assets. Acid rains damage forests; carbon emissions into the atmosphere trap heat; industrial seepage and discharge reduce water quality in streams and underground reservoirs; sulphur emissions corrode structures and harm human health; and so on. The damage inflicted on each type of **asset** (buildings, forests, the atmosphere, fisheries, human health) should be interpreted as depreciation. For natural resources depreciation amounts to the difference between the rate at which they are harvested and their regenerative rate (Appendix 4); the depreciation that pollutants cause on natural resources is the difference between the rate at which pollutants are discharged into the resource-base and the rate at which the resource-base is able to neutralize the pollutants. The task in either case is to estimate depreciation. Economists have tried to estimate the damage an additional ton of carbon in the atmosphere is likely to inflict over time on agricultural production, submerged coastal habitats, **human health,** and so on (the estimates today range from 50 to 500 US dollars). Resources are "goods," while pollutants (the degrader of resources) are "bads." Pollutants are the reverse of resources and polluting is the reverse of conserving.[10]

In common parlance assets are often called capital goods. But economists typically confine the use of the term to **assets** that are material (tangible), durable, and alienable (whose ownership is transferable) - a piece of furniture, an orchard, a lathe, and so on. One reason for the latter restriction was the desire to work with **assets** that can be measured and compared with one another. It's not enough to say that houses can be measured in physical units (floor space, say), they need to be compared with other capital goods, such as automobiles. We need a common unit. Valuing assets is a way to do that. Goods that are both material and alienable can be exchanged in markets. As a possibly crude, first approximation, market prices offer a measure of their value to us.

In recent decades economists have introduced Nature into economics by viewing it as a capital good. In Part II we build on that way of looking at things and apply it to the biosphere as a whole. Of the myriad of assets that make up Nature, land is perhaps the most familiar. The market price of a hectare of agricultural land at the outer edges of a farm includes the ecological services it provides the rest of the farm, but excludes the services the neighbouring farm may enjoy (the latter would be a positive externality). So the market price of the hectare is likely to be an underestimate of its worth to the economy as a whole. Valuing a watershed is harder, but again, it is manageable. Among the services it offers neighbouring communities is purifying water. One way to value that service is to estimate the cost of purifying that same flow of water by other means (water purification plant). In a well known study of a degraded watershed (the Catskills in upper New York State),

10 A formal demonstration of the equivalence is in Dasgupta (1982).

Chichilnisky and Heal (1998) found that constructing a water purification plant so as to provide New York City with clean water would have cost far more than restoring the watershed (fortunately, the latter is what had been decided by the authorities). Of course, a watershed provides many other services, whose value is often not measurable. So we would have an underestimate of the value. Then there are objects of Nature that communities would rightly refuse to value (sacred groves). That measurability is perforce incomplete is no argument for not trying measurement. Economists call the assets that comprise Nature, “natural capital.”

What about a person's knowledge, skills, reputation, and state of health? They are non-alienable (knowledge, skills, and reputation are also intangible). Nevertheless contemporary economists include them on the list of capital goods by calling them “human capital.” We economists use market wages and salaries to estimate the value of human capital to the individual possessing it. The term "human capital" reminds us that assets can be ends, they can be means to ends, or they can be both. Reading is a pleasurable activity, but it is also necessary in a job that requires literacy. Similarly, a person's health is both a desired end for him and a means to employment. These examples suggest that valuing human capital on the basis of their market prices is to under-estimate their worth. Working with biassed estimates can nevertheless be revealing. Recent work on human capital using market prices has shown that as a share of national wealth it is larger than other capital goods by at least two orders of magnitude (Arrow et al., 2012).

Once you include knowledge, skills, reputation, and health in the category of capital goods, it's hard to know where to stop. Should you not include institutions, such as the State or the market system? **They too are assets**. There is then the temptation to go for broke and speak of institutional capital, knowledge capital (science and technology), and cultural capital. Today some people even refer to religious capital. You could then take an extreme position by calling *all* durable goods capital goods. I first discuss the advantages in doing that. I then show that there are severe disadvantages.[11]

In a ground breaking work, Putnam (1993) studied the role social networks play in facilitating civic engagement. The underlying thesis was that network activities help to create trust among members (by, among other ways, enabling members to learn who are trustworthy), which in turn helps people to engage in civic activities. Using Italy as his laboratory, Putnam uncovered contemporary data on memberships of choral societies and football clubs in each of the 20 states in the country. Calling the trust that is created in such networks "social capital," he found that the networks not only discipline the nations's state governments in their role as suppliers of public services, but that they also have a long temporal reach. Regions where civic engagement was greater hundreds of years ago enjoy greater levels of civic engagement and better governance today. Putnam identified civil society

11 There is a huge literature in which cross-country data are used to show that various types of capital goods are factors in economic performance, usually growth in GDP. See for example Jones and Romer (2010), who find that the quality of institutions matters for economic performance.

as the seat of social capital. His finding says that while civic engagement contributes directly to one's well-being, they are also of instrumental value: accumulation of one type of asset (social capital) improves the quality of another type (state government).[12]

As these examples show, there are many ways of classifying assets. However, only one has proved useful in empirical work. It is one thing to recognise that a durable object has worth, it is another thing to measure that worth. Try for example to compare the value to a nation of good governance with the real estate value of its capital city. The requirement that a durable good be measurable if it is to be called a "capital good" was at the basis of the complaints Arrow (2000) and Solow (2000) made of attempts to regard social capital on par with buildings, roads, labour, and land. We take heed of their caution in our study, by creating a three-way partition of capital goods: *produced capital* (buildings, roads, ports, machines, instruments), *human capital* (population size, health, education, reputation, knowledge and skills), and *natural capital* (ecosystems, sub-soil resources). The remaining set of assets (institutions and practices, more generally, social capital; publicly available knowledge) we will call *enabling assets*, because they confer value to the three classes of capital goods by facilitating their use. Moreover, knowledge in the sciences, technologies, and the arts and humanities – they are all enabling assets - are created and acquired, but they are created and acquired by people (human capital) in combination with produced capital (books; laboratories and equipment) and natural capital (raw material). We study the real economy here. Financial capital facilitates exchange (among people and across time), and so in our reckoning it too is an enabling asset. In our classification, there are three categories of capital goods and a wide range of enabling assets.[13]

3.2 Inclusive Wealth and Social Well-Being

At this point we do not specify the economy under study. The economy could be a person, a household, **village,** town, district, nation, the whole world. In Part II we will apply the account of

---

12 Networks can have negative social value, for example criminal gangs. Dasgupta (2000) studied the formal economics of social capital by introducing it in models of economic growth and development. Grootaert and von Bastelaer (2002) is a collection of empirical papers on the measurement of social capital. Ghate, Jodha, and Mukhopadhyay (2008) contains applied studies on the strengths and weaknesses of social capital as a lubricant of human interchanges. The authors do that by studying the management of local common property resources in South Asia.

13 Here I am following the classification in Dasgupta and Mäler (2000). Of the three categories of capital goods, produced capital has the oldest pedigree, being the substance of Classical Political Economy of the 18th and 19th Centuries. (Land of various qualities served as an additional, indestructible factor of production.) The paper that placed human capital in the category of capital goods and created the modern literature on human capital is Schultz (1961). A huge contemporary literature in ecology and environmental economics has introduced natural capital in economic reasoning.

population ethics we develop in Part I to the world as a whole. Here we seek to find a way to aggregate the three classes of capital goods in the economy, for without aggregation we would be left with a catalogue of capital goods, and no idea how they relate to one another in people's activities and engagements.

In order to aggregate disparate objects, we need a common currency. The currency we use in this essay is well-being, and by the value of a capital good we mean its social worth (Section 2). In the body of the Arrow Lecture we appeal **to Well-Being Consequentialism for defining** the social worth of all goods and services, not only capital goods. But the propositions I discuss here have a far wider reach. They apply to any conception of social well-being which recognises that we are dependent on assets (be they capital goods or enabling assets) for anything we do and any kind of person we strive to be.

It transpires that the right way to aggregate the social worth of our three classes of capital goods is to add them. That's the social worth of the totality of capital goods to which the economy has access at any given date. By social worth we mean not only **the** worth to people **who are** alive at that date, but also to future people. We call the social worth of the three classes of capital goods *inclusive wealth*. The qualifier signals that the notion of wealth adopted here differs from the one in common use in two ways: (i) the social worth of capital goods - they are called *accounting prices* - are not necessarily market prices; (ii) in addition to produced capital, wealth includes human capital and natural capital. The inclusion of human capital says among other things that personal characteristics of the individual with access to a piece of capital good matter. A piano in the possession of someone who can play the piano has a higher accounting price than it would have were it to be in the possession of someone innocent of music, other things equal of course.

A person's well-being is shaped by the extent to which her projects and purposes are realized. They in turn are rooted in her engagements, both with her own self and with others (we develop the idea further in Section 4), which means accounting prices of goods and services are person-specific. It also means that the degree of fairness in the distribution of well-being influences accounting prices. So as to avoid repetition, we will sometimes drop the qualifier "inclusive" from inclusive wealth.

An economy's institutions and practices endow capital goods with their social worth, which is why we are calling them enabling assets. The same portfolio of capital goods would have greater worth to people if the society in question were to bring about changes to its institutions and practices that, for example, create greater trust among people. Those changes would express themselves through an altered set of accounting prices. A writing desk has a higher accounting price in someone's study than in a war zone. An economy can become wealthier simply by improving the quality of its enabling assets. A more familiar means of wealth creation is to ensure that (net) investment, aggregated across the three classes of capital, is positive. These claims are proved formally in Appendix 4.

Why should we be interested in inclusive wealth? The reason is this:

Call the numerical measure of well-being across the generations, *social well-being* (Sects. 8-9). The accounting price of a capital good is the extent to which social well-being would increase if the economy were provided with an additional unit of it. The idea of accounting prices generalises quite naturally to all goods and services. We have now nearly reached the point to which we have been heading. If assets in their totality are the basis of social well-being, and if accounting prices of capital goods measure the (marginal) contribution they make to social well-being, there must be an intimate connection between inclusive wealth and social well-being. That is indeed so. It can be shown that social well-being increases over time if the corresponding measure of inclusive wealth increases over time, and declines over time if inclusive wealth declines over time.

The equivalence of inclusive wealth and social well-being holds also in policy evaluation. At any given date, a change in policy that increases (resp. decreases) social well-being also raises (resp. lowers) inclusive wealth. Of particular interest are investment projects, which directly reallocate capital goods from their use under the status quo to their use in the project. Acceptance of a project amounts to a change in policy, marginally of course if the project happens to be small relative to the size of the economy.

Inclusive wealth and social well-being are thus two sides of the same coin whether we ask if society has progressed over a period of time (the question of interest in sustainability analysis) or whether we evaluate alternative policies at a moment in time (that's where social cost-benefit analysis applies). We will call this the *Wealth/Well-Being Equivalence Theorem*. The theorem is stated formally and proved in Appendix 5. The theorem implies that inclusive wealth is the right measure of social well-being, not gross domestic product (GDP) nor any of the other measures that have been suggested in recent years, such as the United Nations' Human Development Index.[14]

---

14 The UN's Human Development Index (UNDP, 1990) is a linear aggregate of GDP per capita, life expectancy at birth (human capital), and literacy (human capital). The equivalence of movements in social well-being and movements in (inclusive) wealth in a general setting was proved by Dasgupta and Mäler (2000) and Arrow, Dasgupta, and Mäler (2003a-b). Dasgupta (2004) is a book-length treatment of the idea of sustainable development and the role that the wealth/well-being equivalence theorem plays in it. Arrow et al. (2004) study the connections between sustainable and optimum development. In order to facilitate empirical work, the authors of these works identified circumstances in which average well-being across the generations moves in the same way as per capita wealth. Dasgupta (2014) and Irwin, Gopalakrishnan, and Randall (2016) are non-technical accounts of the wealth/well-being equivalence theorem and its extensions. In an interesting and important paper Yamaguchi (2018) unearths the differences between conceptualizing social well-being in its total and its population-average forms. Arrow et al. (2012, 2013) applied the theorem by estimating movements in wealth in five countries (Brazil, China, India, USA, and Venezuela) over the period 1995-2000. Yamaguchi, Sato, and Ueta (2016) applied the theorem to a region in Japan. UNU-IHDP/UNEP (2012, 2014) and Managi and Kumar (2018) contain pioneering estimates of the wealth of more than 120 countries, akin to estimates the World Bank provides annually of the GDP of nations. Tomlinson (2018) has applied the theorem to the recent economic history of Nigeria. The theorem has also been used to motivate the

The wealth/well-being equivalence theorem directs us to expand our common-sense notion of investment. To invest in a capital good is to increase it beyond what it would be if there were to be no investment in it. We are talking of *net* investment here, that is, investment net of depreciation. And because we are interested in inclusive wealth, we may call net investment "inclusive investment" (Hamilton and Clemens, 1999, called it "genuine saving"). Our common-sense notion of investment, however, carries with it a sense of robust activism. When the government invests in roads, the picture drawn is of bulldozers levelling the ground and tarmac being laid **by men in hard hats**. But the notion of capital goods we are using here extends beyond produced capital to include human capital and natural capital. That training people to be teachers is investing in human capital is simple enough. To leave a forest unmolested may not suggest much like investment, but it is an investment. It enables the forest to grow. To allow a fishery to re-stock under natural conditions is to invest in the fishery; and so on.

That suggests inclusive investment amounts to deferred consumption, but the matter is subtler. Providing additional food to undernourished people via, say, food guarantee schemes not only increases their current well-being, it enables them also to be more productive in the future and to live longer. Because their human capital increases, the additional food intake should count also as investment. Note though that food intake by the well-nourished doesn't alter their nutritional status, which means the intake is consumption, not investment. Our equivalence theorem suggests that by "net investment" in an asset we should mean the value of the change in its stock.

We are talking of ends and means here. Despite the equivalence theorem, ends (enhancing social well-being) are the right starting place for population axiology, as is the case in Sidgwick's Utilitarianism. Ends (human flourishing) are also the starting point of Rawls' contractual theory of justice. His theory is emphatic in regarding persons inviolable in certain respects of their personhood and insists persons are to be treated equally (Appendix 3). Ends are the right starting point because they are antecedent to means. One can articulate ends even without asking whether they can be realized, but it makes no sense to talk of means if the ends they are meant to advance aren't articulated first. The wealth/well-being equivalence theorem doesn't deny the antecedence of ends; what the theorem says is that if the means to a set of ends have been identified, it doesn't in principle make any difference whether we examine the extent to which the ends have been (or are likely to be) furthered by a change to an economy or whether we estimate the degree to which the means to those ends have been (or are likely to be) bolstered by that change: the two point in the same direction. The wealth/well-being equivalence theorem draws attention to the fact that no matter what conception of ends citizens may adopt, the source of the means to those ends lies in a society's capital goods. Their accounting prices serve to tie them to the ends. The theorem says the weapons and human capital deployed by the state to abduct awkward citizens from their homes in the middle of the night are to be

---

development of methods for estimating of accounting prices (Appendix 5).

awarded very large, possibly unboundedly large, negative accounting prices. Likewise, steel put to use in making ploughs differs from steel used to manufacture guns. Accounting prices of capital goods depend on their location and the use to which they are put. The equivalence theorem is utterly wide in its reach.

The equivalence between inclusive wealth and social well-being holds as tightly in a society where the ends are far from being met owing to mis-allocation of the means or unjustified usurpation of the means by the powerful, as it would in a society where they are met as far as is possible under the prevailing scarcities of the means. The equivalence of ends and means will be confirmed repeatedly in this essay.

That inclusive wealth is equivalent to social well-being is not an empirical law, it is an analytical proposition. Being an equivalence relationship, it does not say whether a society is doing well or badly, whether it is well governed or badly governed. But both theory and experience say that it is commonly easier to measure the means to the ends than it is to measure the ends themselves, which is why in empirical work we are drawn to the means. Consider for example a proposal to build a social housing project on an unused piece of wetland. The project's feasibility report contains estimates, expressed in dollars perhaps, of the investment involved (draining the land, constructing the buildings and ancillary structures). The report estimates that construction requires so many labour hours in each year of the investment phase, so many types and quantities of machines and equipment and intermediate goods, and so on. The report also estimates the expenditure incurred by various parties and the transfer of resources that the project envisages among them during the project's life (government, tax payers, people who will occupy the homes). If the public are environmentally minded, they will insist the report also itemises the services the wetland currently provides to the local community (e.g. filtering water) and a description of the animal and bird populations that make it their habitat. Their disappearance will be seen as a loss. All that information, and it is usually very detailed, are in the form of flows of goods and services. The problem is to evaluate the project.

It won't do to simply ask whether the project will enhance social well-being, for that would be to re-ask whether the project should be accepted or rejected. The idea is to put the project data to use by applying accounting prices to the items. Decades ago welfare economists proved that the project can be evaluated by estimating the present-value of the flow of social profits (net social benefits) that arise from it. By "present-value" we mean a weighted sum of the flow of the project's benefits and **costs** over time. The weights are called "social discount factors," which transport benefits and **costs** in the future to the present (**Section 9.1;** Appendix 5). It can be shown that the project should be accepted if its present-value is estimated to be positive (unless, that is, a close variant of the project with an even higher present-value is identified), but rejected if it is found to be negative. That's known as the "social PV-criterion." The word "social" is crucial here, for it means that accounting prices are to be used to value goods and services, not market prices. The criterion also tells us that the accounting price of a capital good is the present-value of the flow of social benefits that would be enjoyed if the economy were to be provided with an additional unit of the good.[15]

The present-value of a project's flow of social profits (dollars per year) has the dimensions of stock (dollars, period). You will have by now guessed that a project's "present-value" is none other than the change to inclusive wealth brought about by the redeployment of capital goods from elsewhere to the project. It thus transpires that the criterion for project evaluation that was developed by welfare economists decades ago is implied by the wealth/well-being equivalence theorem. Notice though that the social PV-criterion for investment choice is a far cry from the demand commonly made by economists and political commentators that economic policies should be so chosen as to enhance GDP. A macroeconomics that sees growth and distribution of GDP as the aim of public policy is inconsistent with basic welfare economics. I have no explanation for how and why that incongruity has been allowed to fester in economics.

There is a subtler reason for evaluating policies and assessing the progress or regress of economies in terms of the means to our ends. It is a mistake to suppose we come armed with our ends. For the most part they are inchoate in our minds. Studying means clarifies the tradeoffs involved in promoting our ends. The interplay of the "is" and the "ought", to use that well-worn distinction, helps us to better understand our ends. In Part II we conduct sensitivity analysis so as to take advantage of that interplay.

3.3 Placing a Value on Opportunity Sets

The wealth/well-being equivalence theorem also puts into perspective controversies over the objects of interest in distributive justice, for example whether they should be personal well-beings or resources or Rawlsian primary goods or opportunities or human capabilities.[16] The equivalence theorem says that the choice is a matter of convenience and context. We are talking of convenience in practical work, not the relative status of the two notions. Well-being is a measure of human flourishing. It serves as the primary notion in the ethical reasoning we pursue here.[17]

In contrast consider "human capabilities," which have been defined as the alternative combinations of functionings that are feasible for a person to achieve (Sen, 1984, 1992, 1999, 2009). Functionings, earlier we called them "engagements," are in turn the various doings and beings a person rationally values, and are frequently interpreted by Capabilitarians as different kinds of life. A

---

15 Prest and Turvey (1965) is an early survey of the then existing literature on social **cost-benefit** analysis. Little and Mirrlees (1968, 1974), Arrow and Kurz (1970), and Dasgupta, Marglin, and Sen (1972) are book-length treatments of the theory and of ways to make use of the social PV-criterion in project evaluation. Accounting prices of environmental resources are especially hard to estimate, but there is now a sophisticated literature establishing techniques for estimating them. See Freeman (2002) and Haque, Murty, and Shyamsundar (2011).

16 Rawls (1972), Dworkin (1981a-b), Sen (1985, 1992, 1999, 2009), Cohen (1989), and Barry (1990), among many others.

17 In Section 13 we study the place of the value of Nature, qua Nature, in our sensibilities.

person's capability represents the effective freedom she enjoys in selecting from different functionings, meaning that capabilities are *sets* of functionings. Despite the theory's acknowledgement that functionings are objects that people rationally value, the distinguishing feature of Capabilitarianism is its view that the value of a capability to someone is not derived from the worth to her of the functionings that are included in the set (otherwise Capabilitarianism would be simply a version of Utilitarianism); instead, the set is valued directly. (Sen, 1999, contains a clear statement of this line of reasoning.) There would seem to be a reason behind the move. The theory has been proposed as a rival to Utilitarianism and to the reasoning Rawls (1972) deployed for identifying citizens' choices behind the veil of ignorance.

Capabilities are examples of what economists call "opportunity sets".[18] Imagine that a person is able to place a value, measured in terms of her well-being, on every element in an opportunity set presented to her. The value she ascribes to the opportunity set would then be the worth to her of the element in it she values most highly. It follows that if the person were offered a choice between alternative opportunity sets, she would be able to rank them and identify the one most valuable to her.

Capabilitarianism doesn't subscribe to that way of reasoning. It values capabilities directly, which is why the theory faces an insuperable problem: it admits no machinery for comparing capabilities if none is a subset of any of the others. The theory yields partial orderings of the objects of choice, but is insistent in not offering complete orderings.[19] The theory is certainly able to say that slavery is bad, that health is a basic need, that in the modern world education is necessary for a flourishing life, and that freedom of speech is a cherished goal; but it is unable to say much more. And those particular value judgments are reached by all moral theories it has been designed to contend with.

Why should someone wish to value opportunity sets if she knows the value of their elements to begin with? Why doesn't she simply choose the best element from the opportunity sets available to her? The reason is that the person typically has insufficient knowledge about herself and the world she lives in, but knows she will learn more about both with the passage of time. Which is why she is aware she would be able to make more satisfactory choices if she waited rather than tie her hands by choosing an ill-informed best element from an opportunity set now. No doubt waiting has a cost, but if the cost is smaller than the gain from keeping her options open for a while, she would rationally choose to keep her options open. That's why there is a case for selecting an opportunity set now and waiting to choose an element from it when she can better identify the element she rationally desires

18 In economics a familiar example of an opportunity set is the set of commodity bundles an individual can afford to purchase in a market economy.

19 Let R be a binary relation of elements of a set X. We say R is a *complete ordering*, or simply an *ordering*, if it is reflexive (for all x in X, xRx), transitive (for all x, y, z in X, xRy and yRz implies xRz), and complete (for all x, y in X, either xRy or yRx). The binary relation R is said to be a *partial ordering* is it is reflexive and transitive, but not complete.

most. This involves backward induction, a style of reasoning I deploy when developing population ethics.

Consider that the acquisition of skills (an aspect of "human capital") involves the use of resources, which means there are tradeoffs among them. But not all skills are of equal value to all, nor even to the same person. Numeracy and literacy are basic skills in the modern world, they prove vital to people no matter what they wish to be and do and no matter what circumstances they face. Investing in education in the early years of one's life (the decision is of course made on the child's behalf) is a way of keeping her options open on the choice of further education and the profession she pursues when she has become an adult. The child's future options are the elements of the opportunity set her parents are investing in on her behalf today.

A theory that values the capacity to form life plans but doesn't relate that value to the realization of those plans and the experiential states that go with them as and when they are carried out throws away material of ethical substance. Capabilitarianism does that in abundance and does it as a matter of principle. The theory says for example that the reason headaches are bad is that they limit one’s ability to function, it appears to be uninterested in the fact that headaches are painful. Capabilitarians could no doubt retort that people have good reasons to abhor headaches, but Capabilitarianism doesn’t admit information on a person's ranking of states of affairs she wants rationally to avoid. How is the Capabilitarian to know, for example, how headaches are to be compared to toothaches? Patrick Suppes' exploration of the normative force of freedom (Suppes, 1987) showed that freedom to choose can't carry ethical weight unless there is an independent machinery for valuing the objects of choice.

The line separating an instrumental from an intrinsic value can be wafer thin when the instrument advances a value that is deeply held. What is taken to be an intrinsic value could well be an instrument for advancing a more deeply held value. Nevertheless, we should suppose, as many do, that freedom has an intrinsic value, and that, the very sense of a flourishing life includes freedom. An engagement that is freely chosen would then contribute more to a person’s well-being than that same engagement would were it to be thrust on her. The characteristics of the engagement under the two circumstances would thus differ. Moreover, the value of an engagement to someone would depend on what other engagements were available to her. These additional considerations can be readily embedded in the account of capabilities we have reviewed here (see Sher, 2018).

Arrow (1995) appealed to the instrumental value of freedom to show that Capabilitarianism is an expression of Utilitarian reasoning. I have focussed on that line of reasoning here because to add an intrinsic value to freedom would not change the character of the argument. To choose a functioning before the world reveals itself any further is to lock oneself into a situation from which there is no recourse if the choice proves to be incongruent with one's more informed values. To choose instead a capability (it's a set of functionings, remember) is to keep one's options open until things become clearer. But in order to know which capability is the best against an uncertain world, it is necessary to value, howsoever imperfectly, the consequences to the person of her choices in alternative, uncertain states of affairs. Functionings themselves have to be valued. The abiding attraction of Utilitarianism is that it offers a way to do that. The theory does that by using the person's well-being as currency.

Arrow's analysis showed that choosing capabilities is akin to purchasing options in the market for securities. Such choices reflect a desire for flexibility against future uncertainty. The argument is formalised in Appendix 6.[20] In Appendix 3 this line of reasoning is extended to show that Rawls (1972) was entirely right to base his theory of justice on "primary goods." In Rawls' account primary goods are the objects citizens behind a veil of ignorance know they will need no matter what they will subsequently discover about themselves, including their own good, and their social world. More particularly, he saw primary goods as necessary for self-respect, by which he of course meant *informed* self-respect. That Rawls' focus in his great work was on primary goods should not detract from the fact that human flourishing is at the heart of his theory of justice. I also argue in Appendix 3 that Rawlsian primary goods are the ingredients of inclusive wealth.[21]

4 Synopsis

"Classical Well-Beingism" reads distinctly odd. So we will call the theory that evaluates states of affairs in terms of the sum of individual well-beings, Total Utilitarianism. In part I we study the foundations of population ethics. Sections 5-6 explore how far Total Utilitarianism is able to guide population ethics before running into trouble. In doing that I respond to several strands of criticisms that have been levelled against the theory. To uncover what Total Utilitarianism is able to deliver, I put it through its paces in a timeless world endowed with a stock of natural capital. In the formal models we work with below, I shall be thinking of the world as a whole. Natural capital is therefore to be interpreted as the biosphere.

4.1 The Genesis Problem and Actual Problems

Total Utilitarianism is least open to problems **of interpretation** if it is applied to a world devoid of people. In Section 7 I argue why. Nevertheless, in Section 5 I assume there is a Decision Maker (DM). DM knows that if humans work on the biosphere, they can produce the consumption good. DM also knows that the larger is the number of people created, the greater would be the output, albeit at a rate that diminishes with increasing numbers. The model is presented in Sections 5.1-5.2. We should think of the model as responding to the Genesis Problem.

Choices studied in population ethics differ from those in (standard) decision theory. A central idea in population ethics is of a life that goes neither well nor not-well (Sect. 5.1). Life at that border

---

20 Arrow and Fisher (1974) and Henry (1974) applied the idea of option values to provide a new argument in support of conservation science. Arrow (personal communication) attributed the idea to Weisbrod (1964).

21 Rawls **observed** that **a** way to identify one's good **is to acknowledge** the Aristotelian Principle (pp. 424-433): **"... (O)ther things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and the enjoyment increases the more the capacity is realized, or the greater its complexity.**"

is the point of reference against which DM deliberates additional births. Sidgwick (1907: 124) spoke of "neutral feeling" (he also called it the "hedonistic zero") as the point from which positive, or for that matter negative, happiness is measured. The notion can be extended to a person's life; and Sidgwick did that. We may then speak of a "neutral life" as a life in which happiness algebraically aggregates from birth to death to hedonistic zero. The interpretation of a neutral life I am drawn to, however, differs from the one Sidgwick offered (Sects. 5.4 and 6).

Decision theory studies choices that don't affect population numbers. Numbers being the same no matter which policy is chosen, comparisons of personal well-beings are the only things that have mattered to decision theorists. Decision theory does not recognise the idea of a neutral life because it is not needed. In population ethics it is an essential ingredient.[22]

The notion of a neutral life can be transferred to lives evaluated in terms of consumption activities. Meade (1955) called the living standard in a neutral life, "welfare subsistence." It is natural to refer to it here as "well-being subsistence." I assume DM knows that a person's well-being is a function of her consumption level, and that marginal well-being is a declining function of consumption.

DM is a Total Utilitarian, and knows in advance that humans will be identical to one another in every respect (their productivity, their ability to convert consumption into well-being, and so on). The optimum distribution of consumption is therefore an equal distribution of total output. The problem before DM is to determine the optimum tradeoff between population size and the representative person's consumption level. Its solution is Total Utilitarianism's answer to the Genesis Problem, the subject of Section 5.

In the model I deploy to test Total Utilitarianism, DM commends a large population. The optimum living standard is shown to be proportionately not much higher than well-being subsistence. For a central class of parameter values the ratio is shown to be less than *e* (the natural base of logarithms). We study the sensitivity of optimum population numbers to alternative values of the parameters defining personal well-being functions and production possibilities (Sect. 5.5). In a series of publications Parfit (1976, 1982, 1984, 2016) and others following him have faulted Total Utilitarianism for the tradeoff it commends between population size and average well-being. The authors have also suggested modifications to the theory so as to avoid what they see are paradoxes arising in the theory. Parfitian modifications are studied in Sections 5.6 and 5.8.

I have found no suggestion in Sidgwick (1907) that the idea of a neutral life is independent of a person's social environment. In contemporary writings on population axiology the question of whether it is or it is not is ignored. But if the neutral life is independent of the social environment, so

22 Formally, decision theory requires only that personal well-beings are uniquely measurable up to positive affine transformations. Luce and Raiffa (1957) and Raiffa (1968) are classic expositions. In contrast, we will need to demand of personal well-beings that they are uniquely measurable up to positive linear transformations (Sect. 5.3).

is well-being subsistence. That assumption, in wide use amongst economists, is hard to square with a commonplace human experience. Whether a person sees his life as going well depends in part on his expectations and aspirations; those in turn are (again, in part) dependent on what others in his society aspire to and are able to achieve. I take a minimalist approach to this range of issues and sketch a world in Section 5.7 and Appendix 1 where personal well-being is socially embedded. Well-being subsistence in the model increases with the average living standard in society. I study ways in which population ethics is affected by that feature of the human experience.[23]

In Section 6 we find that Sidgwick (1907) erred in his reading of a "neutral life." The error can mislead one into thinking that when someone says Total Utilitarianism advocates large populations (as I have just done), they must also mean that the theory can commend states of affair where people have lives that are barely worth living. To make that inference is wrong.

In Section 7 a modified version of the plight of Sleeping Beauty is used to show that Total Utilitarianism's weaknesses don't lie where contemporary population axiologists claim they do, but elsewhere. The problem with the theory is that it speaks to Genesis, not to actual choices made in an inhabited world. One may even say that potential parents face a far more difficult set of choices than DM. DM can afford the luxury of conflating the good with what one should do, potential parents don't enjoy that privilege. I explore a way to amend Total Utilitarianism for what should be called Actual Problems, as opposed to the Genesis Problem. The amendment I work with involves in an otherwise Utilitarian reasoning the adoption of an attenuated form of agent-centred prerogatives over choices open to people. The influence of Thomas Nagel and Bernard Williams will be transparent, although both authors have shown in their writings a far more critical attitude toward Utilitarian-Consequentialism than I do here. But the modern philosophical source that comes closest to what I am after is Scheffler (1982). Potential parents evaluate states of affairs on the basis of a weighted sum of personal well-beings, where the weight they place on potential well-beings of children they could have is less than the weight they place on their own well-being and on the well-being of children they already have, knowing in advance that they will want to share resources with the children they produce on an equal basis with themselves. The amended theory is Generation-Relative Utilitarianism. The backward induction involved in this reasoning is essential to the theory. The gap between ex ante and ex post reasoning distinguishes Generation-Relative Utilitarianism from Total Utilitarianism; the latter admits no gap.

Generation-Relative Utilitarianism invokes a weak form of agent-centred prerogatives. It curbs births, it doesn't sanction applying a lower weight on others' well-being even on grounds of prerogatives. But depending on the relative weights deployed by decision-makers, the gap between ex ante and ex post reasoning can have huge implications for optimum population size.

23 Williams (1976) and Nagel (1976) have shown that whether a life is exemplary depends on what life throws at us. I abstract from this in the text.

In Section 8 Generation-Relative Utilitarianism is put to work in a timeless world where present people deliberate how many further people to create. As in the model studied in Section 5, the world in Section 8 is endowed with a capital asset, the biosphere, that can be used to produce a consumption good. Generation-Relative Utilitarianism is shown to recommend a smaller population and a higher living standard than Total Utilitarianism. Using stylized values of ethical parameters we find that Generation-Relative Utilitarianism can recommend high living standards and correspondingly small populations.

Parfit (1976) and following him Broome (2004: 157-162) have argued that Generation-Relative ethics is incoherent because it doesn't yield a binary relation among states of affair. In Section 8.3 it is shown that the theory only appears to be incoherent, and that it does so because Parfit insisted that decision-makers view states of affair from nowhere. That's a natural requirement to make of DM, but not of we mortals. Generation-Relative Utilitarianism evaluates states of affairs from inside the states of affairs themselves.

It may seem odd that a model that reflects a timeless economy could be of use in applying population ethics to an economy that moves through time. In fact a timeless economy is near-identical to an economy that moves through time over an indefinite future but in which economic variables (e.g. population and consumption) are constant. The latter is known as a "stationary economy." The connection is demonstrated in Section 8. When we come to study the stationary economy, the asset people are assumed to work with is taken to be an aggregate of all capital goods, including the biosphere.

Discounting for time and the generations in ethical deliberations over saving and fertility is not the same as exercising agent-centred prerogatives. In Section 9.1 I show why. I also argue there that a world where investment, when chosen wisely, has a positive return directs us on grounds of fairness to discount the well-beings of future generations.[24]

So as not to contaminate agent-centred prerogatives with time discounting, I eschew the latter in Section 9.3 and thereafter. It's a modelling strategy designed to develop Generation-Relative Utilitarianism in a minimalist setting. As in the timeless economy of Section 8, I assume that personal well-being is a function of personal consumption: people are egoists. In each period adults apply labour to capital goods to produce and then share with their children the all-purpose consumption good. We find that the recommendations of Generation-Relative Utilitarianism can be carried out only if there is an implicit understanding among the generations that each generation will choose the size of the next generation and act as a trustee of the biosphere it has inherited. Being trustees, the generations protect the biosphere from excessive use; not to mention, from unacceptable damage.

24 Most writings on the ethical basis of intergenerational saving have been built on Utilitarian thinking. But in his contractual theory of justice John Rawls also sketched a principle of saving (Rawls, 1972: ch. 5, sects. 44-45). We study the principle of just saving in *A Theory of Justice* in Appendix 3.

4.2 The Biosphere as a Commodity

Part II moves away from that scenario. Section 10 summarises studies that have recorded substantial declines in the biosphere's productivity over the past decades. Those declines can be traced to the environmental and reproductive externalities people inflict on one another. Today, growth in atmospheric carbon concentration is the canonical expression of adverse externalities, but humanity faces wider and deeper threats to our future from the biological extinctions now taking place **(see e.g., Cardinale et al., 2012)**, which are also morally more reprehensible. Proximate causes of extinctions **include global climate change; but they also include** destruction and fragmentation of natural habitats and over-exploitation of biological communities residing there. We are converting land into farms and plantations, destroying forests for timber and minerals, applying pesticides and fertilizers so as to intensify agriculture, introducing foreign species into native habitats, and using the biosphere as a sink for our waste. And they are taking place at scales that are orders of magnitude greater than they were even 250 years ago.

Adverse externalities arising from our use of the biosphere in great measure arise because Nature is mobile: birds and insects fly, water flows, the wind blows, and the oceans circulate. That makes it hard to establish property rights to key components of the biosphere. By property rights I don't only mean private rights, I include communitarian and public rights. Which is why much of the biosphere is an "open-access resource," meaning that it is free to all to do as we like with it. Hardin (1968) famously spoke of the fate of unmanaged common property resources as "the tragedy of the commons." But while Hardin's analysis was entirely appropriate for global commons (the atmosphere, the oceans), it was less than applicable to geographically confined resources such as woodlands, ponds, grazing fields, coastal fisheries, wetlands, and mangroves. Because local commons are geographically confined, their use can be monitored by community members. There were exceptions of course, but in times past those resources were managed by communities, they were not open-access resources. Reviewing an extensive literature, Feeny et al. (1990) observed that community management systems enabled societies to avoid experiencing the tragedy of the commons. Social norms of behaviour, including the use of fines and social sanctions for misbehaviour have guided the use of local common property resources.[25]

In poor countries the commons continue to supply household needs to rural people (water, fuelwood, medicinal herbs, fruits and berries, manure, and fibres and timber for building material). Some products are also marketed (fish, fuelwood, dung, wood and fibre products). But as in so many other spheres of social life, communitarian practices have over the years strengthened in some instances (e.g. community forestry in Nepal) and weakened in others. They weakened for example when communal rights were overturned by central fiat. In order to establish political authority after

25 The literature on this is extensive. See Ostrom (1990), Marothia (2002), Ostrom et al. (2002), and Ostrom and Ahn (2003).

independence (and also earn rents from timber exports), a number of states in sub-Saharan Africa and Asia imposed rules that destroyed community practices in forestry. Villages ceased to have the authority to enforce sanctions on those who broke norms of behaviour. But knowledge of local ecology is held by those who work on the commons, not by state officials, who in addition can be corrupt. Thomson et al. (1986), Somanathan (1991), and Baland and Platteau (1996), among others, have identified ways in which state authority damaged local institutions and turned local commons into seemingly open-access resources. Then there are subtle ways in which even well-intentioned state policy can cause communitarian practices to weaken (Balasubramanian, 2008; Mukhopadhyay, 2008).[26]

4.3 Common-Property Resources and Fertility Intentions

Even when a common is managed by the community and outsiders are kept at bay we should ask whether access to it is based on household size or whether each household has a fixed share of its output. In Appendix 2 it is shown that when larger households are entitled to a greater share of the commons' goods and services, households have an incentive to convert natural resources excessively into private assets, **which includes** household size. In sub-Saharan Africa larger households are (or until recently, *were*) awarded a greater quantity of land by the kinship group. That practice encourages fertility. What is true in the case of local commons to which households have access regardless of their size holds true in the case of global commons, to which we all have access regardless of our household size. Even humane systems of property rights can give rise to adverse externalities.

How important are local commons in household income? Despite the importance of the question there is little in the form of quantitative evidence. Casual empiricism suggests they are less significant in advanced industrial countries than in poor rural societies. In the former, local resources are either owned privately or under the jurisdiction of local authorities or, as in the case of places of especial aesthetic value, national parks. That is not so in rural areas in poor countries. In a pioneering **study**, Jodha (1986) reported evidence from semi-arid rural districts in Central India that among poor families the proportion of income based directly on local commons was 15-25%. Cavendish (2000) arrived at even higher estimates from a study of villages in Zimbabwe: the proportion of income based directly on local common property resources was found to be 35%, the figure for the poorest quintile being 40%. To not recognize the significance of the local natural-resource base in poor

26 In recent years democratic movements among stakeholders and pressure from international organisations have encouraged a return to community based systems of management of the local commons. Shyamsundar (2008) is a synthesis of the findings in nearly 200 articles on the efficacy of a devolution of management responsibilities - from the state to local communities - over the local natural-resource base. Her article focuses on wildlife, forestry, and irrigation. The balance of evidence appears to be that devolution leads to better resource management, other things equal. Shyamsundar of course offers a discussion of what those other things are.

countries is to not understand how the poor live.[27]

4.4 Our Impact on the Biosphere

Being a measure of the social worth of an economy's capital goods, inclusive wealth is a stock. In contrast GDP, which the market value of the final goods and services an economy produces, is a flow variable. The rogue word in GDP (gross domestic product) is "gross", as the index doesn't include the depreciation of capital. Which is why it is possible for GDP to grow for a period while inclusive wealth declines. Even if produced capital and human capital were to grow in magnitude, inclusive wealth would decline if natural capital were to decline in quality, or quantity, at a high enough rate. But if wealth were to continue to decline, GDP would eventually have to decline. You cannot degrade the biosphere indefinitely and expect living standards to rise continually.

Section 10 contains evidence that humanity's enormous economic success in raising GDP and the many benefits that have come with it in recent decades has involved an unsustainable conversion of natural capital (the biosphere) into produced capital and human capital. Population ethics directs us to look not only at the state of the world as it is today, at what has been achieved, it also requires of us to peer into what lies ahead. There are intellectuals who insist we are living in the best of times (Pinker, 2018), but in Part II we find evidence that we may well be living simultaneously in the worst of times.

Ehrlich and Holdren (1971) introduced the metaphor, I=PAT, to draw attention to the significance of the biosphere's carrying capacity for population ethics. The authors traced the *i*mpact of human activities on the Earth system to *p*opulation, *a*ffluence (read, the standard of living), and the character of *t*echnology in use (including knowledge, institutions, social capital). We can imagine that our impact on the biosphere is proportional to the demands we make of it. The demands we make of it in turn increase with rising economic activity, as measured by global output of goods and services. That even today's poorest societies can be expected in time to make fertility transitions to population replacement levels (perhaps even to below replacement levels for a while) is no reason to think that humanity's demands for the biosphere's goods and services will cease to exceed its ability to supply them. That is why it is a mistake to ignore the Ehrlich-Holdren observation that the biosphere responds to the demands we make of it, not to changes in the demands we make of it (e.g. those that accompany declines in fertility rates), nor to changes in the rate of change in the demands we make of it (those that accompany declines in the rate of growth of the global population). A long run global population of 11.2 billion (the UN's projection of global population in 2100) is likely to make a vastly greater demand on the biosphere than a population of, say, 2.5 billion (global population in 1950).

27 I have previously tried to build an account of the lives of the rural poor in poor countries in a treatise (Dasgupta, 1993) and more recently in a brief introduction to economics (Dasgupta, 2007a). Jodha (2001) is an outstanding collection of essays by the author on coping mechanisms that have been adopted by the rural poor in fragile ecosystems of Central India.

The Ehrlich-Holdren study also tells us by population number we should mean "weighted population," each person being weighted by the impact he has on the biosphere.

In order to give precision to the idea of the biosphere's human carrying capacity, I identify it as the maximum population that can sustain itself at a standard of living equal to well-being subsistence (Sect. 5). That move will be rejected by commentators who question that it is meaningful even to talk of human carrying capacity, let alone estimate it. They say the constraints imposed by a finite biosphere can be overcome if humanity were to accumulate other forms of capital assets at sufficiently high rates. Dasgupta (1969: Sect. 3) applied Total Utilitarianism to a model of consumption and accumulation in an economy endowed with a fixed factor of production (land) to find that indefinite accumulation of produced capital, even if feasible, would not be desirable. It was shown that no matter what levels of capital stocks the economy may have inherited from the past, Total Utilitarianism would recommend a population and savings policy that, over time, takes the economy to a stationary state. The argument extends to a world guided by Generation-Relative Utilitarianism. That finding informs our exercises here.

The notion of the biosphere's human carrying capacity is developed in Section 10, where I use statistics on the state of the global environment to put flesh into population ethics. In Section 11 Generation-Relative Utilitarianism is applied to data on the global demand for the biosphere's services. I use that data to estimate the maximum demand we can make of today's biosphere on a sustainable basis. I then arrive at quantitative estimates of optimum population and the optimum living standard in a world constrained by *today's* state of the biosphere. This is an unsatisfactory research strategy, but I know of no study that seeks to determine sustainable supplies of ecological services were the biosphere allowed to improve in quality from what it is now. So I am obliged to take the current state as the basis of the computations.

Because we study global estimates here, I am able to lay bare differences between Total Utilitarianism and Generation-Relative Utilitarianism. I also test to see how sensitive Generation-Relative Utilitarianism is to the choice of well-being subsistence and the weight placed on generation-centred prerogatives. In a central case we find that optimum population size is lower and the optimum standard of living higher than they are (respectively) in the contemporary world. In view of the pervasiveness of the externalities that we inflict on one another, there is a strong case for consumption policies, twinned with public expenditure to help households to plan their families in an informed way. Successful programmes in today's poor countries have involved not only the government, but also charities and non-government organizations. Community engagement in family planning is an essential ingredient in any such programme.[28]

It is today commonly thought that accumulation of produced capital and advances in technology will see us through, that they are the means of avoiding further dependence on the

---

28 This is one of the conclusions in the accompanying paper with Aisha Dasgupta.

biosphere even as global living standards rise. The suggestion is the subject of Section 12. There I review historical and archaeological studies where scholars have identified successes and failures of past societies to meet local environmental problems. We are concerned in this essay with our global impact on the biosphere. And that should make one circumspect. The evidence collated in Sections 10-11 suggests that the international economy is so interconnected and extensive, that whatever progress we may enjoy in economizing on one resource base can be expected to lead to greater pressure on some other resource base. Globally, reductions in consumption and reproductive externalities cannot occur without international engagement. A pre-condition of any agreement the global community is able to reach is a common acceptance that every generation is a trustee of the biosphere it has inherited.

Like Total Utilitarianism, Generation-Relative Utilitarianism offers a way to get a measure of the loss that is suffered when someone dies. And like Total Utilitarianism, the theory can be used to get a measure of the loss that would be entailed if the human race were to become extinct. Loss in the latter case would be forward looking, it would reflect the value of lives that would be foregone on account of extinction, as viewed from a generation-relative perspective. In a remarkable work the writer Jonathan Schell drew our attention away from that line of reasoning and spoke of the loss each of us would suffer if we were to learn that no one will follow us. In Section 13 I review Schell's argument and modify it by enlarging the scope of that loss.[29] By extending the sphere of human motivations we avoid a free-rider problem to which Schell's ethics could be vulnerable. I then use the deep human need to live through time rather than in time to arrive at a view of stewardship of Earth. Acknowledgement of stewardship points also to an implicit understanding among the generations to protect and promote the biosphere's ability to support life. But it arrives at it from a different direction from the one we are led to by Generation-Relative Utilitarianism. The amended account of a person's well-being is not meant to be a substitute for Generation-Relative Utilitarianism, it is a complement to it. Nor is the implicit understanding among generations contrary to Generation-Relative Utilitarianism's idea of agent-based prerogatives. The viewpoints may speak different dialects, but they speak the same language. Or so I will argue.

Part I

Foundations

5 Genesis Under Total Utilitarianism

Total Utilitarianism was applied to a timeless economy by Meade (1955) and extended by Dasgupta (1969) to a world facing an indefinite future. In the latter publication it was shown that the theory encourages large populations. I first reproduce that finding by working with a stripped-down version of the model in Section 3 of Dasgupta (1969).

29 The passages there are taken from Dasgupta (2005a). In a penetrating essay Scheffler (2013) has further developed that line of thought.

5.1 Production and Consumption Possibilities

We imagine a timeless world, endowed with a finite stock of assets, of size K. We may think of K as an aggregate measure of produced capital and the biosphere, but often, to stress the salience of of the biosphere in population axiology, I shall refer to K as the biosphere. In the latter case we may think of K as being measured in units of biomass (tons, say). People are both producers and consumers. For convenience I keep knowledge and institutions separate from K (see below).

The Decision Maker (DM) knows that humans will be identical to one another in every respect (their productivity, their ability to convert consumption into well-being, and so on). DM also knows that when it it is applied to the stocks of assets, human labour can produce an all-purpose consumption good. Each person supplies a fixed amount of labour. As the world is timeless, stocks (e.g. the biosphere) and flows (ecological services) are the same. We will distinguish stocks from flows when we come to model the world economy moving through time (Sect. 9; Appendix 4).

Let Q be output of the consumption good. If population size is N, we follow Dasgupta (1969: Sect. 3) and Arrow's letter #3 and assume that

$$Q = AF(K,N), \qquad A > 0 \tag{1}$$

In equation (1) F is assumed to be homogeneous of degree 1, increases with K, N at diminishing rates, and $F(0,N) = F(K,0) = 0$.

A is a parameter in the model, not a variable. It is called "total factor productivity" in the economics literature, and can be interpreted as an aggregate measure of the society's knowledge base and its institutions. Because the model is timeless, I take K to be a parameter as well. That rules out accumulation or decumulation of all assets in the economy, a move that requires justification. Justification for working with a timeless economy can be found in Dasgupta (1969: Sect. 3), which studied an economy where production possibilities at any moment in time have the same structure as that in equation (1), but which moves through time, so that produced capital can be accumulated if it is so desired. The biosphere was taken to be a fixed factor of production, rather as the "indestructible land" that was imagined by Classical Political Economists of the late 18th century. In Dasgupta (1969: Sect. 3) it was shown that Classical Utilitarianism recommends that the economy be steered toward a stationary state in the long run. The stationary state in question depends on production possibilities and ethical parameters. In a stationary state nothing changes over time. So it is rather like the timeless economy we are studying here. I want to focus on population size, N. I want to do that in order that we can uncover in a simple way how the various other features of an economy affect N's optimal value.

We assume that either K or A (or both) is "large," which implies that the optimum value of N is large; which in turn means that it is a good approximation to regard N as a continuous variable. N is the sole variable in the model. Figure 1 depicts Q as a function of N.[30] We write marginal output

30 All functions are assumed to be twice continuously differentiable. That enables us to

($dAF(K,N)/dN$) as AF_N. It follows from the properties of the F-function that output per person ($AF(K,N)/N$) exceeds output of the marginal person (AF_N) and that both are declining functions of N. Figure 2 depicts average output and marginal output as functions of N.

If C is someone's consumption level, her personal well-being is $U(C)$. The U-function is assumed to increase with C but at a diminishing rate. We write marginal well-being ($dU(C)/dC$) as U_C, and the marginal of marginal well-being, d^2U/dC^2 as U_{CC} . Thus $U_C > 0$ and $U_{CC} < 0$.[31]

Positive well-being records life as good for the person, negative well-being records life as not good. U is positive at large values of C but negative at small values of C. It follows that there is a unique value of C at which U is zero. We interpret C to be the living standard and write the standard of living at which $U = 0$ as C^S. Thus $U(C^S) = 0$. Meade (1955) referred to C^S as "welfare subsistence." It would be more accurate to call C^S "well-being subsistence,", which is what I do here. Figure 3 depicts a U-function with those features.

As the economy is timeless, there is no scope for investing or disinvesting. All that is produced is consumed.

5.2 Social Well-Being and the Sidgwick-Meade Rule

In applying Total Utilitarianism's reasoning to population ethics it is cleanest to apply it to Genesis. I argue below that Genesis is also the scene where Total Utilitarianism doesn't run into any meaningful trouble. So we imagine Earth to be devoid of people.[32]

Let N be the number to be created. From equation (1), we know that total output $Q = AF(K,N)$. Because people are identical and marginal well-being declines as the living standard increases, an equal distribution of Q among all who are created is the ideal. If N people were created, each would receive Q/N units of the consumption good. Thus $C = Q/N = AF(K,N)/N$.

Social well-being is an aggregate of personal well-beings. If we denote social well-being as V, Total Utilitarianism says

$$V = NU(C) = NU(AF(K,N)/N) \qquad (2)$$

We are in search for the value of N at which V is at its maximum. Differentiating V in equation (2) with respect to N and setting the differential coefficient equal to zero tells us that optimum N satisfies the condition

$$U(C) = ((AF(K,N)/N)\text{-}AF_N)U_C = (C\text{-}AF_N)U_C > 0 \qquad (3)$$

---

use the calculus for solving for optimum population size. To assume otherwise would lead us into matters that are inessential to population axiology.

31 Edgeworth (1881) routinized the assumption of diminishing marginal well-being.

32 In his letter #3 of 6 June 2016 Arrow complained that he could not tell who the decision-maker is in the model. In my conversation with him in Cambridge later that month I reminded him that I had for many years interpreted the model as being pertinent to a Total-Utilitarian Decision Maker at Genesis.

Equation (3) is fundamental to Total Utilitarianism. We will call it the Sidgwick-Meade Rule. The intuition behind it is this:

At the optimum neither a small hypothetical increase in population nor a small hypothetical decrease would alter social well-being. Suppose now a marginal increase in numbers is contemplated by DM (the argument associated with a marginal decrease is analogous). The additional person would share Q equally with the population that was originally contemplated. The increase in additional well-being that would obtain if that person were created would be her well-being, which is $U(C) = U(Q/N)$. But there would also be a decrease in well-being because all others would have slightly less consumption, amounting to the difference between average consumption and the contribution the additional person would make to total output. That potential loss in well-being in symbols is $(AF(K,N)/N\text{-}AF_N)U_C$. At the Genesis Optimum the potential gain and the potential loss in social well-being must be equal. The Sidgwick-Meade Rule asserts this.

Denote the solution of equation (3) by C^O and the corresponding population size by N^O. So $AF(K,N^O)/N^O = C^O$. N^O is the optimum population size and C^O is the optimum living standard. Let $\hat{C}$ be the consumption level at which average well-being equals marginal well-being. The Sidgwick-Meade Rule says $C^O < \hat{C}$. That is shown in Figure 3.

5.3 Measurability, Comparability, and the Aggregation of Personal Well-Beings

There is no absolute scale for measuring personal well-being. Standard decision theory for example yields a measure that is unique up to positive affine transformations. That says if U_i is a measure of individual *i*'s well-being and if α is a positive number and β is a number of either sign, then $\alpha U_i+\beta$ is an equally valid measure of *i*'s well-being. The ordering of states of affair represented by U_i is the same as the ordering represented by $\alpha U_i+\beta$. In the latter scale α represents the unit and β the level. We say U_i in that case is a "cardinal" measure. Concerned as it is with well-being differences among alternative states of affair, the theory does not require the notion of zero well-being.[33]

Population ethics in contrast requires that lives that go well be distinguished from those that don't go well. A life that goes neither well nor not-well is calibrated to be zero (in Sidgwick's system of ethics it is the hedonistic zero). Personal well-being is "strongly cardinal" if its measure is unique up to positive linear transformations (i.e. proportional transformations). That says if U_i is a measure of *i*'s well-being and if α is a positive number, then αU_i is an equally valid measure of *i*'s well-being. A population ethic that is derived from Total Utilitarianism requires individual well-beings to be strongly cardinal.

When applying Total Utilitarianism to the Genesis Problem we made an implicit assumption, that well-beings are fully comparable across individuals. There was no need to mention it because

33 For fixed populations Harsanyi (1955) built Utilitarianism on the basis of the theory of choice under uncertainty.

people were assumed to be identical in all respects. In a heterogeneous population the assumption should be made explicit. Without full comparability of personal well-beings it would not be possible to construct an ethics with which to derive population policies.[34]

The idea of full comparability we need is familiar from weights and measures. The weights of objects (measured, say, in a vacuum-sealed flask at ground level in a given latitude) are fully comparable. Suppose we find *x* to be a heavier object than *y*. If that is to be a meaningful finding, the units in which they are measured must be the same; it's no good measuring *x* in ounces and *y* in grams. Reference to grams and ounces tells us that we can say a lot more than merely that *x* is heavier than *y*: we can say how much heavier *x* is proportionately than *y*. The reason we can is that if *x* is found to be twice as heavy as *y* using one system of units (ounces), it will be found to be twice as heavy as *y* using any other system of units (grams). And that's because an ounce is proportional to a gram. We can move from one system of units to another with impunity so long as the corresponding transformations (grams to ounces) are applied consistently.

We can say even more. Because physical theories tell us that addition and subtraction are legitimate (even required) operations on weights, each scale in the set of admissible scales is proportional to any other scale in the set, and all scales that are proportional to a scale in the admissible set are also in the admissible set. Non-linear scales don't belong.

Personal well-beings are taken to be fully comparable here if multiplying each individual's well-being by a constant positive number α is ethically of no significance. Other things equal the Total Utilitarian would rank the pair of personal well-beings {4,11} above the pair {5,9}, because 15 is bigger than 14. Full comparability says that if we multiply each individual's well-being by a positive number α, the ranking of the resulting pairs remains the same. They remain the same because $4\alpha+11\alpha$ ($= 15\alpha$) is bigger than $5\alpha+9\alpha$ ($= 14\alpha$) for all positive α. The Total Utilitarian can choose any value of α he likes without compromising his ethics.

Measurability and comparability of personal well-beings are closely related to possibilities of aggregating them into a measure of what we have been calling social well-being. There would be few restrictions on the social well-being function if personal well-being was measurable in an absolute scale. As there *is* no absolute scale, DM has to be circumspect before identifying an ethically defensible social well-being function. Social well-being functions cannot be chosen willy-nilly.[35]

---

34 Sen (1970) is the classic on the measurability and interpersonal comparability of individual well-beings in fixed populations. By extending his analysis to population ethics, it is easy to show that dropping full comparability would yield only a partial ordering of population policies. Partial orderings may be the best that the ethicist can hope for in practice, but no theory of social ethics should start with partial orderings.

35 Koopmans (1972), Hammond (1976), d'Aspremont and Gevers (1977), Maskin (1980), and Blackorby and Donaldson (1982) are studies on the connections between measurability, comparability, and aggregation of personal well-beings.

Total Utilitarianism aggregates personal well-beings by summing them. The theory requires personal well-beings to be strongly cardinal and interpersonally fully comparable. The reasoning confirms that the Sidgwick-Meade Rule is invariant with respect to α.[36] Well-being subsistence, C^S, remains the same under all proportional transformations of the U-function. That's because $\alpha U(C^S) = 0$ for all positive α. In the model we have used here to study the Genesis Problem, α was set equal to 1 so as not to add another symbol.

5.4 Zero Well-Being

Sidgwick (1907: 124-125) spoke of "neutral feeling" when suggesting ways to identify C^S:

"If pleasure ... can be arranged in a scale, as greater or less in some finite degree, we are led to the assumption of a hedonistic zero, or perfectly neutral feeling, as a point from which the positive quantity of pleasures may be measured... For pain must be reckoned as the negative quantity of pleasure, to be balanced against and subtracted from the positive in estimating happiness on the whole; we must therefore conceive, as at least ideally possible, a point of transition in consciousness at which we pass from the positive to the negative. It is not absolutely necessary to assume that this strictly indifferent or neutral feeling ever occurs. Still experience seems to show that a state at any rate very nearly approximating to it is even common: and we certainly experience continual transitions from pleasure to pain and *vice versa*, and thus (unless we conceive all such transitions to be abrupt) we must exist at least momentarily in this neutral state."

Zero well-being is a defining notion in population ethics. Sidgwick's reference to neutral feeling invites us to assess life from the inside. But the idea of neutral feeling and the corresponding idea of a neutral life also point, even if several steps removed, to a comparison of life with non-existence. That latter exercise requires calibrating well-being in terms of something outside our experience, which is why the exercise is regarded as questionable by some philosophers. Yet, when in deep despair people have been known to say they would rather not have been born, an utterance that doesn't sound incomprehensible. Comparison of life with non-existence is an unavoidable exercise in population ethics because reproduction is never a certainty. Suppose a couple understands that there is a 90% chance of producing a happy child and a 10% chance they will be unable to conceive. Neutral feeling in Sidgwick's sense cannot cover the latter event in the couple's reasoning because the intended child doesn't exist there.[37]

Nagel (1979b: 2) famously suggested that death is not an unimaginable condition of the living person: "... the value of life and its contents does not attach to mere organic survival: almost everyone

36 If $U(C)$ is replaced by $\alpha U(C)$ in equation (3) and U_C by αU_C, α cancels from the two sides.

37 There are philosophers who advocate evaluating states of affair in terms of an impersonal good (Griffin, 1986, offers an account of that viewpoint). As the example in the text suggests, even for them comparison of life with non-existence would seem to be a required move.

would be indifferent (other things equal) between immediate death and immediate coma followed by death twenty years later without reawakening."[38] He also suggested that death is a mere blank, and that it can have no value whatever, positive or negative. I am unable to tell whether by that Nagel meant the blank can't be used as a benchmark against which other states of affair are compared; but he went on to suggest that one can imagine non-existence by imagining being in a coma for the rest of one's life. For Nagel non-existence is the real blank, being totally unconscious for the rest of one's life is a simulation of that blank.

We are thinking of someone's life as a whole, not her life at a moment in time. There are thinkers who believe the whole is the sum of its momentary parts; there are others who believe the whole to be a non-linear function of those parts. We don't need to adopt a position on that because excepting for the model we analyse in Sections 9 and 11, there will be no occasion to slice someone's life into parts. The U-function here is calibrated by using Nagel's "blank" as a point of reference. Without loss of generality I am attaching the number 0 to that point. Zero well-being is therefore the measure of a person's life that, taken as a whole, goes neither well nor not-well for her. In view of the additive structure in Total Utilitarianism's conception of aggregate utility, $U = 0$ is also the level of well-being at which, in DM's judgement, an additional life adds no further value to the world that contains it.

I have heard it said that $U = 0$ is the point of indifference between dying and continuing to live, or the point of indifference between life and death. In Section 6 we uncover the reason that interpretation is misconceived. The reason also steers us away from the thought that in the contemporary world C^S is an absolute poverty line, in the sense the notion has been articulated in such figures as the World Bank's "1.90 dollars-a-day." As DM would see it, every life of poverty represents a bad state of affairs for that individual, and so makes the state of the world less good than it would otherwise be. In the contemporary world C^S would be higher than the World Bank's poverty line. C^S is the living standard at which life is neither good nor not-good. Determining C^S in a society will always prove contentious, but the move can't be bypassed in population ethics. It involves a deep and difficult value judgment. I discuss C^S further in Section 5.7 in the context of socially-embedded well-being functions.

5.5 Optimum Population Size

Total Utilitarianism says social well-being is the product of population size (N) and average well-being (U). Loci of $\{N,U\}$ pairs for which NU is constant are called isoquants of NU. The isoquants of NU, being rectangular hyperbolae, asymptote to $U = 0$ as N tends to infinity (Fig. 4). As Rawls (1972: 162-163) noted, that means tradeoffs between N and U in the product NU are such that no matter how small is U, so long as it is positive a sufficiently large N can compensate for a

38 The point is develped in Section 7, but I apply it to a different purpose from Nagel's.

reduction in U (as long as it remains positive of course). Parfit (1984: 425-441) found that feature of Total Utilitarianism repugnant. So he called it the Repugnant Conclusion (RC). There is a reverse of the Repugnant Conclusion, aptly known as the Reverse Repugnant Conclusion (RRC). It too is a feature of Total Utilitarianism, and it says for any world where a given number of people have lives that of excruciating distress, there is a worse world where a vast number of people have lives that are just below $U = 0$ (Mulgan, 2002). But it is RC that has given rise to an enormous literature, so we study it in detail.[39]

The question arises whether Total Utilitarianism commends large populations when it is tested in a world facing resource constraints. Dasgupta (1969) had previously shown it does; the theory advocates population numbers at which the standard of living (C^O) is proportionately not much above well-being subsistence (C^S). I rehearse the argument here and conduct a sensitivity analysis of the optimum with respect to various parameters of the model.

The simplest expressions of $AF(K,N)$ and $U(C)$ are power functions:

$$AF(K,N) = AK^{(1-\rho)}N^{\rho}, \qquad A > 0, \quad 0 \leq \rho < 1 \tag{4}$$

$$U(C) = B\text{-}C^{-\sigma}, \qquad B > 0, \quad \sigma > 0 \tag{5}$$

Equation (4) is widely used by economists to reflect production possibilities. The parameter ρ reflects the productivity of labour. Output is an unbounded function of population numbers, but Nature imposes a restraint on the rate at which output can expand with population. The latter is reflected in the condition $\rho < 1$. ($(1\text{-}\rho)$ is the productivity of K.) It will be confirmed below that optimum population size is a continuous function of ρ so long as it is less than 1. Arrow ("Some Random Thoughts on Birth and Death," letter #3) observes that optimum population is discontinuous at $\rho = 1$ and remarks that there is something wrong with Total Utilitarianism. I have excluded the case from equation (4) purely on empirical grounds, but I suggest below that we should not even otherwise be bothered by the discontinuity.

The other limiting case ($\rho = 0$) corresponds to a world where the biosphere offers a pure consumption good. That too is an unrealistic assumption, but if kept within the classroom for illustrative purposes it is not an absurd assumption.

Ideal national income accounting would interpret ρ to be the share of total output attributable

39 See, among others, McMahan (1981), Sikora (1981), Hurka (1983), Sterba (1987), Temkin (1987), Cowen (1989), Arrhenius (2000), Räikkä (2002), Tännsjö (2002), Broome (2004), Huemer (2008), and Parfit (2016). Arrhenius, Ryberg, and Tännsjö (2017) contains an outstandingly clear account of RC and its ramifications for population axiology. I am grateful to Christopher Cowie for discussions on both the Repugnant Conclusion and the Reverse Repugnant Conclusion.

to labour. There is an enormous empirical literature offering estimates of ρ. They tend to lie in the range 0.6-0.7. Because I am including the biosphere in the accounts, ρ should be taken to be smaller. In numerical exercises I will assume, solely for computational ease, that $\rho = 0.5$.

$U(C)$ in equation (5) is defined by two parameters, B and σ, both positive numbers (Fig. 5). U is bounded above, the least upper bound being B. Ramsey (1928) spoke of B as Bliss. $1+\sigma$ is the absolute value of the percentage rate at which marginal well-being changes with each percentage rate of increase in consumption (i.e. $1+\sigma = -\mathrm{dlog}(U_C)/\mathrm{dlog}C$). Which is why $1+\sigma$ is called the "elasticity of marginal well-being with respective to consumption." The elasticity exceeds 1 in the U-function of equation (5). Note also that $C^S = B^{-1/\sigma}$.[40]

Let e be the base of natural logarithms. We know from elementary number theory that it approximately equals 2.72. On using equations (4)-(5) the Sidgwick-Meade Rule reduces to

$$C^O/C^S = [1+(1-\rho)\sigma]^{1/\sigma} < e^{(1-\rho)} < e \approx 2.72 \qquad (6)$$

Equation (6) is the Rule in a convenient form; it relates the optimum living standard directly to well-being subsistence.[41]

For ecologists it proves more natural to recast equation (6) in terms of population numbers. Let N^S be the maximum human population the biosphere can support if everyone consumes at the level C^S. We call it the biosphere's *human carrying capacity*. It follows that N^S can be determined from the relationship

$$C^S = AK^{(1-\rho)}[N^S]^{\rho}/N^S = AK^{(1-\rho)}[N^S]^{\rho-1} = A(K/N^S)^{1-\rho} \qquad (7)$$

The bigger is K (or for that matter, A), the larger is N^S. Using equation (7) in equation (6) yields

$$(N^S/N^O)^{(1-\rho)} = [1+(1-\rho)\sigma]^{1/\sigma} < e^{(1-\rho)} < e \approx 2.72 \qquad (8)$$

It will prove useful to combine equations (7)-(8) and write them jointly as

$$C^O/C^S = (N^S/N^O)^{(1-\rho)} = [1+(1-\rho)\sigma]^{1/\sigma} < e^{(1-\rho)} < e \approx 2.72 \qquad (9)$$

Neither the size of the biosphere (K) nor A plays a role in determining C^O/C^S, nor for that matter N^S/N^O.[42] If K (or A) were to be bigger, Total Utilitarianism would respond entirely by adjusting the optimum population size N^O. This could be seen as an expression of pro-natalism. But pointing in the opposite direction gives a different impression. Suppose, for example, K was to be smaller. The same feature of equation (9) now says that the adjustment to the optimum would be made entirely through a reduction in population size; the standard of living would be fully protected. When applying

40 The U-function in equation (5) is bounded above but unbounded below. In applied work the latter property should be an unacceptable feature because unboundedness is inconsistent with the axioms of choice under uncertainty (Arrow, 1965). But here we are engaged in theoretical reasoning. The U-function in equation (5) has the virtue of revealing the structure of Total Utilitarianism in a very simple manner. I return to this presently.

41 To confirm equation (6), use equations (4)-(5) and the fact that $[C^S]^{-\sigma} = B$ in the Sidgwick-Meade Rule.

42 I owe this observation to Kenneth Arrow in a previous correspondence.

the theory to global economic statistics in Section 11, I return to this feature of the calculus of Total Utilitarianism.

Sensitivity of C^O/C^S and N^S/N^O to σ

Equation (9) says Total Utilitarianism favours large populations. The optimum living standard is less than $e^{(1-\rho)}$ times well-being subsistence no matter what value of σ DM settles on. The equation also says that carrying capacity is less than e times the optimum population size. To illustrate, suppose $\sigma = 1$ and $\rho = 0.5$. Then $C^O/C^S = (N^S/N^O)^{(1-\rho)} = 1.5$, which means optimum living standard (C^O) is only 1.5 times well-being subsistence and carrying capacity is only 2.25 times the optimum population size (N^O). If the biosphere's carrying capacity is reckoned to be 12 billion people, the optimum global population size would be about 5.3 billion.

No doubt the idea of the biosphere's carrying capacity points to a range, not to an exact figure. Technological advances and improved efficiency in the way resources are used in production raise carrying capacity (Sect. 12). The notion, moreover, is not independent of ethical values (eq. (7)). The point remains though that Earth's life support system is bounded. N^S reflects that fact. Whether Total Utilitarianism should be seen to be advocating overly large populations ("pro-natalist", as some would say) depends entirely on the standard of living (C^S) at which $U = 0$. Identifying C^S involves a value judgement. Once it has been identified by DM, the factors determining C^O are the ethical parameter σ and labour productivity ρ.

Utilitarian value pluralists have interpreted 1+σ also as a measure of the aversion to inequality in the distribution of living standards.[43] Moreover, Utilitarian decision theorists say that when a person's well-being is measured on the basis of her attitude to risk in her living standard, 1+σ is also a measure of her risk aversion.[44] In order to obtain a feel for the way 1+σ influences the optimum living standard (simultaneously, the optimum population size), we subject the Sidgwick-Meade Rule to a full sensitivity test. To do that we compute C^O/C^S in equation (9) using alternative values of σ. Notice that, to vary σ is to vary well-being subsistence (eq. (5)), but that's another acknowledgement that identifying well-being subsistence involves a value judgment.

How does C^O vary in relation to C^S as σ assumes different values? Equation (9) says that C^O/C^S (= $(N^S/N^O)^{(1-\rho)}$) is a declining function of σ. That is intuitive. The larger is σ, the greater is the curvature of the U-function, which means the lesser is the normative significance of average consumption relative to population numbers in social well-being $NU(C)$.

Equation (9) also says that the larger is σ, the closer is C^O/C^S (= $(N^S/N^O)^{(1-\rho)}$) to 1.[45] If the

43 Pioneering contributions to the modern literature on income inequality measures are Kolm (1969) and Atkinson (1970).

44 The classics are Pratt (1964) and Arrow (1965). Dasgupta (2008) offered reasons why the plausible range of values for σ is (0,2]. Empirical studies of choice under uncertainty have usually revealed σ to lie in the range (0,1].

45 That's because $[1+(1-\rho)\sigma]^{1/\sigma} \rightarrow 1$ as $\sigma \rightarrow \infty$.

biosphere's carrying capacity is 12 billion people and σ is taken to be large, optimum population is close to 12 billion. Strong social aversion to consumption inequality (alternatively, strong risk aversion) goes with large optimum populations. I have known this result since my student days but still find it puzzling that inequality aversion (alternatively, risk aversion) should play so influential a role in population ethics. Admittedly, the theory we are invoking here says that all who are born are to be treated equally; even so, it isn't *a priori* obvious why an attitude toward consumption inequality (alternatively, consumption risk) should so influence the ideal size of population. *Ex post* it is obvious of course.

In equation (5) σ has been assumed to be positive. The limiting case, $\sigma = 0$, corresponds to the logarithmic U-function. The elasticity of U_C equals 1 in this case. To study that limiting case, let us without loss of generality choose the units of consumption so that

$$U(C) = \log_e C \qquad (10)$$

That means $C^S = 1$. Notice that $\log_e C$ is unbounded both below and above. Using equations (4) and (10) in the Sidgwick-Meade Rule yields $C^O = e^{(1-\rho)}$. Note also that $e^{(1-\rho)}$ is the upper limit of $[1+(1-\rho)\sigma]^{1/\sigma}$ (eq. (9)). We have therefore confirmed that $\log_e C$ corresponds to the U-function in equation (5) for the case $\sigma = 0$.

There remains the case where the elasticity of U_C with respect to C is less than 1. The U-function is

$$U(C) = C^{\varphi}-D, \qquad 0 < \varphi < 1, \qquad D > 0 \qquad (11)$$

The elasticity of U_C in equation (11) is $1-\varphi$, and $C^S = D^{1/\varphi}$. Notice that the U-function in this case has exactly the opposite asymptotic properties to those of the U-function in equation (5). In the previous case U was bounded above but unbounded below. In the present case the U-function is unbounded above but bounded below. That makes a difference to the character of optimum population.

Straightforward application of equations (4) and (11) in the Sidgwick-Meade Rule yields

$$C/C^S = (N^S/N)^{(1-\rho)}) = [[1-(1-\rho)\varphi]^{-1/\varphi} \qquad (12)$$

It is simple to check that C/C^S, when expressed as a function of φ, is not bounded above. But then, nor is C^S bounded above. Correspondingly, N^S is small when φ is small, and so is N. In what follows, we confine our numerical calculations to the case where U satisfies equation (5).

Equations (5), (10)-(11) exhaust the class of U-functions for which the elasticity of U_C is independent of C and for which marginal well-being declines with C. As class room examples, iso-elastic U-functions are useful, but they are not believable. That's because any U-function that is unbounded at either end is vulnerable to St Petersburg Paradoxes. Arrow (1965) proposed that the elasticity of marginal well-being should be expected to be less than 1 for small values of C but greater than 1 at large values of C. Therefore, at some intermediate value of C the elasticity of U_C would equal 1. Any such U-function would be bounded at both ends. The Sidgwick-Meade Rule (equation

(3)) is valid quite generally and covers *U*-functions that satisfy Arrow's proposed conditions.[46]

Sensitivity of C^O/C^S and N^S/N^O) to ρ

In letter #3, Arrow uses equation (9) to study the sensitivity of C^O/C^S to ρ. He notes that C^O/C^S is a declining function of ρ and tends to 1 as ρ tends to 1. Because ρ is a measure of labour productivity relative to the biosphere, the finding says that an increase in labour productivity lowers the optimum living standard and correspondingly raises the optimum population size. Arrow finds that disturbing.

But we can put the matter in reverse and say that C^O increases if ρ is reduced. Which is why it is hard to know what our prior intuitions should be on the matter. We know that as a function of ρ, the range of C^O/C^S is bounded. Should we on intuitive grounds have expected the ratio to move non-monotonically within that range as ρ is made to increase? The key finding is equation (9), which says that the range within which C^O/C^S (equally, N^S/N^O) lies is narrow. That is another way of saying that the bounds on the optimum are tight. The finding had been a surprise to me as a graduate student, but I had no reason even then to expect the result to have been otherwise.

Arrow observes that C^O/C^S is discontinuous at $\rho = 1$. He notes that if $A > C^S$, then $C^O = A$ and N^O is infinity, but that $N^O = 0$ if $A < C^S$. In a world where the average product of labour is independent of population size ($\rho = 1$), the biosphere imposes no constraint on labour productivity. It seems to me that is so unusual a case to consider, that no ethical theory should be discounted merely because it displays features there that are unexpected. In his famous paper on optimum saving in an economy facing an infinite future, Ramsey (1928) discovered that Total Utilitarianism is incoherent for the problem because it awards the well-beings of future generations the same weight as the weight awarded to the present generation. The incoherence is a reflection of the fact that infinite sums of well-beings do not converge in that setting (I discuss this further in Section 9.1). But the problem of optimum saving in the economic model Ramsey worked with has a solution if a positive rate is used to discount the well-beings of future generations. We can say that Total Utilitarianism's response to the question of optimum saving is discontinuous at zero discount rate.[47]

5.6 Critical-Level Utilitarianism

So as to avoid what Parfit regards as a Repugnant Conclusion (RC) of Total Utilitarianism

---

46 I know of only one study of intertemporal public policy in which the *U*-function is taken to satisfy the above Arrow conditions and a further one that needs to be invoked so as to keep optimal consumption away from 0, namely, that $U_C \to \infty$ as $C \to 0$. And that exception is Arrow and Priebsch (2014).

47 The greatness of Ramsey's paper does not lie in that observation, but in showing how the Utilitarian criterion can be so altered as to allow for an answer to the question of optimum saving even when no discounting is permitted. Roughly speaking, Ramsey re-normalized the Utilitarian aggregator by subtracting an infinity from the infinite utility sum, in such a way that the reconstructed Utilitarianism had an answer to the question he had started with; which was, "How much of a nation's output should it save for the future?" (Ramsey, 1928: 543)

(Sect. 6), Kavka (1982) and Parfit (1984) explored the idea that, other things equal, the creation of an additional person should be judged to be good by DM only if her well-being were to exceed a critical level $U^C > 0$. The proposal sidesteps in one move what Parfit (1984) called the Mere Addition Paradox, a close cousin of RC. Blackorby and Donaldson (1984) arrived at the proposal from a set of normative axioms. The theory is called Critical-Level Utilitarianism.[48]

There is a problem with it. Suppose DM selects births sequentially. Suppose also that the well-being of each of the first N people is expected to be $2U^C/3$. As this is a positive number, each person's life would be good by their own reasoned reckoning. Critical-Level Utilitarianism however tells DM that, other things equal, creating a further person whose well-being would also be $2U^C/3$ is a bad idea. I cannot imagine why DM should be persuaded.

A possible interpretation of the theory was suggested to me at a workshop on population ethics. We are to imagine that the authors of Critical-Level Utilitarianism re-calibrated the personal well-being function U, so that their U^C is the hedonistic zero (i.e. Classical Utilitarianism's $U = 0$). In that interpretation life is not good if U lies in the interval $[0,U^C)$. Under the re-calibration states of affair would be evaluated using $U(C)$-U^C as the basis of comparison. But if Critical-Level Utilitarianism is a mere renumbering of Total Utilitarianism, it is the same theory as Total Utilitarianism and is therefore unable to serve as an escape route from Parfitian concerns.

In support of Critical-Level Utilitarianism it has been suggested that a life in which a person's well-being is below U^C is not one she deserves as a human being (Feldman, 1995).[49] But in the present example the first N people will enjoy $2U^C/3$. There will be no inequality even when there are N+1 people. Moreover, DM knows they all will acknowledge that by their own reckoning their lives are good. Nevertheless, Critical-Level Utilitarianism requires DM to say they deserve better and to believe that the world would be a worse place should another person be born with the same quality of life as the N people who preceded. I know of no reason why DM should accept the evaluation.

Once a figure for well-being subsistence is reached on the basis of individuals' reasoned conception of their well-being, DM should acknowledge that life for a person is good at any standard of living exceeding it (the higher, the better, of course). And a person's life is good if she enjoys $2U^C/3$. Of course, it could be that the underlying economy is mismanaged and there are sufficient resources to enable people to enjoy a life in excess of U^C. But that's a different matter, it points to bad governance, it doesn't provide a reason for modifying Total Utilitarianism.

5.7 Socially Embedded Well-Being

There is room for the notion of desert in population axiology if we recognise that people are not egoists. It pays to study that.

We should not imagine that C^S is independent of society's experiences. Personal aspirations

48 See also Blackorby, Bossert, and Donaldson (1997). Broome (2004: 199-202) provides a supportive account of the theory.

49 I owe this reference to Krister Bykvist.

and what reads as success or failure are influenced by what is reachable, they are not entertained in vacuum. So too with what constitutes a life that goes neither well nor not-well. It appears to be deep in our psychology that our living standard relative to that of others, especially perhaps our peer group, matters to us. Veblen (1925) is a classic on this line of inquiry. He pointed to the competitive side of our nature. Douglas and Isherwood (1979) is another, more recent, classic in which the authors pointed to the sociability in us.[50] In either case C^S would be higher, the higher is the general living standard. That means zero well-being in the distant past would have been calibrated at a lower standard of living than it would be today. It also means that in today's poor societies C^S can justifiably be taken to be lower by people there than it is by people in rich societies. And it means that if the overwhelming majority of a society's population are very well off, a new-born in a poor household can be acknowledged to deserve better.[51]

In the formal model in which I derived the Sidgwick-Meade Rule, people are egoists. The assumption makes for simplicity of analysis, which is why we will continue to study that world. But it is as well to check that the Rule extends to cover worlds where personal well-being is socially-embedded.

Let $C\circ$ denote the average living standard. We assume now that a person's well-being is a function not only of her consumption level C, but also of $C\circ$. Thus $U = U(C,C\circ)$. In a Veblenesque world, U would satisfy

$$\partial U/\partial C = U_C > 0 \tag{13a}$$

$$\partial U/\partial C\circ = U_{C\circ} < 0 \tag{13b}$$

In order to have a meaningful problem for population ethics, we should also assume

For all $C\circ$, $U(C,C\circ) < 0$ for low enough values of C, (13c)

For all $C\circ$, $U(C,C\circ) > 0$ for high enough values of C. (13d)

Conditions (13c-d) imply that for all $C\circ$ there is a C^S such that

$$U(C^S,C\circ)=0 \tag{14}$$

It is immediate from conditions (13a-d) and equation (14) that C^S is an increasing function of $C\circ$. Write the function as $C^S(C\circ)$. In Figure 6 $C\circ$ is measured along the horizontal axis and C^S along the vertical axis. The curve $C^S(C\circ)$, which is upward sloping, intersects the 45° line at $\underline{C}\circ$.

---

50 Bourdieu (1984) is a deep sociological study of consumption behaviour. The evidence suggests that others influence our choices in every sphere of consumption. We coordinate our activities by appealing to social norms. Those norms could be the average of what others in one's society do. The norms are therefore determined within the social system. Dasgupta et al. (2016) is a formal enquiry into the implications of Bourdieu's empirical findings for public policy.

51 Veblen's observation on human psychology found a telling expression in a remark attributed to a Garry Feldman of Stamford, Connecticut, one of the wealthiest towns in the USA: "I might be in the top one per cent, but I feel that I am in the bottom third of the people I know." (*The Guardian*, Saturday 16 February 2013).

Empirically the interesting region in the quadrant is where $C\circ$ exceeds $\underline{C}\circ$.[52] It is an easy matter to confirm that the Sidgwick-Meade Rule (eq. 3) generalizes to read as

$$U(C,C\circ) = [(AF(K,N)/N)-AF_N](U_C+U_{C\circ}) \tag{15}$$

Equation (15) identifies the optimum living standard. The latter is above $\underline{C}\circ$, but the Rule commends a large population. It is large all the more so because a high consumption level doesn't amount much to the average person when all others also enjoy a high consumption level. Appendix 1 contains an explicit form of *U*-functions satisfying conditions (13a-d).

We mustn't use the production function of equation (4) to study the implications of Veblenesque well-being functions for Total Utilitarianism. At an optimum in our model everyone consumes the average. Using equation (4) in equation (15) tells us that the optimum living standard is independent of *A* and *K*. That's self-defeating for the purpose in hand. Note though that equation (4) represents an extreme special case, useful for pen-on-paper calculations; which is why I will continue to use it here for studying population ethics. To understand the implications of Veblenesque well-being functions, however, we need to make use of more general production functions, consistent with equation (1). In general settings the optimum living standard under Total Utilitarianism increases with *A* and *K*.

In contrast to conditions (13a-d), the *U*-function in the world described by Douglas and Isherwood could be modelled as follows:

Define $J(C,C\circ) = (C-C\circ)^2$ and assume that $U = U(C,J)$, with the properties $U_C > 0$ and $U_J < 0$. We now suppose that for all $C\circ$ there is a C such that $U(C,J) > 0$ and also a C such that $U(C,J) < 0$. Together they imply that there is a C^S such that $U(C^S,J(C^S,C\circ)) = 0$. Call the solution of this equation $C^S(C\circ)$. A reasoning similar to the one deployed in Appendix 1 then allows us to identify *U*-functions for which $C^S(C\circ)$ is upward sloping so long as $C\circ$ is not too small.[53]

5.8 Non-Archimedean Intuitions and Non-Additive Social Well-Being Functions

Parfitian concerns are directed at the tradeoffs Total Utilitarianism advocates between population size (*N*) and average well-being (*U*) at large values of *N*. To explore alternatives to the theory, we imagine DM begins by expressing social well-being as *V*. We write the *V*-function as

$$V = V(N,U) \tag{16}$$

Total Utilitarianism requires DM to value the world in terms of the function

---

52 $\underline{C}\circ$ may be interpreted as a physical-subsistence level of consumption. Below that a person withers away. C^S is less than $C\circ$ to the right of $\underline{C}\circ$. The Sidgwick-Meade Rule commends a standard of living above $\underline{C}\circ$.

53 In the April 2016 version of this essay I was tentative when remarking on socially embedded *U*-functions, even though Aisha Dasgupta and I had made use of it in analysing the idea of fertility desires. I am grateful to Kenneth Arrow and Robert Solow for encouraging me to explore socially-embedded *U*-functions in population axiology. For a study of the implications for optimum saving when consumption preferences are socially-embedded, see Arrow and Dasgupta (2009) and Dasgupta et al. (2016).

$V = NU$ (17)

In contrast, an economics literature in the first half of the twentieth century took J.S. Mill's Average Utilitarianism to be the basis for population ethics.[54] Average Utilitarianism says

$V = U$ (18)

Average Utilitarians postulated a production function unlike equation (1). They assumed that marginal productivity of population, AF_N, is an increasing function of N when N is small and a decreasing function of N when N is large, other things equal. Optimum population is at the point where average productivity equals marginal productivity (Gottlieb, 1945; see also Dasgupta, 2005a). Contemporary population ethicists have made convincing objections to the theory. Moreover, it recommends a vanishingly small population in the economic model we are working with here. So I ignore the theory.

As a way of avoiding the Repugnant Conclusion and related Parfitian puzzles such as the Reverse Repugnant Conclusion and the Mere Addition Paradox, population axiologists have enquired after tradeoffs between N and U that are intermediate between those that are reflected in equations (17)-(18). To avoid the Repugnant Conclusion some scholars have appealed to a non-Archimedean intuition, which says that no number (no matter how large an N) of very small goods (U only just above zero) can exceed the value of a smaller number (small N) of greater goods (U greatly above zero). Parfit (2016) has an elaborate and largely sympathetic discussion of the intuition and develops a supporting idea that there is an ethical hierarchy of the factors that give rise to U. The claim is that a world containing a population no matter how small, enjoying only goods and engagements that are somewhere in the hierarchy is better than a world containing a population no matter how large, enjoying only goods and engagements lower in the hierarchy. This amounts to a lexical ordering of the hierarchy of goods and engagements. We are told that virtuous activities are higher in the hierarchy than mere pleasurable ones, that reading Homer is superior to reading Spillaine, that Mozart beats muzak, and so on.[55]

I may not be alone in reading the directives as finger-wagging from the Common Room window. Many of us aspire to a balanced life. A life of all Homer and no Spillaine would be dull, and a life entirely of virtue is to be avoided if we want to maintain our sanity. Many people try to locate a combination of engagements that is ideal for their own reasoned projects and aims. And for nearly all of us socio-economic constraints, misjudgments about our abilities, and bad throws of the die curb our ability to get there. More than 60 years of empirical research by sociologists and economists has shown that people choose on the basis of their personal tradeoffs between luxuries and necessities, work and leisure, Giselle (Live from the Royal Opera House, Covent Garden) and Strictly Come

54 Gottlieb (1945) is a well-known account of the literature.

55 Arrhenius (2005) contains a lucid discussion of the claim that there is a hierarchy of engagements. Carlson (2007) constructed a mathematical formulation of the hierarchical structure of engagements, but didn't put it to work on population axiology.

Dancing, and so on. And in democracies they vote in line with those tradeoffs in mind. For them the tradeoffs are embedded in their U-functions. Which is why it pays to chase the non-Archimedean intuition as applied to $\{N,U\}$ pairs.

When we do that, the intuition requires the isoquants of $V(N,U)$ to asymptote to distinct values of U as larger and larger Ns are considered - the higher the isoquant, the greater the asymptotic value of U. Consider

$$V(N,U) = g(N)U,$$

in which

$$\mathrm{d}g(N)/\mathrm{d}N \equiv g_N > 0;\ g(N) = \mathrm{o}(N) \text{ as } N \to 0;\ g(N) \to \pi > 0 \text{ as } N \to \infty \tag{19}$$

This can be read as Total Utilitarianism, but under the proviso that population numbers are measured using a non-linear scale based on the application of the operator "g" on N. An example is $g(N) = \tan^{-1}N$.

Let a be a real number. In equation (19) N and U are related on the isoquant of $V = a$ as

$$U = U_a(N) = a/g(N) \tag{20}$$

Figure 7 presents isoquants of V, which imagines for concreteness that $g(N) = \tan^{-1}N$. Equation (20) says that social well-being V is positive on isoquants in the upper panel and is negative in isoquants on the lower panel.[56]

The isoquants of V resemble the rectangular hyperbolae associated with $V = a$ in equation (17) in all essential ways but one. The exception is that $U_a(N) \to a/\pi$ as $N \to \infty$; which means the higher is an isoquant, the greater is its asymptotic value of U as N tends to infinity. DM can avoid both the Repugnant Conclusion (RC) and the Reverse Repugnant Conclusion (RRC) by adopting the V-function in equation (19).[57]

On using equations (1) and (19), the rule that optimum population must satisfy in this modified version of Total Utilitarianism reads as

$$U(C) = g(N)[F(K,N)/N - F_N]U_C/Ng_N > 0 \tag{21}$$

It is a simple matter to check that equation (21) is invariant to positive linear transformations of the U-function (Sect. 5.3). The equation also tells us that the question isn't whether we can avoid RC or RRC in our ethical reasoning; we can. But there is a price to pay: the intuitive meaning of the V-function in equation (19) is unclear. In any case, as I argue below, for we humans Total Utilitarianism's problems lie elsewhere.

6 Death

---

56 Note also (i) $U_{-a}(N) = -U_a(N)$, meaning that the isoquants on the upper panel are mirror images of those on the lower panel; (ii) $NU_a(N) = aN/g(N) \to a$ as $N \to 0$, which means $U_a(N)$ behaves like a/N for small N; (iii) $\mathrm{d}U_a(N)/\mathrm{d}N = -[a/(g(N))^2]g_N$, which means the isoquants on the upper panel are negatively sloped, while those on the lower panel are positively sloped.

57 The Mere Addition Paradox (Parfit, 1984: ch. 19), which is related to RC, is also avoided.

In a moving discourse on the place of autonomy and responsibility in personal well-being, Williams (1993: 50-102) drew attention to an aspect of personal responsibility that starts not from what others may demand of someone, but from what that someone demands of himself. Williams reminded readers that Sophocles had reported that Ajax, being slighted by the award of Achilles' arms to Odysseus, had intended to kill the leaders of the Greek army. To prevent the massacre, Athena made Ajax mad. It is significant that Ajax's condition didn't affect his purposes; rather, it altered his perception. Thinking that he was killing Odysseus and the others, Ajax slaughtered the army's flocks of sheep and cattle. In Sophocles' account the despair arising from the shame Ajax felt on awakening left no option open to him but to take his own life. And Williams observed (p. 76) that when Ajax says he must go, "... he means that he must go: period."

Sidgwick in contrast offered a view of life that is at odds with Sophocles' account. In the chapter that introduces Utilitarianism to his readers, Sidgwick (1907: 414-415) wrote:

"... I shall assume that, for human beings generally, life on the average yields a positive balance of pleasure over pain. This has been denied by thoughtful persons: but the denial seems to me clearly opposed to the common experience of mankind, as expressed in their commonly accepted principles of action. The great majority of men, in the great majority of conditions under which human life is lived, certainly act as if death were one of the worst of evils, for themselves and for those whom they love..."[58]

Nagel (1979b) concluded that if death is an evil, it is the loss of life that is objectionable. The conclusion is incontrovertible to the secular mind, but there are at least three circumstances of death that should be distinguished, and they don't point in the same direction. There is death that comes naturally to one in the fullness of time; there is death that comes not from one's hands before one's time; and there is death that is brought on one by one's own deliberate action. Nagel contrasted the

---

58 Odysseus' tale of his encounter with Achilles in Hades gives striking expression to the thought that death is the worst evil that can befall someone. He reports having greeted Achilles by exclaiming that no one in the world will have been more blest than Achilles, and that as Achilles now lords over the dead with all his powers, he has no cause to grieve at dying. To which, as Odysseus recounts, Achilles retorted,

> "No winning words about death to *me*, shining Odysseus!
> By god, I'd rather slave on earth for another man -
> some dirt-poor tenant farmer who scrapes to keep alive -
> than rule down here over all the breathless dead."

(***The** Odyssey*, trans. Robert Fagels, Book 11 (lines 555-558), Penguin Books, 1996).

At the other extreme is the Chorus in *Oedipus at Colonus*, who were reported by Sophocles as having observed: "Not to be born is, past all prizing, best; but, when a man hath seen the light, this is next best by far, that with all speed he should go thither, whence he has come." (R.C. Jebb, *Oedipus at Colonus*, trans. R.C. Jebb (lines 1225-1229), Cambridge University Press, 1917).

For a remarkable tour of visions of the after-life from ancient to modern times in literature and in religious teachings, see Casey (2009).

first two, but didn't speak to the third. And it is the latter that should make us pause before accepting Sidgwick's conclusion that life, all in all, is a positive good for most people.

Religious prohibition, fear of the process of dying (the possibility of suffering pain, the feeling of isolation), the thought that one would be betraying family and friends, and the deep resistance to the idea of taking one's own life that has been built into us through selection pressure would cause someone even in deep misery to balk. It may even be that no matter what life throws at us we adjust to it, if only to make it possible to carry on. But the acid test for Sidgwick's inference that "life on the average yields a positive balance of pleasure over pain" is to ask ourselves whether we shouldn't pause before creating a person so as to imagine the kind of life that is likely to be in store for the potential child.[59] The desire to procreate springs from our deep emotional needs, and the direct motivation we have to create children can be traced to a wide variety of reasons (we noted a few in Section 1; see also the accompanying paper with Aisha Dasgupta), but here we are concerned only with the life of the prospective child.

In the passage in which he presented the Repugnant Conclusion, Parfit (1984: 388) recognized Total Utilitarianism's pro-natalist character, but then interpreted it in the following way:

"For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living."

The play on words in the passage has always baffled me. We are being asked to consider a figure for world population that will almost certainly be reached in the second half of this century (a figure that is unlikely to be sustainable at reasonable material comfort; Section 11), and are then made to imagine an Earth where, because of population pressure, people scramble for resources so as to eke out an existence, having *lives that are barely worth living*. But someone whose life is barely worth living doesn't enjoy a life of positive quality, she suffers from a life that is not only not good (as experienced by her), but is positively bad. In the contemporary world over a half billion people are malnourished and prone regularly to illness and disease, many of whom are also debt ridden, but who survive and tenaciously display that their lives are worth living by the fact that they persist in wishing to live. If you were to say that you would not wish the circumstances they endure on anyone, I wouldn't take you to mean their lives aren't worth living; I would take you to be saying that their circumstances are so bad that you wouldn't wish them on even your worst enemy, that something ought to be done to improve their lives, that if you were to disregard the countervailing needs you and

---

59 Nagel acknowledges that the loss encountered when someone's life is cut short can't be balanced on a one-to-one basis by replacing that someone with a new birth. In my readings on the subject I have found him (and Narveson, 1967, and Heyd, 1992) to be among the very few who (at least until recently) acknowledged the asymmetry. Below (sect. 7) I make use of the asymmetry in developing population ethics.

your household may have, you wouldn't want to *create* children facing those circumstances.[60]

Death *relieved* the intolerable pain Ajax experienced on awakening from the madness Athena had inflicted on him. Ajax knew it would, which is why he chose it. It was better for him that he paid the price of death than that he carried on. The inference Sidgwick drew from the fact that death is generally thought to be one of the worst evils, namely, that life on the average yields a positive balance of pleasure over pain, is altogether unfounded, and I cannot imagine how so profound and careful a thinker could have made such an elementary arithmetical error. That death is a horror to most people doesn't imply that life is on balance pleasurable. On the contrary, the greater is the horror that taking one's own life poses to someone (betrayal of one's family and friends, revelation of one's misery to others when one wants it to remain undisclosed even after death), the *more* he would be willing to carry on in a state of misery.[61]

To illustrate Sidgwick's error, imagine that in the units chosen to measure *U*, the horror of suicide for someone is -300. The person would choose to continue to live so long life offered her a value in excess of -300; and that could be as low as -299.99.

One way to interpret life in the range between the point at which a person takes his life and well-being subsistence is to view it as *bearable* but not good. The person would not contemplate suicide, but could wish he hadn't been born, that he didn't have to go through his life's experiences. Zero well-being is then the transition point from the bearable to the good.[62] That is why estimating well-being subsistence from people's behaviour or responses to questionnaires would be a mistake.[63]

---

60 Parfit's interpretation of well-being subsistence hadn't changed over the years. On the Repugnant Conclusion, he wrote recently, "Compared with the existence of many people who would all have some very high quality of life, there is some much larger number of people whose existence would be better, even though these people would all have lives that are barely worth living." Parfit (2016: 110).

61 Suicide isn't that rare a phenomenon. Annually over 1 million people take their own life worldwide (approximately 1.8 per cent of deaths are suicides) and 20-30 million people attempt suicide.

62 I owe this interpretation to Robert Solow.

63 The presumption that life on balance is inevitably good ($U > 0$) or that it is invariably bad ($U < 0$) precludes reasoned discussion on population policies. The former commends a large (even infinite) population; the latter commends an Earth devoid of people. That alone shows that unlike decision theory, population ethics requires of us to specify the zero of the *U*-function.

As far as I can tell, numerical measures of subjective well-being have been taken to be non-negative in all the large-scale surveys on happiness and life satisfaction that have been conducted in recent years (see Diener, Helliwell, and Kahneman, 2010; and Helliwell, Layard, and Sachs, 2013, for reviews of the findings). It would look as though designers of the questionnaires that have shaped the literature on empirical hedonism have also been much influenced by the thought that life satisfaction can never be negative (otherwise why do people continue to want to live?). I do not know whether people would respond differently to questionnaires on life satisfaction if, instead of a scale from, say 0 to 10, respondents were offered a scale from -5 to +5.

7 A Problem Like Sleeping Beauty

In the Genesis Problem Earth is devoid of people. The domain of discourse consists only of potential persons. In any Actual Problem Earth is inhabited. The question is whether further people should be created. The two settings are wildly different.

Consider the problem of Sleeping Beauty. She is alive but in a state of total unconsciousness (Sidgwick would regard her state to be the hedonistic zero). What makes her an interesting case for us here is that her life expectation is the prevailing average, in that with a modicum of medical attention she can be expected to complete the natural lifespan in a state of total unconsciousness. A small expenditure can however revive her immediately and fully, in which case it is confidently expected her lifetime well-being will be U^* (> 0). That's option X for her parents. Option Y is for them to conceive another child who would enjoy a lifetime well-being equal to U^*. However, under Y Sleeping Beauty will remain unconscious. We suppose that the couple have no special feelings for Sleeping Beauty. She was born unconscious, so they have never got to know her. Assume now that in all other respects X and Y have the same consequences. What should Sleeping Beauty's parents do?

If, as in a literal interpretation of Sidgwick's Utilitarianism, agreeable consciousness is the sole good and if the fact that something good would be the result of one's action is the basic reason for doing anything (the two together give rise to agent-neutrality in ethical reasoning) the couple should be indifferent between X and Y. But there are a number of reasons X should be viewed as the right option. They all involve taming the Utilitarian-Consequentialist reasoning I have been following here with further considerations.

7.1 Imperatives

I first list four reasons that have been deployed by philosophers studying problems similar to that of Sleeping Beauty and discuss their extensions. I then develop the fourth reason (Parental Projects) by invoking arguments in Scheffler (1982) to apply agent-centred prerogatives to population axiology in Section 8.[64]

(i) *Parental Obligation to Sleeping Beauty*

X is the right option because Sleeping Beauty's parents have an obligation toward her which they don't have toward a potential child. They have an obligation toward Sleeping Beauty because they were responsible for conceiving her. People don't have an obligation to become parents, but they

---

Daly et al. (2011) have found in one study that a major determinant of suicide is being at a well-being level significantly lower than others, thus implying that living in a country where well-being is high is a risk factor for committing suicide. This can be explained if the coefficient of variation in incomes in two countries, one rich the other poor, is the same and personal well-being accords with Veblen (1924), that, other things equal, U is a decreasing function of the average living standard relative to a person's own (Sect. 5.7 and Appendix 1).

64 Nagel (1986: 164) sees the central problem of ethics as, "(How) the lives, interests, and welfare of others make claims on us and how these claims, of various forms, are to be reconciled with the aim of living our own lives."

acquire one if they choose to become parents. It can even be argued that parents have an obligation toward their offspring that no one else has. Parental obligation provides an argument for choosing *X* over *Y*.[65]

(ii) *Sleeping Beauty's Claim on Her Parents*

Relatedly, Sleeping Beauty has claims on her parents. She is a person, it is *her* well-being that is subject to parental choice. She has a claim to be acknowledged by her parents, to be heard by them even though she is unconscious. The potential child has no comparable claims. Notice that this is not a neutral claim - that would be in the realm of deontology - the claim here is agent-relative (Sleeping Beauty's claims on her parents).[66]

(iii) *The Value of Life's Experiences*

Nagel (1979b: 2) suggested that life has a value to the person living it that is independent of anything she may experience:

"The situation is roughly this: There are elements which, if added to one's experience, make life better; there are other elements which, if added to one's experience, make life worse. But what remains when these are set aside is not merely *neutral*: it is emphatically positive. Therefore life is worth living even when the bad elements of experience are plentiful, and the good ones too meager to outweigh the bad ones on their own. The additional positive weight is supplied by experience itself, rather than by any of its contents."

Experience has value to a person over and above its content. Call that stand-alone value, Ω. When we say life is precious, even sacred, we probably point to Ω. Here we note that Ω gives content to the idea of autonomy. That life has a stand-alone value Ω is not a sufficient reason for creating a person, but Ω provides a reason for awakening Sleeping Beauty. She has a claim to life's experiences which she will miss if she isn't revived. To deny her that option is to deny her autonomy.[67]

(iv) *Parental Projects*

In his comments on the population problem, Seabright (1994) spoke of the parental desire to promote their family. Sleeping Beauty is their child. Over and above her parents' obligation toward her, their own common project includes protecting the integrity of their family. Other considerations aside, responding to Sleeping Beauty's needs trump's option *Y*. Choosing *X* is an agent-relative response on the part of her parents.

7.2 Wider Cases

---

65 Simon Beard has pointed out to me that the example can be expanded to encompass a wider range of obligations, for example, if the chooser between *X* and *Y* was Sleeping Beauty's friends (they would have no obligation to make a new friend).

66 Gosseries (2010) has a good discussion of this. He studies the obligations people have to one another, not simply parents toward their children. I am ignoring the obligation people may have (at least to themselves) to procreate in special cases, such as a population at risk of extinction from non-fecundity. See Section 13, on existential risks.

67 I am most grateful to Itai Sher and Krister Bykvist for helpful correspondence on this.

A problem like Sleeping Beauty gives rise to a number of variations. Consider a case where *X* is not available, but *X*' is. In *X*' Sleeping Beauty is revived at a small cost, but her life will not be good. Even though she will be able to function, she will suffer from discomfort and pain and will not lead a fulfilling life. All in all it is expected that her lifetime well-being will be U^{**}, a negative number whose absolute value is not large. Her parents can, as can others such as the physicians caring for her, point to Nagel's Ω and require that she be revived. That's one way to justify *X*'. Some would go a slightly different route and argue that Sleeping Beauty should be allowed to exercise her agency, her right to life's experiences.

Suppose even *X*' is not an option. Her parents, or for that matter anyone else, could argue that she should be released from the indignity of living in coma for the rest of her natural life. In that case they would give their consent to having her life-support system disconnected.[68]

Consider another variation. Option *Y* remains, but in place of *X* we are faced with *X*'', which involves reviving Sleeping Beauty with the understanding that her lifetime well-being will be $U^*-\varepsilon$, where ε is a small positive number. Any one of reasons (i)-(iv) is sufficient for her parents (or society, more generally) to choose *X*'' if ε was sufficiently small. Assuming her parents view states of affairs in a continuous manner, there is a value of ε, say ε^*, such that it is a matter of normative indifference to them whether *Y* is chosen, or whether Sleeping Beauty is revived with the understanding that her lifetime well-being will be $U^*-\varepsilon^*$. To acknowledge that ε^* is positive is to say that, other things equal, reviving Sleeping Beauty with an expectation of $U^*-\varepsilon$ (where $0 < \varepsilon < \varepsilon^*$) would have priority over creating a person with a lifetime well-being of U^*.

8 Generation-Centred Prerogatives in the Timeless World

Considerations (i)-(iii) in the problem Sleeping Beauty raises come close to what I am after, but don't quite get there. Consideration (iv) gets there, but only in so tangential a form for my needs here that I am reluctant to use it. The general formulation is in Scheffler (1982), who uncovered agent-centred concerns that people can justifiably use as prerogatives over agent-neutral demands when deliberating over courses of action open to them (p. 20):

"... a plausible agent-centred prerogative would allow each agent to assign a certain proportionately greater weight to his own interests than to the interests of other people. It would then allow the agent to promote the non-optimal amount outcome of his own choosing, provided only that the degree of its inferiority to each of the superior outcomes he could instead promote in no case exceeded, by more than the specified proportion, the degree of sacrifice necessary for him to promote the superior outcome."

In a society of *N* persons, those prerogatives would apply reciprocally, meaning that the state of affairs that would ensue would be the outcome of *N* choices, each having been guided by agent-

68 Removing Sleeping Beauty's life support system is not the same as killing her.

relative concerns. In population axiology the force of those prerogatives works unidirectionally.[69]

We are to imagine that each generation reasons collectively and chooses the size of the next generation. Therefore, the "agent" in each generation is that "generation", not the set of all generations beginning with that one. Because generations enter sequentially, the prerogatives we study are unidirectional, not reciprocal. Moreover, the restrictions that I impose on account of those prerogatives are only a weak version of Scheffler's account; they only involve the place of potential well-beings in present people's moral considerations. I modify Total Utilitarianism by considering a weighted sum of personal well-beings, the weight awarded by existing people to potential well-beings being lower than the weight they award their own well-beings. But people in the modified theory display the same concern for actual future people (one's descendants) as they do for themselves (that requirement is the moral we took away from the case of Sleeping Beauty). We will call the modified perspective Generation-Relative Utilitarianism.

I first put the theory to work on the timeless model of Section 5. Stocks and flows are indistinguishable. In Section 9 the theory is extended to an economy moving through time. Stocks and flows have to be distinguished there.

8.1 Two-Stage Decision

We imagine there are N_0 people, identical in every respect. They are the existing population. By assumption there are no environmental nor reproductive externalities. In that situation an individual's choice reflects the choice of all. The biosphere is of size K.

As the model we study here is timeless, each individual's life cycle is embedded in it. If C is someone's consumption level, his personal well-being is $U(C)$. The problem facing existing people is to determine the number of people to add to their world. We assume again that the world is thinly populated. It means K is large relative to N_0. It also means that if they were not to create further people, their standard of living would be high. Adding to their numbers will cost them some, but because they are Generation-Relative Utilitarians, they decide how many to add by appealing to the theory's calculus.

The decision facing the existing population involves two stages. At the first stage the number of additional people (N_1) is chosen. In making that choice people know that once born, the additional population will join them, both as producers and consumers, and enjoy the same status as they. Production and consumption take place in the second stage.

---

69 Arrow (1981) used this reasoning to develop a theory of income distributions arising from voluntary contributions. He (personal communication) was moved to advance Agent-Relative Utilitarianism by the words of the first century sage, Rabbi Hillel: "If I am not for myself, then who is for me? And if I am not for others, then who am I? And if not now, when?"

Recall (sect. 6) also Ajax's shame and the action he took to meet the obligation he had to himself.

Recall the production structure in equation (1). If N_1 people are born, total output will be $AF(K,N_0+N_1)$. As that output will be shared equally, each person's living standard will be $AF(K,N_0+N_1)/(N_0+N_1)$, which is C. That's the start of a backward-induction, an essential requirement in Generation-Relative Utilitarianism.

Without loss of generality the weight awarded by existing people to their own well-being is 1. The weight they award to the well-being of potential people is μ, where $0 \leq \mu \leq 1$. Let V be the representative person's conception of social well-being. Because her perspective plays a role in that conception, we will call V her *social valuation function*. Thus

$$V = N_0U(AF(K,N_0+N_1))/(N_0+N_1)) + \mu N_1U(AF(K,N_0+N_1))/(N_0+N_1)) \qquad (22)$$

The problem before the existing population is to choose N_1 so as to maximize V in equation (22). To contrast it to the Genesis Problem, we will call the maximization of V in equation (22) an Actual Problem.[70] Routine calculations show that the optimum value of N_1 satisfies

$$\mu U(C) = [(N_0+\mu N_1)/(N_0+N_1)][(AF(K,N_0+N_1)/(N_0+N_1))-(\partial AF(K,N_0+N_1)/\partial N_1)]U_C > 0 \qquad (23)$$

Equation (23) is the counterpart of the Sidgwick-Meade Rule in Generation-Relative Utilitarianism.

To contrast the two theories it is simplest to return to the power functions in equations (4)-(5). Using them in equation (23) and labelling the optimum C as C^{OO} and the optimum N as N^{OO} yield

$$C^{OO}/C^S = (N^S/N^{OO})^{(1-\rho)} = 1+[(N_0+\mu N_1)/\mu(N_0+N_1)](1-\rho)\sigma\}^{1/\sigma} \qquad (24)$$

Compare equations (9) and (24). A simple calculation shows that the optimum standard of living (C^{OO}) in the Actual Problem is larger than in the Genesis Problem (C^O). This is shown in Figure 5.[71] Correspondingly, optimum population is smaller in an Actual Problem (N^{OO}) than in the Genesis Problem (N^O). The smaller is μ, the fewer are the additional numbers (N_1).[72]

In Figure 5 $\hat{C}$ is the consumption level at which average well-being equals marginal well-being. The magnitude of C^{OO} relative to $\hat{C}$ depends on μ, ρ, σ, and N_0. It is easy to confirm that if the μ

---

70 μ corresponds to ε* in Section 7.2. Notice that the larger is ε*, the smaller is μ.
V in equation (22) is a direct extension of Total Utilitarianism and would reduce to it if μ = 1. I have arrived at it by applying Scheffler's reasoning to population ethics. Value Pluralism in contrast would require the formula to be derived from ethical axioms on the way individual U-functions are combined to yield the V-function. Welfare axiomatics do that. Blackorby and Donaldson (1982) applied welfare axiomatics to population ethics. Their techniques can be adapted to yield equation (22).

71 Proof: Use the fact that $(N_0+\mu N_1)/\mu(N_0+N_1) > 1$.

72 We may put the matter in a slightly more technical, way: There exists a function $N(\mu)$, which decreases with μ, such that if $N_0 < N(\mu)$, equation (24) yields the optimum population. But if $N_0 > N(\mu)$, the optimum policy is not to add to the existing population. Personal well-being in the latter case is $U(AK^{1-\rho}N_0^{(\rho-1)})$. If the existing population were to have the misfortune of being too many ($N_0 > N^S$), the personal well-being of each of the N_0 individuals would be negative. And *that* is over-population. (We are assuming that killing a fraction of the existing population so as to raise the well-being of those that remain to positive levels is forbidden.)

not much less than 1, C^{OO} is less than $\hat{C}$, but that if μ is small, C^{OO} is greater than $\hat{C}$. In Figure 5, the two possibilities are identified by C_1^{OO} and C_2^{OO}, respectively.

8.2 Reproductive Replacement

We now consider an extreme scenario that will prove useful when we come to study an economy moving through time in a stationary mode. Imagine that the optimum policy for the existing population is to replicate itself. That's when N_0 just *happens* to be a figure for which optimum N_1 equals N_0.

Define N by the equation $N_0 = N_1 = N$. $2N$ is the optimum population and N is the size of each of the two generations. Equation (22) then reduces to

$$C^{OO}/C^S = (2N^S/2N^{OO})^{(1-\rho)} = (N^S/N^{OO})^{(1-\rho)} = [1+(1+\mu)(1-\rho)\sigma/2\mu]^{1/\sigma} \qquad (25)$$

Equation (25) is the counterpart of the Sidgwick-Meade Rule in this model. As expected, equation (25) reduces to equation (9) if $\mu = 1$.[73]

Notice that C^{OO}/C^S is not bounded above by e. If μ is small, $C^{OO} > \hat{C}$ (Fig. 5). To obtain a quantitative sense of the optimum, we need numerical values of ρ, σ, and μ. As previously we set $\rho = 0.5$. σ has been estimated from consumption behaviour under risk. $\sigma = 1$ is at the upper end of the range that has been found in empirical studies. For ease of computation I settle on that.

Stopping rules that are used by households to determine their family size could in principle be used to estimate μ, but I don't know of any study that has gone that route. No doubt household behaviour isn't the exclusive source of ethical understanding (Section 1 drew attention to a number of reasons), but it would be wrong to ignore people's intentions altogether in reaching ethical directives. Casual empiricism on health and education expenditures in time and money on children in the West, especially perhaps on children with special needs, suggest that μ is considerably less than 1. People seem to place far greater weight on the well-being of their children than on the potential well-being of children who might have been born but weren't because couples chose not to have further children. I assume $\mu = 0.05$. This is to take a lunge in the dark, but I am using the figure only for illustrative purposes. Using the figures in equation (25) yields $N^S/N^{OO} \approx 39$, meaning that the biosphere's carrying capacity for humans is (approximately) 39 times optimum population. This is a far cry from the pro-natalism of Total Utilitarianism. Suppose C^S is taken to be below even the average annual income in the world's poorest countries; say 1,500 international dollars (Table 1). In that case equation (25) says C^{OO} = 9,375 international dollars. That's the average annual income in the World Bank's list of low middle-income countries.

Even though equation (25) has been derived for a timeless world, it reflects Generation-Relative Utilitarianism in an economy moving through time in a stationary mode. We confirm that in Section 9.

Notice that I am unable to estimate N^{OO}. That's because we don't know N^S. In Part II we will

---

73 Notice that C^{OO} is unboundedly large and N^{OO} vanishingly small if μ is infinitesimal.

make an attempt to estimate N^S from an on-going research programme of a group of environmental scientists.

8.3 Coherence

Total Utilitarianism requires people to award the same ethical weight to potential people as they do to and present and actual future people. That requirement has been questioned previously. Narveson (1967) noted difficulties in the requirement and recast Utilitarianism by observing that "we (utilitarians) are for making people happy, not for making happy people" (Narveson, 1973: 73). Narveson called his theory Person-Affecting Utilitarianism.[74]

Parfit (1976) observed that the theory yields an incoherent notion of goodness. As his critique would apply equally to any theory that recognises agent-centred prerogatives no matter how attenuated they may be, here is an example of what he, and in endorsing the criticism Broome (2004), meant.

Consider individual A and a potential person B. In what follows we apply the *ceteris paribus* clause on all other people. A's well-being is expected to be 11 in the prevailing state of affairs, or "social state", X. But A has the option of creating person B. Two social states that include B are possible. In one (we label it Y) A's and B's well-beings would each be 6. In the other (we label it Z) they would be 8 and 4, respectively. Y and Z differ in the way A and B share resources, a feature of those states of affair that are taken into account by A when contemplating whether to create B. We re-label X, Y, and Z as {11}, {6,6}, and {8,4}. The problem is to rank them in a situation where A exists but B doesn't. We say that B, were he to be created, would be of a different generation from A (hence, "Generation-Relative Ethics").

Denote ethical dominance by ">" and ethical equivalence by "=". We imagine A to be a Generation-Relative Utilitarian. Suppose the weight A awards to the well-being of a potential person is half the weight she awards her own well-being, the latter being set equal to 1. That means $\mu = 1/2$. It also means that so long as B is a potential person, $X > Z > Y$. But if B were born, he would be an actual person. Person A recognises that the ethical relation between Y and Z, the only remaining social states, would then be $Y = Z$. A's social valuation function is dependent on the state of affairs in which she constructs it. That's Parfit's criticism.

The trouble with the criticism is that it harks back to the Genesis Problem. As the name suggests, the Genesis Problem is viewed by someone residing nowhere. That is why Total

74 Narveson argued that we are under no obligation to procreate, even if, other things equal, it was guaranteed that the person born would have an outstandingly good life. Obligation is a strong word. In this essay I am exploring a particular way of evaluating states of affair, in which "well-being" serves as the coin with which states of affair are evaluated. Generation-Relative Utilitarianism is a lot closer to Total Utilitarianism than Narveson's theory. I have friends however who regard Generation-Relative Utilitarianism to be infested with ontological elitism. The criticism comes close to regarding potential persons as actual persons, a view that the problem of Sleeping Beauty asks us to reject.

Utilitarianism would say $Y = Z > X$ before B is conceived. Parfit took exception to Narveson's Person-Affecting Utilitarianism because the binary relation between social states is dependent on whether B is an actual person. Generation-Relative Utilitarianism has that same feature. In the present example it is A who is evaluating social states. She does that from somewhere, she is not nowhere.

It is key to Generation-Relative Ethics that the ranking of the three social states X, Y, and Z is a function of the state of affairs at which it is arrived at. The ranking corresponds to A's social valuation function. The perspective is from the current state. The normative content of the pair {6,6} in a world where B is only a potential person differs from the normative content of the pair {6,6} in a world where B is an actual person. Denote the ranking of the triplet $\{X,Y,Z\}$ at X (the social state at which only A is present) under Generation-Relative Utilitarianism by the relationships "$>_X$" and "$=_X$", and define ("$>_Y$", "$=_Y$") and ("$>_Z$", "$=_Z$"), correspondingly, at Y and Z.

Because person A subscribes to Generation-Relative Utilitarianism, she would choose not to add to the population, because viewed from X,

$$\{11\} >_X \{8,4\} >_X \{6,6\} \tag{26}$$

If A were to view the triplet from Y, however, the ranking would be

$$\{6,6\} =_Y \{8,4\} >_Y \{11\} \tag{27}$$

and from Z, it would be

$$\{6,6\} =_Z \{8,4\} >_Z \{11\} \tag{28}$$

There is no contradiction here. Once X ceases to be an element of the set of possible outcomes (because B has come into existence), the ranking over Y and Z is reversed.[75]

That a person's social valuation function could depend on the state of affairs she finds herself at is at odds with a long standing tradition in both moral philosophy and welfare economics. It has become common practice in both disciplines to seek a ranking of social states that is reached by the ethicist looking at states of affairs from outside. The incoherence Parfit thought he had noted in someone's ethical reasoning when she views her actions from where she is at has been used by philosophers even to dismiss calibrating zero well-being by reference to non-existence (more accurately, to an absence of all experiential states). Broome (1999: 168), for example, has argued that it cannot ever be true that it is better for a person that she lives than that she should never have lived at all; for if it *were* better, then it would have been worse for her if she had never lived at all; but if she had never lived at all, there would have been no her for it to be worse for, so it would not have been

75 Total Utilitarianism and General-Relative Utilitarianism are not the same normative theory. The binary relations over social states that are implied by the two are different. We should therefore distinguish them. Under Total Utilitarianism the social state in which the ethics is conducted does not matter. The ranking of $\{X,Y,Z\}$ is the same ($Y = Z > X$) regardless of whether B exists. If a temporal structure were imposed on the example, X would not be reachable if either Y or Z were realized; nevertheless the thought experiment that would give rise to inequalities (26)-(28) can still be conducted. I am grateful to Simon Beard and Shamik Dasgupta for clarifying the complexities of the example to me.

worse for her.

The chain of reasoning is correct enough, but the premise reads doctored. It is doubtful that anyone would acknowledge her life to be going well by saying it is better for her that she lives than that she were never to have lived. On those occasions we express a sense of well-being we tend to say that it is good to be alive, or that we are grateful to our parents for having conceived and nurtured us, or that we ought to do something for our society in return for our good fortune. But those are different thoughts, and they are meaningful. When one's well-being is compared with non-existence, the comparison is made from the perspective of the person herself, the comparison isn't made from a universal point of view. The person does not ask herself how she would have felt had she not been born. She knows that to ask that would be absurd.

The preference reversal displayed in inequalities (26)-(28) translates to intergenerational inconsistency. Each generation's ordering over possible future worlds differs from those of future generations over the possible future worlds they in turn will face. Nevertheless, it is possible for the generations to arrive at choices that are consistent with one another's perspective. We now confirm that each generation could both influence and affirm the choices to be made by those that are to follow. We do that by extending the timeless model to construct a world moving through time.

9 Generations Across the Indefinite Future

Generations may be relatively easy to define within dynasties, but because individuals of all ages live side by side, they are a slippery concept when we want to identify them from the demographic structure of an entire economy. What we mean by a generation depends on the purpose to which the notion is needed (year of birth, those of working age, retired people, and so on). In welfare economics generations are often identified with time, as in, "the generation at time t". But justifying that move requires a number of demographic and normative assumptions. In Section 11 we will find a way to distinguish generations in a clean manner, but that will be in an economy in stationary state. Matters become difficult outside stationary states, but the problems are not insurmountable. Excepting in the formal models that follow, I shall keep the term "generation" vague, which is the way moral philosophers would seem to proceed.

Total Utilitarianism evaluates a generation's choice of their successors' numbers on the basis of a weighted sum of well-beings across the generations. Here is Sidgwick (1907: 414) on the matter:

"It seems ... clear that the time at which a man exists cannot affect the value of his happiness from the universal point of view; and that the interests of posterity must concern a Utilitarian as much as those of his contemporaries, except in so far as the effect of his actions on posterity - and even the existence of human beings to be affected - must necessarily be more uncertain."

Discounting for time differs from discounting the well-beings of future generations. An individual could choose to discount his own future well-being at a positive rate but feel that in the choice of public policy the well-being of future people should receive the same weight as his own. Frank Ramsey would appear to have held that viewpoint (see footnote 77). Sidgwick in the passage just quoted spoke against the thought that it is ethically legitimate to discount the well-being of future generations merely on grounds that future generations will appear only in the future. He spoke to future generations' futurity.

We consider three reasons for applying generation-discounting: (i) discounting the well-beings of future generations on grounds that they will appear only in the future; (ii) discounting for risk of extinction of the human race; (iii) exercising generation-centred prerogatives.[76] In Section 9.1 we assume there is no risk of extinction and we identify periods with generations so as to draw out sharply that generations arrive sequentially in time. But that means even adjacent generations don't overlap: time and the generations coincide. That assumption is removed in Section 9.2, and in the rest of the Arrow Lecture we study population ethics in a world of overlapping generations.

So as to keep generation-discounting from contaminating the force of generation-centred prerogatives, I eschew the futurity argument and assume from Section 9.3 onward that only the risk of extinction and generation-centred prerogatives matter. However, so as to expose the underlying structure of Generation-Relative Utilitarianism, even the risk of extinction is eschewed in Section 11, which is when we come to estimate globally optimum population size and the optimum standard of living.

Because I want to stay close to the literature on discounting, we assume in Sections 9.1-9.2 that population is exogenously given (i.e. future population is not subject to choice). We return to population ethics in Section 9.3.

9.1 Discounting Future Generations

Sidgwick's formulation of (Classical) Utilitarianism was the opening observation in Frank Ramsey's famous paper on the optimum rate of saving in a world shorn of risk (Ramsey, 1928). Ramsey regarded the practice of discounting future well-beings in such a world to be ethically indefensible (p. 543) and thought it "arises merely from the weakness of the imagination." In a book that laid the foundations of the modern theory of economic growth, Harrod (1948: 40) judged discounting future generations to be "... a polite expression of rapacity and the conquest of reason by passion." Strong words, and to many economists Ramsey's and Harrod’s strictures read like Sunday pronouncements. Solow (1974a: 9) expressed the feeling exactly when he wrote, "In solemn conclave assembled, so to speak, we ought to act as if the [discount rate on future well-beings] were zero."[77]

---

76 Beard (2018) in unpublished notes discusses further reasons.

77 In a talk on 28 February 1925 before fellow Apostles (a secret mostly-undergraduate discussion society in Cambridge University, known to its fraternity as "The Society"), under the title, “Is There Anything Left to Discuss?”, Ramsey had expressed a different viewpoint:

"My picture of the world is drawn in perspective, and not like a model to scale. The foreground is occupied by human beings and the stars are as small as threepenny bits. I don't really believe in astronomy, except as a complicated description of part of the course of human and possibly animal sensation. I apply my perspective not merely to space but also to time. In time the world will cool and everything will die; but that is a long time off still, and

Parfit (1984) applied sophisticated reasoning in support of the Sidgwick-Ramsey-Harrod view on discounting for time and the generations. That view has influenced one strand of the welfare economics of global climate change (Broome, 1992, 2012; Cline, 1992; Stern, 2006), but it appears to be a minority opinion among economists. Economists generally have adopted a democratic attitude, noting that because people do in fact discount for time and the generations, the social decision maker should follow the practice (Nordhaus, 2007).[78] And there are economists (e.g. Arrow and Kurz, 1970; Arrow, 1999; Dasgupta, 2005b, 2007b) who have questioned the Sidgwick-Ramsey-Harrod view because they have been persuaded by an equally sophisticated reasoning as Parfit's, which showed that to *not* discount the well-being of future generations can have deeply problematic implications. The argument was fashioned in a remarkable set of papers by Koopmans (1960, 1965, 1967, 1972, 1977).

There are two strands to Koopmans' reasoning, and they come at the question of time/generation discounting from different directions. One strand (Koopmans, 1965, 1967) begins by noting that the productivity of investment creates an asymmetry between the present and future. Seedlings grow to become trees, which is to say investment has a positive return (a foregone unit of consumption today can generate more than a unit of consumption in the future). The asymmetry is so built into the arrow of time, that symmetric treatment of well-beings across the generations is questionable. We study that possibility first (the example is taken from Arrow, 1999: 14).

Time is divided into periods: $t = 0,1,\ldots$ ad infinitum. We are to imagine that population is projected to remain constant over time. Generations coincide with periods. So we may speak of "generation-t" and "period t" interchangeably. Imagine that output consists of a steady stream of a completely perishable good, and that generation-0 has an investment opportunity in which a unit of consumption foregone yields a perpetual stream of r units of the consumption good from $t = 1$ onward. The rate of return on the investment, r, is positive.

---

its present value at compound interest is almost nothing. Nor is the present less valuable because the future will be blank." (Ramsey, 1931: 291).

Ramsey was of course speaking after dinner, behind closed doors, to his fraternity of Apostles. But as the talk predated his famous 1928 paper in the *Economic Journal,* I couldn’t tell for a long while whether he had changed his mind by the time he came to compose the paper or had just put on Sunday-best clothes for a journal publication (see also Arrow and Kurz, 1970). Recently Ken-Ichi Akao suggested an entirely convincing third explanation to me. He observed that in his talk to the Apostles Ramsey was giving expression to his personal discounting, whereas in his *Economic Journal* article, concerned as he was with socially optimum national saving, Ramsey was speaking of public discounting.

78 In their work on global climate change, the annual time discount rate chosen by Cline (1992) was 0%; by Nordhaus (1994), 3.5%; by Stern (2006), 0.1%. Stern's 0.1% was attributed by the author to an annual risk of human extinction. Nordhaus' choice of 3.5% a year reflected his reading of consumer behaviour in the United States and the private rate of return on investment there.

To not discount future well-beings would imply that the present-value of returns is infinite. That's because the infinite sum, *r*+*r*+..., is unbounded. We are to understand then that generation-0 incurs a finite loss for a unit of consumption sacrificed, but that the investment yields an infinite gain to future generations no matter how small *r* happens to be. It is obvious the investment should be undertaken, which means consumption should be reduced by generation-0.

Now suppose a similar investment project is also available. The same reasoning says the new project should also be accepted, further reducing current consumption. And so on, until current consumption is reduced to near-zero. Thus *any* consumption sacrifice by generation-0 (short of 100% of available consumption) is good. Many would regard that to be an unacceptable burden on the present generation.

The reasoning goes further. Generation-0, having sacrificed nearly everything for future generations, is followed by generation-1. Imagine that it too is faced with a string of investment projects of the same kind. An identical reasoning would come into play for members of generation-1, who also would be required to consume at a near-zero level. And so on, down the generations. But that means every generation is required to live in penury for the sake of a future that's always just beyond. Koopmans (1965, 1967) likened this consequence of Total Utilitarian reasoning to a "never-ending potlatch," which of course cannot be an optimum policy. But that only shows that Total Utilitarianism is incoherent under the circumstance of the model: there *is* no optimum policy.[79]

The argument extends to models with durable capital goods. There are classroom models in which an optimum saving policy under Total Utilitarianism exists but which requires each generation, even if the early generations are very poor, to save nearly 100% of GDP so as to accumulate wealth for the generations that are to follow (Dasgupta, 2005a, 2007). That's not a never-ending potlatch (despite the high saving rate, consumption grows with time and the generations), but it comes pretty close to it.

The moral would seem to be this: In such complex exercises as those involving the use of resources in a world with a long time horizon, it is foolhardy to regard any value judgement sacrosanct. One can never know in advance what it may run up against. A more judicious tactic than Sidgwick's and Ramsey's would be to play off one set of ethical assumptions against another in not-implausible worlds, see what their implications are for the distribution of well-being across generations, and appeal to our varied ethical senses before arguing over policy. Koopmans (1965, 1967, 1977) promoted this reasoning process when formalising the idea of optimum economic

79 The flavour of the argument remains if the horizon is long but finite. The present-value of a *T*-period consumption flow, *r*, without time discounting is *Tr*. If *T* is large, Total Utilitarianism demands an attenuated form of consumption postponement: An optimum exists, but the overwhelming bulk of consumption is to be enjoyed by the last few generations. In the limit, if $T \rightarrow \infty$, the never-ending potlatch emerges, and there is no optimum policy.

development. Rawls (1972) called the end point of the process, a "reflective equilibrium."

There is another strand to Koopmans' reasoning, which provides a further argument in favour of generation-discounting.[80] If U_t denotes generation-t's well-being, $\{U_0,U_1,...,U_t,...\}$ is an infinite well-being stream. The problem is to compare alternative well-being streams. Let V_0 denote social well-being as perceived by generation-0. In Ramsey's model V_0 is the sum of every generation's well-being (U_1+...+U_t+...). Ethical comparisons of infinite well-being streams in Ramsey's theory are derived from their sums, which, when they exist, are real numbers. But infinite sums don't necessarily exist. That is a problem Ramsey faced squarely when investigating the optimum rate of saving for a nation, and found an ingenious way round it. The second strand in Koopmans' reasoning flips the way deliberations are conducted. The primitive concept there is a complete ordering (henceforth, ordering) of infinite well-being streams by generation-0, not a social well-being function from which an ethical ordering is to be inferred. The tactic is to impose ethical conditions on orderings so as to determine, if possible, the mathematical structure of social well-being V_0.[81]

We say an ordering is "continuous" if, in an appropriate mathematical sense, infinite well-being streams that don't differ much are close to one another in the ordering. We say an ordering is "monotonic" if one stream of well-beings is regarded as better than another if no generation enjoys less well-being along the former than along the latter and if there is at least one generation that enjoys greater well-being in the former than it does in the latter.

Both assumptions are intuitively attractive. Lexicographic orderings notwithstanding, I know of no convincing argument against continuity. Of course Rawls (1972) placed priority rules and the lexicographic orderings that come with them at the centre of his theory of justice, but that's proved to have been one of his most contentious moves. The richness and depth of his analysis would not be lessened if small tradeoffs were admitted between the objects of justice. And it's hard to find reasons against monotonicity. Even Rawls, whose work was so pointed toward distributive justice, insisted on monotonicity.

But it can be shown that orderings over infinite well-being streams that satisfy continuity and monotonicity have generation discounting built into them. It would seem the real numbers are not rich enough to accommodate infinite well-being streams in a manner that respects continuity and monotonicity even while awarding the well-beings of all genenerations equal weights.[82]

I have digressed into generation-discounting because it could be thought that the practice can only be justified by an appeal to agent-centred prerogatives. The reasoning deployed by Tjalling

---

80 The exposition follows Dasgupta and Heal (1979: Ch. 9) and Dasgupta (2005b).

81 Let V be a numerical function on X, which is to say the V-function awards a numerical value to each element of X. We say V is a numerical representation of the ordering R if, for all x, y in X, $V(x) \geq V(y)$ if and only if xRy.

82 The result is in Diamond (1965) and was attributed by the author to Menahem Yaari. It strengthened a theorem on positive time discounting in Koopmans (1960).

Koopmans shows otherwise. So I want to keep the two separate.

9.2 Dynastic Well-Being

It pays to see how the apparatus Koopmans constructed can be used to study dynastic saving, not only national saving. That will ease the way to Generation-Relative Utilitarianism in Actual Problems, as opposed to the Genesis Problem.

Imagine that the economy consists of N identical households. In a stationary state household numbers remain constant over time. Suppose parental well-being is a function of parents' own consumption and the well-being of their children. Then parental well-being can be expressed as $U(C,U)$, where the former U is the well-being of parents, C is the consumption enjoyed by parents, and the latter U is the well-being of their children. But parents know that their children's well-being will in turn depend on the consumption they enjoy and the well-being of their children; and so on down the generations. By recursion we have $U = U(C,U(C,U(...)$. The formula says informed parents include the well-being of all their descendants in their own well-being.

Now suppose the economy faces no risk of extinction. Koopmans (1972) imposed a set of intuitively attractive ethical conditions on the recursive form $U(C,U(C,U(...)$ for parents of generation-0 and showed that it reduces to

$$U(C,U(C,U(...) = U_0 + U_1/(1+\delta)^{-1} + ... + U_t/(1+\delta)^{-t} + ..., \qquad \delta > 0 \qquad (29)$$

Equation (29) says that dynastic well-being as perceived by parents is the discounted sum of theirs and their descendants' personal well-beings, and that the discount rate is a positive constant, δ. Define θ = 1/(1+δ). θ is called the discount factor associated with the discount rate δ. Equation (29) can now be re-written as

$$U(C,U(C,U(...) = U_0 + \theta^{-1}U_1 + ... + \theta^{-t}U_t + ..., \qquad 0 < \theta \leq 1 \qquad (30)$$

The subject of discounting for time and the generations has not been confined to Utilitarian thinkers. Rawls in his contractual theory of justice spoke to intergenerational saving and Sidgwick's arguments against discounting (Rawls, 1972: Ch. V, Sects. 44-45). It will take us afield in the text of this essay to present his formulation of just saving. An account of the ties that bind the generations in Rawls' work is presented in Appendix 3.

9.3 Overlapping Generations

Because I am adopting Generation-Relative Utilitarianism in this essay and trying it for size, I eschew generation-discounting merely on grounds of futurity in the remainder of the essay. I also assume that generations overlap in time. That connects with the world as we know it; in particular, to the empirical observations in Section 1.

The horizon is taken to be "conditionally infinite", by which we mean there is a positive risk of extinction of the human race at t should it have survived until then. The risk of extinction provides a reason for discounting future well-beings conditional on humanity having survived until then (that's the smaller weight to be awarded to a future generation relative to one's own). The reason is that the risk increases the farther we peer into the future. Sidgwick, in the passage quoted previously, recognised this.[83]

People live for two periods. In their first period (childhood) they are maintained by their parents. At the beginning of their second period (adulthood), should humanity survive until then, they choose how many children to have and how to share the output they produce. The way I have phrased matters could suggest that the society we are studying is authoritarian and centralized. It isn't. Rather, we suppose as previously that people of the same generation are identical in all respects. That means adults in any period have the same ethical motivation. So, even though it is individuals who do the choosing, it is as though each generation chooses the size of the next generation. Normative restrictions arising from agent-centred prerogatives translate here into generation-centred prerogatives.

People die at the end of their second period, which means there are two overlapping generations in any period. That isn't good demography, but it turns out not to matter. To include realistic life cycles would not add to the substance of population ethics, it would only add to the notation. Let $N(t)$ be the number of adults at t, should the world have survived until then. I will call them "generation-t". We now conduct backward induction.

Denote the number of children born at t by $N(t+1)$. Total population is therefore $N(t)+N(t+1)$. Only adults work, children are pure consumers. But adults and children have the same well-being function, $U(C)$.[84] That means total output at t is $Q(t) = AF(K,N(t))$ and consumption per capita is $Q(t)/(N(t)+N(t+1))$. The transfer of output is therefore from the old to the young (Lee, 2000, 2007).

We will study the economy in a stationary mode. K can then be interpreted to be an aggregate measure of all capital assets other than labour. It includes not only natural capital, but also produced capital such as roads, buildings, and machines. Appendix 4 shows how to break away from assuming an aggregate measure of produced and natural capital; but because I want to stress the significance of the constraints Nature imposes on population axiology, I once again refer to K as the biosphere.

Because marginal U declines with C, the allocation rule in each period is simple: everyone gets to consume the average. In period $t+1$, assuming humanity survives, the $N(t+1)$ children of the previous generation are adults (they constitute generation-$(t+1)$); and if they give birth to $N(t+2)$ children, they produce $Q(t+1)$ and share $Q(t+1)$ equally. And so on.

The probability that humanity will suffer extinction at t conditional on having survived until then is called the "hazard rate" at t. We assume that the hazard rate is a constant, $\delta > 0$. As noted in Section 9.2, δ functions like a "time discount rate." Define $\theta = 1/(1+\delta)$. θ assumes the role of a "time discount factor." The random process defined by δ should be interpreted as governing the possibility of a global catastrophe, such as the impact of a massive meteorite. We assume that other than humanity's survival, there is no future uncertainty. In Part II we consider catastrophic risks that are

83 Lind (1982) is an early collection of essays on discounting for time and risk.

84 To assume otherwise would be to add further notation.

brought about by our own activities.[85]

That the hazard rate is constant is not believable. We are reliably informed that Sun will be a hospitable star only for a finite length of time, which means the hazard rate we should work with remains small for a long while but will rise toward infinity as Sun runs out of sufficient fuel to serve life's needs. It can be shown though that optimal policies for a considerable number of periods into the future are insensitive to whether the hazard rate is a small positive constant or whether it is a small positive constant for a long stretch of time but then rises sharply toward infinity. But that is another way of saying that to model the extinction process in terms of a small constant hazard rate is not misleading, at least not for a long stretch of time.[86]

Total Utilitarianism regards social well-being (that is, well-being across the generations) to be the expected sum of each person's well-being from the date of evaluation, t, to infinity. Denote that by $V^U(t)$. As $N(t)+N(t+1)$ is the number of people alive at t and $U(C(t)) = U(Q(t)/(N(t)+N(t+1)))$ is the flow of well-being per person at t,

$$V^U(t) = {}_{u=t}\sum^{\infty}[(N(u)+N(u+1))U(C(u))]\theta^{(u-t)} \tag{31}$$

("${}_{u=t}\sum^{\infty}$" denotes summation as u goes from t to infinity.) Equation (31) is the Total Utilitarian DM's conception of social well-being at t.

The uncanny similarity in discounting between equations (30) and (31) is inescapable. The discounting is identical in form, but its content is different. In Koopmans' representation, θ is to be interpreted as the discount factor for time. I touched on this in Section 9.2. However, in contrast to Koopmans' work I have eschewed discounting for time in equation (31) and have invoked discounting for the risk of extinction. The equation of course can't tell the difference between the two. Which is why equation (31) is consistent with Koopmans' axioms.

Among the many reasons Koopmans' work on the representation of social well-being functions is of huge significance is that it has taught us that the mathematical structure of Total Utilitarianism can be supported by more than one interpretation of its contents. With suitable adjustment of Koopmans' axioms on the V-function, we can even arrive at Generation-Relative Utilitarianism (eq. (32) below).

9.4 Generation-Relative Utilitarianism

To formalize Generation-Relative Utilitarianism, let $V(t)$ be the social valuation function of a representative member of generation-t. In our model there are no actual future people for her to consider, all future people are potential. Generation-centred prerogatives are the reason each generation awards a lower weight to potential well-beings than to their own well-being. So we have

$$V(t) = N(t)U(C(t)) + \mu N(t+1)U(C(t)) + \mu\{{}_{u=t+1}\sum^{\infty}[\theta^{(u-t)}[(N(u)+N(u+1))U(C(u))]\},$$

85 Bostrom (2002) and Rees (2003) discuss the possible causes of such a global disaster.

86 In his study of the optimum price of annuities, Yaari (1965) constructed a model of individuals facing the hazard of death in very much the way we are modelling the extinction process.

$$0 < \mu, \theta < 1; \qquad t = 0,1,... \tag{32}$$

$V(t)$ is generation-t's social valuation function. $V(t) = V^U(t)$ if $\mu = 1$.[87]

Agent-relative theories could appear to advance self-indulgence; worse, callousness toward others. So it is useful to note that in equation (32) the representative member of generation-*t* advances personal projects that are especially meaningful to her by assigning a smaller weight to the potential well-beings of her descendants. To be sure, the smaller weight serves to raise her personal well-being, but it does so only by lowering the number of new births (that's confirmed below), not by discriminating against the interests of others. The backward-induction we have used to arrive at equation (32) formalizes the intention of the potential parent to allocate resources between herself and her children in such a way (complete equality there) that they too will be able to advance projects that are especially meaningful to them.

$V(t)$ guides the choices of generation-*t*, and does so for all $t \geq 0$. Our analysis starts at $t = 0$. Because the economy has a stationary structure, the generation at any t can reason that future generations will have a motivation like their own. Generation-0 is $N(0)$ in number; it is a datum. If previous generations hadn't been moved by Generation-Relative Utilitarianism, $N(0)$ would not have been chosen with that motivation. No doubt generation-0 will want to ask what their ancestors' intentions had been; they may even feel obliged to choose so as not to be entirely at odds with them, but such considerations would be a distraction here. So I assume generation-0 regards the past as past, and is entirely forward looking. I imagine also that all generations starting at $t = 0$ adhere to Generation-Relative Utilitarianism. In that case the motivation embodied in the notion of expected well-being across the generations in equation (32) binds members of each generation to all their descendants.

9.5 Intergenerational Consistency

The preference reversal displayed in inequalities (26)-(28) translates here into inconsistent ethical values across the generations. The analysis that follows is informal. Beyond supposing that $N(0)$ is small relative to K (otherwise generation-0 will choose not to reproduce), I avoid technicalities.

Consider an arbitrary t ($t \geq 0$). The size of generation-*t* will have been chosen by the previous generation. That means $N(t)$ will not be subject to choice at t. But generation-*t* chooses $N(t+1)$, which is the size of the generation that follows. And so on. Why doesn't generation-*t* choose the size of all subsequent generations? It has a conception of well-being across the generations, which is $V(t)$; so

87 Almost all formal models of economic growth and development assume that population size is exogenously given; that is, all future people are actual future people. They give rise to what Parfit (1984) called Same Numbers Choices. Generation-Relative Utilitarianism in a world with a mixture of actual future people and potential people comes in two forms. In the Weak version present people award the same weight to actual future people but a lower weight to potential people; in the Strong version all future people, whether actual or potential, are awarded a lower weight.

why not choose the size of all subsequent generations?

There are two reasons; one practical, the other ethical. The practical problem has to do with implementation. Generation-t has no viable means of enforcing its will on future generations. Moreover, later generations will have every incentive not to reproduce in accordance with generation-t's bidding. Because $\mu < 1$, their grounds of binding reason will differ from that of generation-t, and in turn from those of subsequent generations. That's intergenerational inconsistency.[88]

Turning to the ethical reason, there is the deontological requirement that generation-t will have no right to enforce its will on future generations even if a method was found to enforce it. Taken together the two considerations imply that each generation chooses the size of the generation that is to follow by reasoning progressively into the future. An optimum demographic profile would then be self-fulfilling. We now formalise the idea.

Consider generation-t. Its evaluation of well-being across the generations is $V(t)$. We start with generation-0. An ideal population profile under Generation-Relative Utilitarianism is a stream of reproductive choices $\{N^*(1),...,N^*(\mathrm{t}), N^*(t+1),...\}$ such that, for all $t \geq 0$, $N^*(t+1)$ maximizes $V(t)$ on the supposition that all generations following t (should humanity survive until then) will choose in accordance with that profile.[89]

The common supposition has the status of an implicit understanding. Because each generation has an understanding of the choices its descendants will face conditional on humanity surviving, it can anticipate the choices they will make. Its own choice is based on a forecast of its descendants' choices. The exercise need not be exclusively forward looking. Each generation in a sequence of Generation-Relative Utilitarians may wish to look back at the choices its ancestors had made, if only to reassure itself of the implicit understanding and to confirm that its own choice is guided by it.[90] The birth profile $\{N^*(1),...,N^*(t+1),...\}$ is intergenerationally consistent, even though the ethical motivations of the generations are not consistent with one another. Formally, the birth profile $\{N^*(1),...,N^*(t+1),...\}$ is a Nash equilibrium across the generations.

9.6 The Stationary Optimum

Characterizing the implicit understanding for arbitrary values of $N(0)$ is hard work. So it pays to study population and consumption profiles that are stationary. It is reasonable to conjecture that if the U-function is the one in equation (5), the stationary profile (we confirm presently that it is unique) is dynamically stable, meaning that the optimum population profile from any $N(0)$ near the stationary value converges to it. Very rough calculations suggest that the conjecture is correct, but I haven't been

---

88 The rate at which generation-t would be willing to trade the potential well-being of generation-(t+1) against the potential well-being of generation-(t+2) differs from the tradeoff generation-(t+1) would be willing to make between their own well-being and the potential well-being of generation-(t+2).

89 The corresponding consumption profile, $\{C^*(1),...,C^*(t),...\}$ is implied by the condition that at each t, $C^*(t) = AF(K,N^*(t))/(N^*(t)+N^*(t+1)))$.

90 I return to this point in Appendix 3 when exploring Rawls' principle of just saving.

able to nail it down. But dynamic stability of non-stationary profiles is the reason we should be interested in the stationary profile. In a stationary profile the generations replicate their numbers, conditional on the world surviving of course. Let N^* be that number.

The idea underlying the profile $\{N^*,N^*,...\}$ can be expressed in the following way: If $N(0)$ just happens to equal N^*, then N^* would be the reproductive choice of generation-t if it were to suppose that all subsequent generations will in turn choose N^*. It follows that N^* is a self-enforcing reproductive policy for every generation. The understanding that implements the policy is only implicit; but because N^* is a self-enforcing choice for each generation, there is no call for a constitutional directive. We now study the condition that the profile $\{N^*,N^*,... \}$ must necessarily satisfy.

In each period there are N^* adults and N^* children, making a total population of $2N^*$. Only adults work. So output $Q = Q^* = AF(K,N^*)$. As both adults and children consume the same amount, each person will enjoy $Q^*/2N^* = AF(K,N^*)/2N^*$ units of consumption. Write that as C^*.

Return once again to the laboratory we have been using, consisting of equations (4)-(5). It is simple to confirm that the optimum stationary profile satisfies

$$C^*/C^S = (N_S/N^*)^{(1-\rho)} = (1+\sigma P)^{1/\sigma},$$

$$P = (1+\mu+2\mu\theta(1-2\rho))/2\mu(1+\theta) \qquad (33)$$

Equation (33) is the counterpart of the Sidgwick-Meade Rule in the stationary state commended by Generation-Relative Utilitariansim. In Part II we study the sensitivity of the optimum values of consumption and population to the various parameters in equation (33), ethical presumptions (C^S), and the productivity of capital, $AK^{(1-\rho)}$.[91]

9.7 Extreme Theories

Total Utilitarianism represents one extreme set of values. The theory says $\mu = 1$. At the other extreme is a theory that is easy to describe but is also questionable. It says $\mu = 0$. That viewpoint was explored by Enke (1966) in his study of social cost-benefit analysis of family planning programmes in poor countries. Enke sought ways to measure the economic value of prevented births, and he took the

---

91 To obtain equation (33), note that generation-t has direct control over $N(t+1)$, nothing else. But $N(t+1)$ appears only in the first four terms of the infinite series in equation (32). Differentiate those four terms partially with respect to $N(t+1)$ and set the sum of the partial derivatives equal to zero. Now set $N(t) = N(t+1) = N(t+2)$. As this operation is to be conducted recursively by all generations, we obtain the common value, N^*, say. But $C = AK^{1-\rho}N^{*(\rho-1)}/2$. Equation (33) follows.

As an exercise, suppose we apply Classical Utilitarianism to equation (33). With $\mu = 1$, $\rho = 0.5$, $\sigma = 1$, and θ very slightly below 1, $C^*/C^S = 1.5$. As another exercise, let $\mu = 1$, θ very slightly below 1, and $\rho = 0$. The latter implies there is no production, that manna can be consumed directly. In that case equation (33) and equation (9) become identical.

measure to be the discounted sum of the differences between an additional person's consumption and output over his lifetime. Children in Enke's theory have value only if they pay their way.

Enke's is an extreme point of view, as is Total Utilitarianism. Generation-Relative Utilitarianism lies between two extremes and reflects the strength of each without giving in to the weaknesses of either. It prescribes neither a large population nor a small population. Instead, it offers a wide space in between, within which more detailed ethical considerations can be embedded. We should not expect an ethical theory to do more.

Part II

Applications

10 The Biosphere as a Renewable Natural Resource

Equation (33) gives the ratio of Earth's (human) carrying capacity to the optimum population size, but it doesn't say what the carrying capacity is. No doubt "carrying capacity" is a range of values, not a precise figure; but it helps to use formal, deterministic models so as to think of it as a specific number. Once we have done that, we can loosen the model to construct a range round the figure.[92] To obtain estimates, I first collate salient findings on the current state of the biosphere. I then use estimates of our aggregate demand for goods and services from the biosphere relative to its ability to supply them on a sustainable basis. The procedure won't allow us to nail down Earth's human carrying capacity, even as a range, but it's about as close we can get to it with the information I have been able to obtain.[93]

10.1 Open-Access Resources

Mutually-adverse environmental externalities are of an acute form when, as is the case with much of the biosphere, Nature is an open-access resource; meaning that it is free to all to do as we like with it. Gordon (1954: 135) famously wrote that a resource that is everybody's property is nobody's property. Using the example of fisheries, he showed that the scramble to extract open-access resources dissipates all potential rents from them; and he used the dissipation of rents to explain why fishermen working on open waters are usually poor. Hardin (1968) even more famously called the rent dissipation associated with open-access resources, "the tragedy of the commons."

Gordon's and Hardin's analyses were limited to timeless worlds. A complete analysis of the problem of the commons in an economy moving through time, including a proof that population size is a factor influencing whether open access resources *do* give rise to the tragedy of the commons, has recently been offered by Dasgupta, Mitra, and Sorger (2018). Population was taken to be exogenously given. In a canonical setting the authors showed that open-access resources suffer from the tragedy of the commons only when human population is large, other things equal. Property rights to resources

92 I am using Earth and the biosphere interchangeably.

93 Cohen (1995a-b) reports methods that have been deployed to estimate the biosphere's human carrying capacity. The range he arrived at was wide, but Cohen didn't have at his disposal the ominous environmental signatures that have been uncovered in recent years.

don't matter in a thinly populated world, they matter only when population is large relative to the size of the resource base. The finding is used in Section 12 to read the experiences of some past societies.

To the best of my knowledge, we do not yet have a theoretical account of reproductive behaviour in an intertemporal economy containing open-access resources. In Appendix 2 I use the timeless model in Section 8 to confirm that open access to natural resources offer incentives to households to have too many children. In such a property-rights regime households of greater size are permitted to access a larger fraction of the resource base. They respond by converting the commons excessively into private assets - in our model, household size. The analysis is more restrictive than I would ideally like it to be. Being timeless, the model takes Nature to be a fixed factor in production. The biosphere is of course nothing like that.

10.2 Ecosystem Services

The term "biosphere" is an all-encompassing construct. We may think of it as a mosaic of renewable natural resources, not a fixed factor in production. There is more to the biosphere than that (technically, it's part of Earth occupied by living organisms), but it is a useful reduction for our purposes. Agricultural land, forests, watersheds, fisheries, fresh water sources, estuaries, wetlands, the oceans, and the atmosphere are interlocking constituents of the biosphere. I shall refer to them generically as "ecosystems" and, so as to draw attention to populations of species in their habitats, I shall speak of them also, more narrowly, as "biological communities."[94]

Ecosystems combine the abiotic environment with biological communities (plants, animals, fungi, and microorganisms) that form functional units. Individual actors in ecosystems include organisms that pollinate, decompose, filter, transport, re-distribute, scavenge, fix gases, and so on. Nearly all organisms that help to produce those services are hidden from view (a gram of soil may contain as many as 10 billion bacterial cells), which is why they are almost always missing from popular discourses on the environment. But their activities enable ecosystems to maintain a genetic library, preserve and regenerate soil, fix nitrogen and carbon, recycle nutrients, control floods, mitigate droughts, filter pollutants, assimilate waste, pollinate crops, operate the hydrological cycle, and maintain the gaseous composition of the atmosphere.[95]

In a path-breaking set of publications assessing the state of the world's ecosystems, MEA (2005a-d) constructed a four-way classification of goods and services we enjoy from them: (i) provisioning services (food, fibre, fuel, fresh water); (ii) regulating services (protection against natural hazards such as storms; the climate system); (iii) supporting services (nutrient cycling, soil production); and (iv) cultural services (recreation, cultural landscapes, aesthetic or spiritual

94 Tilman (1982) remains an excellent introduction to the processes by which competition among organisms for resources gives rise to the structure of biological communities.

95 Daily (1997) is an early publication of the role ecosystem services play in our lives. Perrings (2014) is a treatise-length account of the economic value of biodversity.

experiences). Cultural services and a variety of regulating services (such as disease regulation) contribute directly to human well-being, whereas others (soil production) contribute indirectly (by providing the means of growing food crops).

Ecosystems differ in composition and extent. They can be defined as ranging from the communities and interactions of organisms in your mouth or those in the canopy of a rain forest to all those in Earth's oceans. The processes governing them differ in speed. There are systems that turn over in minutes, there are others whose rhythmic time extends to hundreds of years. Some ecosystems are extensive ("biomes", such as the African Savannah), there are those that cover regions (river basins), many involve clusters of villages (micro-watersheds), while others are confined to the level of a single village (the village pond). In each example there is an element of indivisibility. Divide an ecosystem into parts by creating barriers, and the sum of the productivity of the parts will typically be found to be lower than the productivity of the whole, other things equal (Loreau et al., 2001; Worm et al., 2006; Sodhi, Brook, and Bradshaw, 2010). Mobility of biological populations is a reason (Sect. 4.2). Safe corridors, for example, enable migratory species to survive.

Ecosystems can regenerate but suffer deterioration (worse, exhaustion) when human expropriation exceeds the rates at which they are able to supply us with goods and services or when they are converted directly into produced capital; which is why they can be thought of as renewable natural resources. Population extinctions disrupt essential ecosystem services. In tropical forests for example, dung beetles play an essential role in recycling nutrients. Excessive hunting of mammals in the forests has been found to be a cause of local elimination of dung dependent beetles (Brook, Sodhi, and Bradshaw, 2008). When subject to excessive stress, once flourishing ecosystems (e.g. biologically rich estuaries) flip into unproductive states (dead zones). The stress could be occasioned by an invasion of foreign species or substance (as in the above example), it could be owing to loss of population diversity (see below), it could be triggered by the demise of a dominant species (see below), and so on. Ehrlich and Ehrlich (1981) likened the pathways by which an ecosystem can be tipped out of a stable regime into an unproductive state to a flying aircraft from which rivets are removed, one at a time. The probability that it will crash increases very slowly at first, but then at some unspecifiable number rises sharply to 1. In Appendix 4 we review the dynamics of renewable natural resources.

Broadly speaking, an ecosystem's productivity (the flow of goods and services they offer us and their ability to withstand shocks) increases with the diversity of the functional characteristics of its species populations; mere headcount of species can mislead.[96] Mutual dependence among the species is a reason. For example, many trees produce large, lipid rich fruits that are adapted for animal dispersal, which means the demise of fruit eating birds can have serious consequences for forest

96 Hooper et al. (2005) is a clear review of the issues. A classic on the subject is Tilman and Downing (1994). See also Tilman (1997), Hector et al. (1999), and Walker, Kinzig, and Langridge (1999).

regeneration. Relatedly, in a study in Costa Rica, Ricketts et al. (2004) found that coffee yield declined with distance from the forest edge because forest bees aid pollination. The authors reported that the bees increased coffee yield by 20% in fields within 1 km from the forest edge. Looking elsewhere, about one-third of the human diet in tropical countries is derived from insect-pollinated plants. But that means a decline in forest dwelling insects has an adverse effect on human nutrition. And so on.

In food webs the relationships are uni-directional. Primary producers in the oceans (phytoplanktons, sea weeds) are at the bottom of the food chain, while species at higher trophic levels consume those that are below. Species whose impact on a community structure is large and disproportionately large relative to their abundance are called "keystone" (Power et al., 1996). They are usually at the top end of the food chain. When human consumption reduces their populations sizeably, the community flips to a different state, as prey populations explode. That reduces diversity of functional characteristics.

Biological communities can influence their abiotic environment. The Amazon for example generates about half of its own rainfall by recycling moisture 5-6 times as air-masses move from the Atlantic across the basin to the west. Mathematicians call this a positive feedback. Deforestation of the Amazon would be expected to reduce rainfall and to lengthier dry season in the region. One estimate has it that 20-25% deforestation of the Amazon can be expected to flip the forests in the east to savanna vegetation (Lovejoy and Nobre, 2018). Palm oil trees are planted increasingly in the Amazon so as to provide substitutes for fossil fuels, only to contribute to a sharp decline in the ability of the Amazon forest to absorb carbon dioxide from the atmosphere. The irony will not escape readers.

The view that the biosphere is a mosaic of renewable natural resources also covers its role as a sink for pollution. As noted in Section 3.1, pollutants are the reverse of natural resources. One way to conceptualize pollution is to view it as the depreciation of capital assets. "Resources" are "goods," while "pollutants" (the degrader of resources) are "bads." Polluting is the reverse of conserving.

In Part I it was assumed that the biosphere (K) remains unchanged. That enabled us to model the global economy in a stationary mode, but it misleads when we move to study imperfect dynamic economies. Ecosystems decline in quality or quantity (or both) if our demand for their goods and services and the rate at which they are transformed directly into produced capital (roads and buildings) exceed the rate at which they are able to provide a sustainable supply of those goods (collapse of fisheries). By the same token, restoration and conservation measures (e.g., creating protected areas for marine fisheries) help to increase the biosphere's productivity, as measured by quality or quantity (or both).

10.3 Erosion of Natural Capital

Humanity's success in the Modern Era (post 1500 CE) in raising the standard of living has in great measure involved mining and degrading the biosphere. Habitat destruction caused by rising demand for Nature's products are the proximate causes of the decline in the biosphere's ablity to supply our needs on a sustainable basis. The conversion of land for the production of food, live stock, and plantation crops is a prime cause of that decline. Conversion of land into produced capital (buildings, roads) is another cause.[97]

Erosion of natural capital usually goes unrecorded in official economic statistics because GDP does not record depreciation of capital assets. Destroy an open woodland so as to build a shopping mall, and the national accounts will record the increase in produced capital (the shopping mall is an investment), but not the disinvestment in natural capital. The example is a commonplace. Even while industrial output increased by a multiple of 40 during the 20th century, the use of energy increased by a multiple of 16, methane-producing cattle population grew in pace with human population, fish catch increased by a multiple of 35 and carbon and sulphur dioxide emissions rose by more than 10. It has been estimated that 25-30 per cent of the 130 billion metric tons of carbon that are harnessed annually by terrestrial photosynthesis is appropriated for human use (Haberl et al., 2007). Although the rise in the concentration of atmospheric carbon receives much the greater public attention, MEA (2005a-d) reported that 15 of the 24 ecosystems the authors had reviewed world-wide were either degraded or are being exploited at unsustainable rates.

Current extinction rates of species in various orders have been estimated to be 10-1,000 times higher than their average rate (about 1 per million species per year) over the past several million years (Sodhi, Brook, and Bradshaw, 2009; Pimm et al. 2014; Ceballos, Ehrlich, and Ehrlich, 2015). The figures are reached from field studies of the decline in numbers of specific groups of mammals, insects, and birds, and from empirically drawn relationships between the number of species in an area and the size of the area. But the relationships are known to vary substantially among communities and habitats, which is why, as the range shows, there are great uncertainties in the estimates. Despite the uncertainties, the figures put the scale of humanity's presence on the Earth system in perspective (Ehrlich and Ehrlich, 2008) and explain why our current times have been recognised as the start of a new epoch, the Anthropocene.[98]

The statistics I have just summarised for sketching humanity's recent doings differs sharply from the one that has been on offer in a string of recent books, in which intellectuals have re-drawn

97 For an analytical framework that maps the processes by which habitat destruction leads to species extinction, see Tilman and Lehman (1997).

98 The term "Anthropocene" was popularized by Crutzen and Stoermer (2000) to mark a new epoch that began with the Industrial Revolution some 250 years ago. Wilson (1992) is a classic on the value of biodiversity. Colbert (2013) is an outstandingly readable account of how stratigraphers uncover geological signatures of abundance and disappearance of species in the distant past. Dasgupta, Raven, and McIvor (2018) is a collection of essays, presented at a joint meeting of the Pontifical Academies of Science and Social Sciences, respectively, on contemporary biological extinctions, ranging from mammals and birds to micro-organisms in the soil.

our attention to the remarkable gains in the standard of living humanity has enjoyed over the past century (Micklethwait and Wooldridge, 2000; Ridley, 2010; Lomborg, 2014; Norberg, 2016; Pinker, 2018). The authors have collated data on growth in scientific knowledge and the accumulation of produced capital and human capital and argued that humanity has never had it so good. But with the exception of rising carbon concentration in the atmosphere, trends in the state of the biosphere accompanying those advances have gone unnoted by the authors. The problem is, global climate change is but one of a myriad of environmental problems we face today. And because it is amenable to technological solutions (innovating with cheap non-carbon sources of energy and, more speculatively, firing sulphur particulates into the stratosphere to reflect sunlight back to space; Pinker, 2018), it is not representative. Global climate change attracts attention among intellectuals and the reading public not only because it is a grave problem, but also because it is possible to imagine meeting it by using the familiar economics of commodity taxation, regulation, and resource pricing without having to forego growth in living standards in rich countries. The literature on the economics of climate change (e.g. Stern, 2006; Lomborg, 2014) has even encouraged the thought that with but little investment in clean energy sources (say 2% of world GDP) we can enjoy indefinite growth in the world's output of final goods and services (global GDP).

And that's a thought to be resisted. **Our economic possibilities are circumscribed, even if several steps removed via technological progress, by the Earth-system's workings. We should ask whether the biosphere could support on a sustainable basis a global population of 9.5-13 billions (which is the error bar round the UN Population Division's population projection of 11.2 billion for year 2100) at the standard of living we are encouraged to seek. In effect we are asked to imagine such population numbers enjoying a living standard that makes a smaller demand on natural capital than we currently make on it and is at the same time one of high quality. I know of no study on the economics of technological change that has explored that question, let alone the question of what lifestyles that would involve. As of now we should be circumspect that the scenario is plausible, because at least as grave a danger facing humanity as global climate change** is the unprecedented rate of biological extinctions now taking place. **Climate change contributes to the extinction process, but it is only one among several major factors that are contributing to it. Other factors include the habitat destruction that accompanies agricultural extension and intensification, deforestation, and the direct transformation of natural capital into produced capital (e.g. roads, dams, ports).** Continued extinctions will damage the biosphere irreparably, **involving unknown numbers of tipping points, which should tell us** that **potential cascades** cannot be **staved off** by **mere** technological fixes. Politics has intervened to prevent even the relatively small global investment that **economic** experts **only a few years ago suggested was required** to **contain** climate change. So we should expect the problem of biological extinctions to remain off the table, at least until citizens take the matter seriously.[99]

10.4 Global Ecological Footprint

Studying biogeochemical signatures of the past 11,000 years, Waters et al. (2016) tracked the human-induced evolution of soil nitrogen and phosphorus inventories in sediments and ice. The authors reported that the now-famous figure of the hockey stick that characterises time series of carbon concentration in the atmosphere are also displayed by time series of a broad class of global geochemical signatures. They display a flat trend over millenia until some 250 years ago, when they begin a slow increase which continues until the middle of the 20th Century, when they show a sharp and continuing rise. Waters et al. (2016) proposed that mid-20th Century should be regarded as the time we entered the Anthropocene.[100]

Their reading is consistent with macroeconomic statistics. World population in 1950 was about 2.5 billion and global output of final goods and services a bit over 8.7 trillion international dollars (at 2011 prices). The average person in the world was poor (annual income was **only a bit** in excess of 3,500 international dollars). Since then the world has prospered beyond recognition. Life expectancy at birth in 1950 was 45, today it is a little over 70. Population has grown to over 7.5 billion and world output of final goods and services is (at 2011 prices) above 110 trillion international dollars, meaning that world income per capita is now more than 15,000 international dollars. A somewhat more than 12-fold increase in global output in a 65-year period helps to explain not only the stresses to the Earth system that we have just reviewed, but it also hints at the possibility that humanity's demand for the biosphere's services has for several decades exceeded sustainable levels.[101]

In a review of the state of the biosphere, WWF (2008) reported that although the global demand for ecological services in the 1960s was less than supply, it exceeded supply in the early years of the present century by 50 per cent. The figure is based on the idea of "global ecological footprint," which is the surface area of biologically productive land and sea needed to supply the resources we consume (food, fibres, wood, water) and to assimilate the waste we produce (materials, gases). **The overshoot in our ecological footprint thus includes the overshoot in our carbon emissions into the atmosphere**. The Global Footprint Network (GFN) updates its estimates of the global ecological footprint on a regular basis. A footprint in excess of 1 means demand for ecological services exceeds their supply. By GFN's reckoning, maintaining the world's average living standard at the level reached

---

99 The really hard problem in the political economy of global climate change involves using the latter's special features to frame the way we should explore the prospects for international agreements. Barrett (2003, 2012) and Barrett and Dannenberg (2012, 2014) are incisive analytical and empirical studies on this.

100 The Anthropocene Working Group has proposed that the immediate post-war years should be regarded as the start of the Anthropocene. See Vosen (2016).

101 This is a more accurate estimate of growth in world income over the period 1950-2015 than the one in Dasgupta and Dasgupta (2017). It has been taken from Barrett et al. (2018).

some ten years ago (roughly 12,000 international dollars) would have required 1.5 Earths.[102]

These are *really* crude estimates, and I feel nervous using them. Figures for such socio-economic indicators as GDP, population size, life expectancy, and adult literacy are reached by a multitude of national and global institutions, who exchange information and coordinate their work. They are rehearsed regularly and governments and international agencies use them routinely when advocating and devising policy. We all take note of their figures and trust them. In contrast, I am obliged here to rely on the estimates of a solitary research group (GFN), albeit one with a network of collaborators. Most people will look askance at their estimates. What matters though is not the exact figure but whether the footprint exceeds 1. On that there should be little question. That there is an overshoot in global demand for the biosphere's goods and services is entirely consistent with a wide range of evidence on the state of the biosphere, some of which I have reviewed here. As the estimates from GFN are the only ones on offer, I make use of them.

GFN's most recent estimate of the global ecological footprint is 1.6. Sustainable development would require that the footprint must on average equal 1 over time. Global demand for ecological services can exceed supply for a period, but not indefinitely (Appendix 4). Economic development during the past 65 years has raised the average living standard beyond recognition even while population has increased by an unprecedented amount; but we have enjoyed that success by leaving a substantially diminished biosphere for future generations. It would appear we are living at once in the best of times and the worst of times.

11 Estimates of Globally Optimum Population

Equation (33) expresses C^*/C^S and N^S/N^* in terms of four parameters: μ, θ, ρ, and σ. So as to appreciate what the numerical exercise on Generation-Relative Utilitarianism didn't display in Section 9, it helps to de-couple the equation and express it as

$$C^* = C^S(1+\sigma P)^{1/\sigma} \tag{34a}$$

$$N^S = N^*(1+\sigma P)^{1/\sigma(1-\rho)}, \tag{34b}$$

$$P = (1+\mu+2\mu\theta(1-2\rho))/2\mu(1+\theta) \tag{35}$$

In our model people live for two periods, but no mention has been made of the number of years that add up to make a period. I need to do that now because we will be applying annual global data to the model and because people live for more than two years. So suppose each period is of T years. That means people live for $2T$ years, the first T years as children (non-workers) and the latter T years as adults (workers). People reproduce the moment they enter adulthood.

We now imagine that equation (32) is the social valuation function of the member of a dynasty (Sect. 9.4). Some dynasties have an identical age structure, others have a different age

102 For pioneering work on the idea of ecological footprint, see Rees and Wackernagel (1994) and Rees (2006). See also Kitzes et al. (2008). Wakernagel, who founded the Global Footprint Network (www.footprintnetwork.org/public), was a lead author of WWF (2008).

structure. For reasons that will become clear presently, suppose there are T sets of dynasties, with the property that those in any given set have the same age structure, but that dynasties not in the same set differ in their age structure. For example, in any calender year, τ, there would be a set of dynasties in which the adults are all T+3 years old and the young are all 3 years old. And so on. That means dynasties in adjacent sets have age structures that are a year apart. There will be an adjacent set in which adults in every dynasty is T+4 years old and the young are 4 years old. And so forth.

The sets are identical in size, say M. That will be shown presently to be a requirement of stationary modes. M is determined within the model. If an economy governed by Generation-Relative Utilitarianism is in a stationary mode, we would have $M = M^*$, where $M^*T = N^*$. Similarly, if the economy's population is the biosphere's human carrying capacity $2N^S$, the number of dynasties in each set would be M^S, where $M^ST = N^S$. Equations (34a-b) can now be used to solve for M^S and M^* in terms of N^S and N^*, respectively.

Equations (34a-b)-(35) represent the stationary state commended by Generation-Relative Utilitarianism in what may be called a "reduced form." They do not display the demographic structure of the population, but they are sufficient for us to determine optimum population and the optimum living standard. We now apply annual global data to estimate, very crudely, the optimum stationary state.

11.1 Earth's Human Carrying Capacity

Equation (35) says μ, θ, and ρ determine P, while equation (34a) says P, σ and C^S (well-being subsistence) determine the optimum living standard C^*. There is no need to consult environmental scientists if we want to estimate C^*.

But when it comes to estimating N^*, even moral philosophers need environmental scientists. The reason is equation (34b), which says P, σ, and ρ determine N^*/N^S. The problem is, the equation doesn't yield N^*. To determine N^* we need an estimate of N^S, and to obtain that we need data on the biosphere's productivity and its regenerative possibilities. The analysis in Appendix 4 tells us that the biosphere can support a range of human population numbers on a sustainable basis depending on the standard of living we want to enjoy. Well-being subsistence is a minimum living standard we should want to consider. That is why in Section 5 we defined the human carrying capacity as the largest population the biosphere can support on a sustained basis at well-being subsistence. But without consulting the environmental sciences we can't tell what that population number is. And without that we wouldn't be able to tell what N^* is.

As of now we have little quantitative knowledge of the biosphere's dynamics. But we know that expanding our stock of produced capital is likely to have environmental consequences. So, with both hands proverbially tied behind my back I regard K to be an aggregate measure of the biosphere and produced capital. Next, I stop both A and K on their tracks and estimate $AK^{(1-\rho)}$ (eq. (4)) on the basis of figures for the global ecological footprint, the current size of the world economy, and our model of global production (eq. (4)). The estimate of $AK^{(1-\rho)}$ includes the social value not only of the biosphere, but also of produced capital, social institutions, and public knowledge. In short, I use equation (4) to estimate the social worth of all capital assets with the exception of labour. That enables me to estimate N^S. I realize that's applying an intellectual sledge-hammer to a delicate problem (even if humanity were to disappear from the face of the earth, the biosphere's dynamics would be shaped by the human imprint of the past), but I have found no other way to get at N^S.[103]

The data being utterly crude, I confine myself to pen-on-paper computations. I assume that the value of the world's production of final good and services draws proportionately on ecosystem services at all levels.[104] World output is currently about 110 trillion international dollars. Using the model of production in equation (4), we therefore have

$$AK^{1-\rho}N^{\rho} = 110 \text{ trillion (international) dollars} \quad (36)$$

For ease of pen-on-paper computation, I assume world population is 7.4 billion, rather than the more realistic 7.5 billion (dividing 7.4 by 2 allows me to stop at the first decimal - see below). So as to remain in step with our demographic model I assume that half our numbers are engaged in production. I continue to assume $\rho = 0.5$. And for simplicity we interpret the factor $AK^{1-\rho}$ in equation (4) as an aggregate measure of all assets other than labour.

Write $AK^{1-\rho}$ as $\boldsymbol{K}$. Equation (36) then says

$$\boldsymbol{K} = 110\text{x}10^{12}/(3.7\text{x}10^{9})^{0.5} \text{ dollars per producer}^{0.5}$$

$$\approx 1.8 \text{ billion dollars per producer}^{0.5} \quad (37)$$

To err on the conservative side of GFN's most recent estimate, I take the current global ecological footprint to be 1.5. That means if the biosphere and all other forms of capital assets barring labour were to be stopped on their tracks, their sustainable value would be $\boldsymbol{K}/1.5$, which I denote by $\boldsymbol{K}^*$. Using equation (37),

$$\boldsymbol{K}^* = 1.2 \text{ billion dollars per producer}^{0.5} \quad (38)$$

We have now calibrated the model. As there are two generations in each period, optimum population is $2N^*$ and human carrying capacity is $2N^S$. So we have all the information we need to estimate $2N^*$.[105]

---

103 That I am creating a short-cut to what should be a dynamic optimization exercise and not inventing an ad hoc method of analysis is shown in Appendix 4.

104 This would be a crude assumption in non-stationary states, because it ignores differences among economic sectors in the value labour adds in production. For assumptions other than proportionality, see Part II of the accompanying, reprinted essay (Dasgupta and Dasgupta, 2017).

105 In the accompanying paper, Aisha Dasgupta and I took a different route for estimating the optimum global population number. We sought the number that could, under present technologies and institutions, be sustained at a living standard of 20,000 international dollars. We also assumed that the biosphere is a durable consumption good ($\rho = 0$). Denoting the biosphere in that model by K_1 and by $2N$ the global population, it meant $K_1 = 2NC = 73\text{x}10^{12}$ international dollars. So, $2N = 73\text{x}10^{12}/20\text{x}10^{3} \approx 3.6$ billion. As the biosphere was assumed to be a pure consumption good in the accompanying paper, K_1 (which is expressed in dollars)

11.2 Parameter Values

The sticking points in determining C^* and $2N^*$ will almost certainly be C^S, μ, and the aggregate of capital assets, $\boldsymbol{K}^*$. Our understanding of $\boldsymbol{K}^*$ at the global level is woefully inadequate, and equation (38) only offers a first cut into it.

C^S and μ involve value judgments in pristine form. Moreover, cultural norms enter both (Sect. 5.7). It would be possible in principle to obtain an intuitive feel for μ from the reproductive stopping rules that couples have been known to follow ("We intend to have only one child so that we can give our baby a good start in life"); but the sad truth is that population ethics has been so comprehensively neglected in the social sciences that we have developed no informed intuition about either C^S or μ. Economists and decision theorists have identified reasons for commending that σ should be in the range (0,1], they have devised techniques for estimating income inequality, social discount rates, the cost of carbon in the atmosphere, and for identifying absolute and relative poverty lines; but on ethical parameters that are crucial to population ethics, we have developed little intuition.[106]

Notice that P (eq. (35)) is a decreasing function of μ. That means C^*/C^S (= $(N^S/N^*)^{1-\rho}$) increases as μ is reduced. To get a quantitative feel for C^*, I work with alternative values of C^S and μ.

Recall $\hat{C}$, which is the consumption level at which marginal well-being equals average well-being per unit of consumption (Fig. 5). Under Total Utilitarianism $\hat{C}/C^S$ is the upper bound of the ratio of optimal consumption to well-being subsistence. It is the natural reference point for judging quantitative differences between the directives of Total Utilitarianism and Generation-Relative Utilitarianism.

11.3 Sensitivity Analysis

Studying the sensitivity of C^* to various parameters in equation (33) is about the best we can do today. But sensitivity analysis alone is not helpful. It informs but doesn't tell us how to proceed from there. One thought is to iterate: (i) start with an arbitrary pair of figures for C^S and P and derive C^* and $2N^*$; (ii) revise the figures and estimate the corresponding C^* and $2N^*$; (iii) continue doing so until the evaluator reaches what Rawls would call a "reflective equilibrium" regarding C^* and $2N^*$. Of course, a pair of figures for C^* and $2N^*$ that "feel" right may still be very wrong; but we would then want to know why they are wrong. That would force us to think some more about the issues.

---

has a different dimension from K^* (the latter is expressed in dollars per producer$^{0.5}$). If we now apply the numbers to the present model ($\rho = 0.5$) for estimating the global population that could sustained at a living standard of 20,000 international dollars, we obtain

$$K^*N^{0.5}/2N = 20{,}000 \text{ international dollars,}$$

which on using equation (32) yields a population size of $2N$ = 1.8 billion. That was the global population about a century ago (circa 1925). For further discussion on how many people Earth can support in comfort, see Dasgupta and Dasgupta (2018).

106 Cowie (2016) has argued that we have no clear sense of what a minimally good life is, and that it makes it all too difficult to nail down C^S. The point I am making in the text is that econometricians haven't been asked to offer their views on C^S.

In what follows I suppose θ is just a shade under 1 so as to ensure that $V(t)$ is a convergent series, but I then ignore the error that arises when setting θ = 1. We continue to work with ρ = 0.5 in equation (4) and σ = 1 in equation (5). With these parameter values, equation (35) reduces to $P = (1+\mu)/4\mu$, and equations (34a-b) together read

$$C^*/C^S = (N^S/N^*)^{0.5} = (1+5\mu)/4\mu \tag{39}$$

For illustration I consider four sets of figures for C^S and μ. Using each pair of values in equation (39) yields C^*. C^S and $\boldsymbol{K}^*$ (eq. (38)) then yield N^S on the basis of the equation that says aggregate consumption equals aggregate output:

$$2N^SC^S = \boldsymbol{K}^*[N^S]^{0.5}. \tag{40}$$

For ease of comparison, recall that average world income per person today is 15,000 dollars and world population 7.5 billion.

(1) μ = 0.01. Equation (35) then says $P = 25.25$ and $C^*/C^S = 26.25$. As a trial, we set C^S = 1,500 (international) dollars, which is less than the average annual income in the World Bank's list of low income (read, "poor") countries (Table 1). Then C^* = 39,375 dollars, which is just below the living standard in the Bank's list of high income (read "rich") countries. Equations (38)-(40) together imply $2N^S$ = 320 billion (we haven't accounted for crowding!) and $2N^* \approx$ 464 million.[107]

In contrast, Total Utilitarianism (μ = 1) would commend C^* = 2,250 dollars and $2N^*$ = 142 billion. The difference between the directives of Total and Utilitarianism and Generation-Relative Utilitarianism, respectively, is striking.

(2) μ = 0.05. Equation (35) then says $P = 5.25$ and $C^*/C^S = 6.25$. Let C^S = 1,500 (international) dollars. Then C^* = 9,375 dollars, which is the per capita income in the World Bank's list of low middle-income countries. As in case (1), $2N^S$ = 320 billion, but the optimum population is considerably larger, at $2N^* \approx$ 8.2 billion.

In contrast, Total Utilitarianism (μ = 1) would commend C^* = 2,250 dollars and $2N^* \approx$ 142 billion.

(3) μ = 0.05. Equation (35) then says $P = 5.25$ and $C^*/C^S = 6.25$. As a trial, we set C^S = 3,500 (international) dollars, which falls within the range of per capita incomes in the World Bank's list of low income countries. Then C^* = 21,875 dollars, which is the standard of living in the World Bank's list of upper middle-income countries. Equations (38)-(40) imply $2N^S \approx$ 57.8 billion and $2N^* \approx$ 1.5 billion.

In contrast, Total Utilitarianism (μ = 1) would commend C^* = 5,250 dollars and $2N^* \approx$ 25.7 billion.

(4) μ = 0.1. Equation (35) then says $P \approx 2.75$ and $C^*/C^S \approx 3.75$. As another trial, we set C^S = 7,300 (international) dollars, which is slightly above the per capita income in the World Bank's list of

107 A simple way to introduce crowding externalities would be to make the parameter B in equation (5) a declining function of N for large values of N. See Dasgupta (1969: sect. 4) for a study of the quantitative significance of crowding on Total Utilitarianism's prescriptions.

lower middle-income countries. Then C^* = 27,375 dollars, which is at the lower end of per capita incomes in the World Bank's high-income countries. Equations (38)-(40) imply $2N^S \approx 13.4$ billion and $2N^* \approx 1.9$ billion.

In contrast, Total Utilitarianism ($\mu = 1$) would commend C^* = 10,950 dollars and $2N^* \approx 5.9$ billion.

Four exercises can't reveal much, but they tell us that Generation-Relative Utilitarianism is most distinct from Total Utilitarianism when μ is small (thereby P is large) and C^S is large. I expected the former but didn't expect it would come tied to the latter.

12 Technology and Institutions

When communities face exceptional resource stress (droughts and pests are only two causes), they can be expected to seek new practices and fashion new institutions. If migration to better locations is a possibility, communities can be expected to try that if all else fails. We shouldn't imagine people taking impending disasters lying down. Boserup (1965) collated evidence from agrarian societies to argue that resource stress generates societal responses that not only fend off disaster but can even lead to prosperity. Exceptional scarcities may raise exceptional "problems," but as the saying goes today, they offer exceptional "opportunities" as well. Boserup's work countered a widespread fear in the early 1960s that our capacity to produce food was being overtaken by growth in human numbers. She saw population growth as a spur to innovations. The Green Revolution that came soon after her publication matched her narrative. Population was dropped from public discourse even as Boserup came to be seen as a counterpoint to Malthus.[108]

Boserup's case studies were about "organic economies" (Wrigley, 2004), where not only food but also most raw materials needed for manufacturing material artefacts are either animal or vegetable in origin. Inevitably, there was sample bias in her choice of examples. Societies that hadn't made the cut would have disappeared or moved to blend themselves among communities that survived; they would be absent from such records as those that Boserup studied. In a study of a modern-day society, Turner and Ali (1996) put together the contrasting concerns of Malthus and Boserup by demonstrating that in the face of rising population and a deteriorating resource base, small farmers in Bangladesh expanded production by intensifying agriculture practises (introducing multiple cropping and, **with government help**, collectively strengthening drainage systems and flood and storm defenses). The farmers haven't been able to thrive, they still live in poverty, but they staved off collapse (they haven't abandoned their villages *en masse* for cities), at least for now.

108 Economic historians refer to our need for energy to make the same point. Human societies have over millennia improved their living standard by finding new sources of energy in the face of rising costs of established resources. The succession of human sweat, animal power, rivers and streams, wind, timber, coal, oil and gas, and most recently the nucleus of radioactive matter is a frequently cited example of the global success in finding ways to harness energy.

If Boserup is a counterpoint to Malthus, Jared Diamond is a natural counterpoint to Boserup. Techniques for reading archaeological records have improved since the 1960s. In a series of case studies drawn from early-to-middle second millennium CE, Diamond (2005) found that a number of societies that had deforested their land had been able to develop successful forest management practices and population measures, but that in contrast there were others, most notoriously Easter Island, that had failed to develop successful management practices, and had collapsed as a result. Diamond also found a common pattern in past collapses: Population growth that followed access to an abundant resource base made people intensify the means of food production (irrigation, terracing, double-cropping) and expand into marginal land. Growing populations led to a mining of the resource base; which in turn left communities vulnerable to climatic variations, as there was little room left for either mistakes or bad luck.[109]

PNAS (2012) contains a Special Feature on historical collapses. Contributors reported twelve studies of past societies that had faced environmental stress. Seven were found to have suffered severe transformation, while five overcame them through changes in their practices. Butzer (2012) reported the ways in which a number of societies in 14th-18th centuries Western Europe displayed resilience by coping with environmental stresses through innovation and agricultural intensification. Like Diamond, he concluded that collapse is rarely abrupt.

That collapse is rarely abrupt suggests that socio-ecological systems are not brittle, but that on facing continual stress they become less resilient in withstanding shocks and surprises. In a study of European Neolithic societies that began some 9000 years ago, Downey, Randall Haas, and Shennan (2016) found that the introduction of agriculture spurred population growth, but societies in many cases experienced demographic instability and, ultimately, collapse. The authors also uncovered evidence of warning signs of eventual population collapse, reflected in decreasing resilience in socio-ecological systems. Scheffer (2016) has given further suppport to the thesis by reporting that there had been warning signs of reduced resilience prior to the great drought in the late 1270s that destroyed the communities that had built the iconic alcove sites of Mesa Verde.[110]

Inevitably, these studies have been about societies with tight geographical boundaries. A community that failed because of population overshoot or bad resource-management practices no doubt destroyed their natural resource base, but it was their *local* resource base they destroyed; societies until modern times were incapable of affecting the Earth system as a whole. Matters are different today. Our presence is so dominant that the biosphere is no longer as modular as it was until

109 The present section is taken from Section 5 of Dasgupta, Mitra, and Sorger (2018), which contains a formal model of the socio-ecological processes identified by Diamond in his study of the success and failure of organic economies.

110 That would happen if the resource stock were near-depleted. Scheffer, Carpenter, and Lenton (2012) is an excellent study of the loss of resilience in non-linear systems when they are near bifurcation points. See Appendix 4 for an example of bifurcation points in non-linear systems.

recently. Disturbance in one location today gets transmitted to other parts in short order. Movements of people and trade in goods have created a transmission mechanism with a long and quick reach. The mechanism's medium has, however, remained the same: Nature is mobile. We weaken the Antarctica ice sheet without ever going there; fish in the North Sea eat micro-plastic originating in markets in the Bahamas; phosphorus discharge from farms in Minnesota contribute to a deadening of the Gulf of Mexico; emissions of black carbon from kitchens in the Indian sub-continent affect the circulation patterns of the Monsoons; the Green Revolution's demand for water, fertilizers, and pesticides pollute the rivers and ground waters of the Indo-Gangetic Plain; and so on.

Economic historians of the Industrial Revolution point to the role institutions played in creating incentives for entrepreneurs to find ways to work round natural resource constraints.[111] The rate at which we are able to reduce our dependence on natural resources has to exceed the growth rate of humanity's consumption level. Otherwise our ecological footprint will not decline. The footprint currently exceeds 1 and is continuing to increase. We can be sanguine about the character of technological advances and consumption patterns we correspondingly adopt only if we personally experience the scarcity value of the biosphere, that is, if we have to pay for its use. Understandably, entrepreneurs economize on the expensive factors of production, not the cheap ones. So long as the biosphere's goods and services remain under-priced, technological advances can be expected to be rapacious in their use. Moreover, technological advances that are patently good can have side-effects that are not benign. The tightening of links that bind the biosphere together has meant that economizing on the use of one resource is frequently at the expense of a greater reliance on some other resource (e.g. sinks for our waste products). The ability to use fossil-based energy at large scales has transformed lives for the better, but it has created the unintended consequence of global climate change. Bull-dozers and chain-saws enable people to deforest land at rates that would have been unimaginable 250 years ago, and modern fishing technology devastate large swathes of sea beds in a manner unthinkable in the past. If technological progress is our hope, it has either to come allied with elimination of environmental externalities and be directed by public investment in research and development (R&D)**.**

The recent focus on global climate change has led us, even if imperceptively, to concentrate on technological solutions. But **as noted in Section 10**, climate change is not paradigmatic of environmental problems. Contaminating the oceans with materials in all probability requires collective behavioural change in parallel with **R&D that is moved by public concerns**. **Advances in bio-technology will be an essential requirement if our increased demand for food and fibres are to be met**. Restoration and conservation measures are ways by which we can reduce the global ecological footprint. Creating safe zones for migratory species is another needed measure. Reducing

111 Landes (1998) is a classic on the subject. See also Mokyr (2002, 2016). **In contrast, Barbier (2011) points to societies' dependence on so-far unexploited natural capital even as technological progress circumvented society's reliance on depleting resources.**

waste is yet another. **But** irreversible losses, arising say from biological extinctions, would act as constraints on the biosphere's ability to recover. Social moves toward consumption and production practices that make smaller demands on the biosphere would be a more direct approach to reducing our impact on the Earth system. That's the basis on which the numerical exercises were conducted in Section 11.

13 Existential Risks and Rational Ends

In a deep meditation on the significance of a possible nuclear holocaust in which humanity suffers extinction, Jonathan Schell distinguished two types of death:

"It is of the essence of the human condition that we are born, live for a while, and then die... But although the untimely death of everyone in the world would in itself be an unimaginably huge loss, it would bring with it a separate, distinct loss that would be in a sense even huger - the cancellation of all future generations of human beings." (Schell, 1982: 114-115)

Schell's book was originally published as a three-part essay in *The New Yorker* in 1981, at a time the Cold War had created an especial chill. Schell was a writer, not a professional philosopher, but he made not one false move in philosophical reasoning in the crucial middle chapter, Second Death. Both Total Utilitarianism and Generation-Relative Utilitarianism measure the loss from the Second Death in terms of the well-being of all who would not exist on account of human extinction. The coin with which the loss would be measured is the Total Utilitarian social well-being function, V^U (eq. (31)). In Sections 9-11 we saw how those losses could be estimated.[112] Schell however made a different move. He spoke of the loss each of us alive today would suffer if we were to discover that there will be no one after we are gone. He located that loss not to any attachment we may have to humanity writ large, but to a devaluation of our own lives. And he used the artist and his art to make the point:

"There is no doubt that art, which breaks into the crusted and hardened patterns of thought and feeling in the present as though it were the prow of the future, is in radically altered circumstances if the future is placed in doubt. The ground on which the artist stands when he turns to his work has grown unsteady beneath his feet." (Schell, 1982: 163)

Schell spoke of the artist, but he could have made the same case for all who create ideas and objects. Future people add value to the creators' lives by making their creations durable. Here the fact of a general assumption that people desire to have children is significant. An artist may regard his work to be far more important than parenting, but he is helped by the presumption that there will be future generations to bestow durability to his work.

The examples Schell pointed to were works of art and discoveries in the sciences. Those creations are public goods, and most people don't have the talent to produce them. Confining attention

112 Sikora (1978) is an early exploration for a response to the question using Total Utilitarianism as the guide.

to public goods is not only limiting, it also raises an ethical dilemma: Suppose we all were indifferent to having children and stared only at the prospective costs of raising them. We would then free-ride, and the artist would be mistaken in his assumption that there will be future people to give durability to his work.

Nevertheless, the direction Schell was pointing to is exactly right. Public goods aren't the only objects of ethical significance. Our values and practices are significant too. Many are private, even confined to the family, and it is important to us that they are passed down the generations. Procreation is a means of making one's values and practices durable. We imbue our children with values we cherish and teach them the practices we believe are right not merely because we think it is good for them, but also because we desire to see our values and practices survive. Those values and practices are not public goods. On the contrary, we cherish them *because* they are intimate. They are stories we tell our children about their grandparents' foibles, of our own joys, sorrows, and discomfiture, and we instruct them on the family rituals we ourselves inherited from our parents. Our descendants do something supremely important for us: they add value to our lives that our own mortality would otherwise deprive them of. That is the reason we would not practise reproductive free-riding even if we found reproduction to be personally costly.[113]

The springs that motivate humankind to assume parenthood are deep and abiding. Their genetic basis explains the motivation, but doesn't justify it. Justification is to be found elsewhere. Our children provide us with a means of self-transcendence, the widest avenue open to us of living through time, not merely in time. Mortality threatens to render the achievements of our life transitory, and this threat is removed by procreation. The ability to leave descendants enables us to invest in projects that will not cease to have value once we are gone, projects that justify life rather than merely serve it. Alexander Herzen's remark, that human development is a kind of chronological unfairness because those who live later profit from the labour of their predecessors without paying the same price; and Kant's view that it is disconcerting that earlier generations should carry their burdens only for the sake of the later ones, that only the last should have the good fortune to dwell in the completed building - or in other words, that we can do something for posterity but it can do nothing for us - are a reflection of an extreme form of alienation.[114]

The motivation transmutes from the individual to the collective. Every generation is a trustee of the wide range of **assets** - be **they** cultural or moral, produced or natural - it has inherited from the past. Looking backward, it acknowledges an implicit understanding with the previous generation, of receiving the capital in return for its transmission, modified suitably in the light of changing circumstances and increasing knowledge. Looking forward, it offers an implicit proposal to the next

113 I explored this viewpoint in Dasgupta (2005a). In a wide-ranging and moving essay on death and the afterlife, Scheffler (2013) has also observed that our own lives would be diminished if there were to be no future people.

114 Rawls (1972: 291) has a characteristically profound criticism of Kant's perspective.

generation, of bequeathing its stocks of **assets** that they in turn may be modified suitably by it and then passed on to the following generation. This perspective is not at odds with either Total or Generation-Relative Utilitarianism. In our account of population ethics in a world moving through time, Generation-*t* would be moved to internalize the potential well-being of its descendants, expressed in $V(t)$. Our descendants are not us, but they are not outside us either.[115]

Schell's reflections point also to the intrinsic value of Nature. It's a mistake to seek justification for the preservation of ecological diversity, or more narrowly the protection of species, solely on instrumental grounds; that is on grounds that we know they are useful to us or may prove useful to our descendants. Such arguments have a role but they are not sufficient. Nor can the argument rely on the welfare of the members of such species (it does not account for the special role that species preservation plays in the argument), or on the "rights" of animals. A full justification bases itself as well on how we see ourselves, on what our rational desires are. In examining our values and thus our lives, we are led to ask whether the destruction of an entire species-habitat for some immediate gratification is something we can live with comfortably. The idea of intergenerational exchange is embedded in the perspective of eternity, but the intellectual source of such exchange is a far cry from the conception that balked Herzen in his effort to locate mutually beneficial terms of trade. The mistake is to see procreation and ecological preservation as matters of personal and political morality. They are at least as much a matter of personal and political ethics.

115 The idea of stewardship would appear to be common among different cultures. On African conceptions of intergenerational ethics, Behrens (2012: 189) writes: "African thought does not limit moral consideration to only the current generation. It conceives of a web of life that transcends generations, and of the environment as a resource shared by different generations. This entails a direct moral obligation to preserve the environment for future persons, since it is a communal good. Africans also expect that the current generation should develop an attitude of gratitude towards their predecessors for having preserved the environment on their behalf. This virtue of gratefulness ought to be realized by the current generation seeking to reciprocate by preserving the environment for future generations, in turn." I am grateful to Simon Beard for this reference.

Appendix 1

Socially-Embedded Well-Being Functions

Studies by anthropologists and sociologists have established that personal well-being is socially embedded, that is, it depends not only on one's circumstances but also on the circumstances of others. We compete with others in some spheres of life (Veblen, 1925) and conform with others on some other spheres (Douglas and Isherwood, 1979; Douglas and Ney, 1998). Veblen exposed conspicuous leisure in the Gilded Age in America, and Douglas and her co-authors returned to the ancient view of consumption as gift exchange and showed that it still guides our consumption activities. In the accompanying essay, Aisha Dasgupta and I study the implications of conformism on reproductive behaviour. Our analysis there exploits the fact that conformist preferences give rise to social settings that game theorists know as "coordination games." Here I develop an account of well-being subsistence for a world where individual preferences are competitive. Such preferences give rise to social settings that have been widely studied by social scientists under the name of the Prisoners' Dilemma.

Let C° denote the average level of consumption in society. The U-function would now depend not only on C but also on C°. So we write it as $U(C,C^\circ)$ and note that the function has the following properties:

$$\partial U/\partial C = U_C > 0, \tag{A1.1a}$$

$$\partial U/\partial C^\circ = U_{C^\circ} < 0, \tag{A1.1b}$$

$$\text{for all } C^\circ,\ U(C,C^\circ) < 0 \text{ for low enough values of } C, \tag{A1.1c}$$

$$\text{for all } C^\circ,\ U(C,C^\circ) > 0 \text{ for high enough values of } C. \tag{A1.1d}$$

Conditions (A1.1a-d) imply that for all C° there exists a C_S such that $U(C_S,C^\circ) = 0$ and that C_S is an increasing function of C°; which is to say that welfare subsistence in Veblenesque societies increases with the general standard of living. Write that functional dependence as $C_S(C^\circ)$.

If $U(C,C^\circ)$ is applied to the production model of Section 4.2, the Sidgwick-Meade Rule is meaningful and reads as

$$U(C,C^\circ) = [(AF(K,N)/N)-AF_N](U_C+U_{C^\circ}) \tag{A1.2}$$

So as to illustrate the dependence of well-being subsistence on the average living standard in society, we consider a particular form of the U-function in equations (A1.1a-d)

$$U(C,C^\circ) = C^\theta/(C^\circ+M)^\eta - T, \qquad M, T > 0, \quad 1 > \theta > \eta > 0 \tag{A1.3}$$

Setting $U(C,C^\circ)=0$ in equation (A1.3) yields

$$C^S = T^{1/\theta}(C^\circ+M)^{\eta/\theta} \tag{A1.4}$$

Equation (A1.4) relates C^S to C°.

To have a well-defined problem, we now assume

$$\theta M^{(\theta-\eta)/\theta} > \eta T^{1/\theta} \tag{A1.5}$$

Condition (A1.5) ensures that the slope of $C^S(C^\circ)$ is less than 1. It also means that $C^S(C^\circ)$ cuts the 45^o line in Figure 7 at a unique value of C°, labelled as $\underline{C}^\circ$. Because people are identical in all respects, consumption is the same for everyone at the optimum. It is routine to confirm that optimum consumption exceeds $\underline{C}^{\circ}$.

Appendix 2

Common Property Resources and Reproductive Choices

Throughout the body of this essay I have identified adverse environmental externalities as a common cause of natural capital depreciation. How do those externalities affect fertility choice? A commonplace intuition, stemming from Gordon (1954) and Hardin (1968), is that imperfectly managed natural resources - they don't have to be open-access resources - are over-used because they can be converted into private capital. There is a similar, informal intuition among anthropologists studying rural poverty in the world's poorest countries, that households have an incentive to convert imperfectly managed local resources (grazing fields, water sources, local forests) into household numbers. Communal land tenure of the lineage in sub-Saharan Africa has been seen by anthropologists to offer an inducement for men to procreate, because a large family can (or at least until recently, *could*) claim a greater amount of land than a small family. The accompanying essay with Aisha Dasgupta lists this as one of the forces driving sub-Saharan Africa's high fertility rate. To the best of my knowledge, though, there isn't a formal support for the intuition. Here I make use of the timeless model of Section 8 to confirm it.

Recall the production possibilities in equation (1). The assumption that the function $AF(K,N)$ is homogeneous of degree 1 in K and N means that it has the property of "self-similarity," which is a way of saying that every feasible mix of inputs in production can be applied at any scale of operation. This fact reveals itself most vividly when F is a power function, as in equation (4), which is what I use in the analysis that follows. Assuming that access to K can be monitored and restricted, I first uncover (Sect. (A2.1)) the system of property rights to K that can implement the population optimum (eq. (25)) in a decentralized world. In Section (A2.2) I study fertility choice when a household's entitlement to K increases with household size. I show that this form of property rights motivates households to reproduce more than is socially optimal. So as to avoid additional notation, I assume reproduction is asexual.

A2.1 Decentralizing Population Optima

As in Section 8.2, the model is timeless. There are N_0 individuals. By mutual agreement households are taken to be the social units irrespective of their size. So each individual is restricted to the fraction $1/N_0$ of the common property resource K. We study the decisions of person i ($i = 1,...,N_0$).

Let n be the number of children each of the remaining people intend to have, and let n_i be the number of children i intends to have. If intentions are realized, total population will be $[(N_0-1)(1+n)+(1+n_i)]$. In view of the entitlement rule we are studying, the quantity of K to which each household has access is K/N_0. I assume that production possibilities are given by equation (4). Without loss of generality we set $A = 1$. That means person i's valuation function V_i is

$$V_i = U((K/N_0)^{(1-\rho)}(1+n_i)^{\rho-1}) + \mu n_i U(K/N_0)^{(1-\rho)}(1+n_i)^{\rho-1})$$

or

$$V_i = (1+\mu n_i)U((K/N_0)^{(1-\rho)}(1+n_i)^{\rho-1}) \qquad \text{(A2.1)}$$

It follows from equation (A2.1) that i's optimal choice of n_i satisfies the condition

$$\mu U(C_i) = [(1+\mu n_i)/(1+n_i)](1-\rho)C_i dU(C_i)/dC_i \qquad \text{(A2.2)}$$

In a social equilibrium i's optimal choice of n_i equals n. As in the replicating population in Section 8.2, we take it that N_0 just happens to be the number for which $n = 1$. Denote that number as N^*. Population size is then $2N^*$. Let C^* be average consumption per person in social equilibrium. From equation (A2.2) we know that C^* satisfies

$$\mu U(C) = [(1+\mu)(1-\rho)/2]CU_C \qquad \text{(A2.3)}$$

For vividness, suppose U satisfies equation (5). Then equation (A2.3) reduces to

$$C^*/C^S = (2N^S/2N^*)^{(1-\rho)} = (N^S/N^*)^{(1-\rho)} = [1+\sigma(1+\mu)(1-\rho)/2\mu]^{1/\sigma} \qquad \text{(A2.4)}$$

Comparison of equations (25) and (A2.4) confirms that $C^* = C^{OO}$ and $N^* = N^{OO}$.

A2.2 Open-Access Resources and Population Overshoot

We now consider a different property-rights regime. Imagine that the community takes household size into account when allocating rights to the commons. In that case i's household would be entitled to $\{(1+n_i)/[(N_0-1)(1+n)+(1+n_i)]\}$-th portion of K. To avoid clutter, write $H = (N_0-1)(1+n)$. Then in place of equation (A2.1), we have

$$V_i = (1+\mu n_i)U((K/(H+1+n_i))^{(1-\rho)}) \qquad \text{(A2.5)}$$

In a social equilibrium i's optimal choice of n_i equals n. And as in the replicating population studied in Section 8.2, I assume N_0 just happens to be the number for which $n = 1$. Denote that number as N^{**}. Population size is then $2N^{**} > 1$. Let C^{**} denote average consumption per person in social equilibrium. From equation (A2.5) it follows that C^{**} satisfies

$$\mu U(C) = [(1+\mu)(1-\rho)/2]CU_C/2N^{**} \qquad \text{(A2.6)}$$

For vividness, suppose U satisfies equation (5). Then equation (A2.6) reduces to

$$C^{**}/C^S = (N^S/N^{**})^{(1-\rho)} = \{[1+\sigma(1+\mu)(1-\rho)/2\mu]^{1/\sigma}\}/2N^{**} \qquad \text{(A2.7)}$$

Comparison of equations (25) and (A2.7) confirms that $C^{**} < C^{OO}$ and $N^{**} > N^{OO}$. The result confirms the suggestions in Sections 8 and 11 concerning population and the environment: Freedom on the commons leads to over-population.

Appendix 3

Notes on Rawls' Principle of Just Saving

Most writings on the ethics of intergenerational saving have been built on Utilitarian thinking. But Rawls (1972: 284-298) sketched a principle of saving in his contractual theory of justice. It will prove useful to see in which ways the reasonings differ from each other. For convenience I quote the two principles of justice Rawls arrived at (p. 302):

"*First Principle*: Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

Second Principle: Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit to the least advantaged, consistent with the just saving principle, and (b) attached to offices open to all under conditions of fair equality of opportunity."

There are priority rules among the principles. They have been much discussed since the publication of *A Theory of Justice*. Far less attention has been paid to Rawls' saving principle, which is what binds the generations and is therefore of interest to us here.

Rawls argued that the Difference Principle (ranking infinite well-being streams in accordance with "greatest benefit to the least advantaged" generation) would not be reached by the parties behind the veil of ignorance because well-designed investment projects have a positive social return (p.291):

"It is now clear why the difference principle does not apply to the saving problem. There is no way for later generations to improve the situation of the least fortunate first generation. The principle is inapplicable and it would seem to imply, if anything, that there be no saving at all. Thus, the problem of saving must be treated in another fashion."

A3.1 Parental Motivations

Rawls found a basis for the required treatment by drawing on the fact that parents care about their children. Here is how he states the just saving principle (p. 287-288):

"The parties do not know which generation they belong to... They have no way of telling whether it is poor or relatively wealthy, largely agricultural or already industrialized... The veil of ignorance is complete in these respects. Thus the persons in the original position are to ask themselves how much they would be willing to save at each stage of advance on the assumption that all other generations would save at the same rates... Only those in the first generation do not benefit ... for while they begin the process they do not share in the fruits of their provision. Nevertheless since it is assumed that a generation cares for its immediate descendants, *as fathers care for their sons*, a just saving principle ... would be adopted." (Italics added)

Rawls (p. 292) tied this to the Difference Principle by summing up thus:

"We now have to combine the just saving principle with the two principles of justice. This is done by supposing that this principle is defined from the standpoint of the least advantaged in each generation. It is the representative men from this group as it extends over time who by virtual adjustments are to specify the rate of accumulation. ... Thus the complete statement of the difference principle includes the saving principle as a constraint."

Rawls was less than transparent in his formulation. "As fathers care for their sons" can be read to mean that a father's personal well-being depends not only on his own living standard but also on his sons' living standards. But it could also be read to mean that fathers care for their sons' well-being. In the latter case parental well-being takes the recursive form we encountered in Section 9.2. The problem is, as Koopmans (1972) showed, a few normative moves reduce the recursive form to equation (29), which is Total Utilitarianism with a positive discount rate (Sects. 9.1-9.2). That appeared to me to be so contrary in spirit to the rest of Rawls' work, that in Dasgupta (1974b) I took the link between adjacent generations to be the sons' living standard. That's a truncated view of what binds the generations, but as Rawls was reviving a contractual theory of justice, not a Utilitarian one, I felt I had interpreted him correctly.

Rawls' ideas on just saving weren't built on any specific population projection. The principles of justice pronounce on the basic structure of society, not population numbers. Instead, Rawls studied the pure saving problem. We may as well then assume that population is constant and for simplicity identify each time period with a generation.

Time is taken to be discrete. Consider someone of generation-t, where $t \geq 0$. If $C(t)$ denotes his consumption level, we are to imagine that

$$U_t = U(C(t)) + \theta U(C(t+1)), \qquad 0 < \theta \leq 1 \tag{A3.1}$$

This is the motivation assumption in Dasgupta (1974b). We now assume that the world is endowed with a single capital **good** that can be consumed or saved. Investment enjoys a positive return. Consumption takes place at the beginning of each period. Thus, if $K(t)$ is the stock at time t (time and generations are labelled identically), the accumulation process is defined by the equation

$$K(t+1) = r(K(t) - C(t)), \qquad r > 1 \tag{A3.2}$$

The rate of return on investment r-1. The saving rate can now be defined as

$$s(t) = (K(t) - C(t))/K(t) \tag{A3.3}$$

from which it follows that

$$K(t+1) = rs(t)K(t) \tag{A3.4}$$

Notice that $s(t)$ lies in the unit interval [0,1]. Because consumption takes place at the beginning of each period, net saving is zero at t if $s(t) = 1/r$. Saving is positive if $s > 1/r$, it is negative if $s < 1/r$.

Because Rawls rejected the Difference Principle for identifying the just saving rule, I studied equilibria that would emerge in a non-cooperative game among the generations with well-being functions (A3.1). I followed Rawls' wording of the saving principle and studied Nash equilibria when each generation chooses its saving rate. Arrow (1973) agreed with the way I had formulated parental motivations (eq. (A3.1)) and the accumulation process (eqs. (A3.2)-(A3.4)), but applied an intergenerational max-min principle to infinite well-being streams. That is why we published our papers separately.

The character of Rawls' saving policies under the two formulations were reviewed in Dasgupta (1974b). Each was found to have questionable features. Nash equilibrium saving rates were found to be intergenerationally inefficient (there are non-equilibrium saving rules under which all generations would enjoy higher levels of well-being). That would be found objectionable behind the veil of ignorance. I also found that if refined versions of the concept of Nash equilibrium were deployed by the generations, there would be a plethora of equilibrium outcomes. Rawls would then require further normative directives for selecting one from among the multiplicity of refined Nash equilibria.

Arrow (1973) showed that if $\theta r < 1$, intergenerational max-min commends that there be no (net) saving (i.e. $s(t) = 1/r$), which means the stock of capital remains constant through time. Zero net saving per se is not to be rejected (Section 11 and Appendix 4 below show why), what is worrying for the theory is that it commends zero net saving irrespective of the capital stock society has inherited from the past. But Dasgupta (1974b) also showed that if $\theta r > 1$, the saving behaviour Arrow had deduced is intergenerationally inconsistent: the expectations of each generation are thwarted by the desired saving rate of the next generation.

Independently of us, Solow (1974b) put intergenerational max-min to work in a model in which a constant population produces output when working with produced capital and an exhaustible resource. But he was under no illusion that Rawls advocated max-min (p. 30):

"In this article I am going to be *plus Rawlsian que le Rawls*: I shall explore the consequences of a straightforward application of max-min principle to the intergenerational problem of capital accumulation."

In a canonical example of production possibilities (the power-function form in equation (4)), Solow showed that if the rents from the exhaustible resource used in production are saved and invested, the economy follows the max-min path of constant consumption. Along the path the exhaustible resource is depleted gradually, but produced capital is accumulated in such a way that inclusive wealth (Section 3.1) remains constant. Net saving is zero.[116] Dasgupta and Heal (1979: Ch. 10) applied the production structure studied by Solow to a society wedded to Total Utilitarianism with zero discounting for time and the generations and showed that the optimum saving rate is positive. Positive saving leads to rising living standard over time and the generations. The exhaustible resource is depleted gradually, but produced capital is accumulated at a sufficiently fast rate to ensure that inclusive wealth grows indefinitely.

Rawls' ideas of intergenerational justice in *A Theory of Justice* were unsuccessful. When modelled formally, they were found to have unsatisfactory saving behaviour. Perhaps because of those findings (I can't be sure, of course) he offered a different reading in a later publication (Rawls, 1993: 274). There he wrote:

---

116 The result was generalised by Hartwick (1977) to wider production structures.

"... the correct principle is that which the members of any generation (and so all generations) would adopt as the one their generation is to follow and as the principle they would want preceding generations to have followed (and later generations to follow), no matter how far back (or forward) in time."

Unfortunately this gets Rawls nowhere; including past generations in the exercise makes no difference. If the motivation assumption in his previous work is retained, as in equation (A3.1), the saving behaviour identified in Dasgupta (1974b) remains, with all its problems intact. On the other hand, if the motivation assumption is dropped, by which I mean the chooser in the original position cares only about his own prospects, an independent notion of fairness has to be introduced if generations are to be motivated to save on grounds of justice. Otherwise, in a world where people of the same generation have the same wealth, no saving can be expected to be forthcoming no matter whether the chooser looks back or looks forward.

English (1977) had previously coaxed positive net saving out of Rawls' revised proposal (Rawls, 1993, acknowledges her priority over the idea), but she did it by arguing, on empirical grounds, that even in a growing economy some people of generation-(t+1) will be less wealthy than some people of generation-t. The wealthier person in generation-t would then be motivated to save for the less wealthy person in generation-(t+1). The problem is, in a growing economy the weighted average wealth of people of generation-(t+1) would be greater than that of people of generation-t. There is thus every likelihood that the worst off person in generation-(t+1) would be better off than the worst off person in generation-t. The Difference Principle would now kick in from the opposite direction. Generation-(t+1) will reason that justice would have been better served had the previous generation saved less. Generation-t will have anticipated that reasoning. That reasoning is always implementable, because saving can be made negative by depreciating assets (by setting $s(t) < 1/r$), or more generally, by mining natural capital (Sect. 10). We should conclude that Rawls didn't have a principle of just saving after all.

I am a theoretical economist, not a moral philosopher. I don't have the facility to resuscitate Rawls' theory of just saving. But it seems to me one does disservice to the idea of personhood if our deep-rooted concern and affection for our children is not acknowledged to be an integral part of what defines us. And so it seems to me we would not wish to shed those concerns of ours when entering the original position. Even if we are not fortunate enough to meet our grandchildren, we will have met our children and will know they will meet and care about their children, and so on, through all our descendants. That's a deep fact of our motivation, and to discard it would be to imagine that we enter the original position with a truncated conception of the self. So I now think I was wrong in 1973 to dismiss the recursive form of personal well-being when modelling Rawls' motivation assumption in *A Theory of Justice*. Taking the lead of equation (29), we could push the consideration and even imagine that the well-being function of someone of generation-t would be

$$U_t = U(C(t),U(C(t+1),U(...) = {}_{u=t}\sum^{\infty}[U(C(u))/(1+\delta)^{-(u-t)}], \qquad \delta > 0, \qquad \text{(A3.5)}$$

where $C(u)$ is the living standard of the person's descendant in generation-u. In standard economic models the intergenerational Nash equilibrium saving behaviour of interest would be the one that maximizes U_t for each t subject to all other generations complying with their roles in fulfilling the the saving policy. The policy would thereby be self-enforcing, a crucial requirement of a contractual theory of justice. But the just saving rule would be indistinguishable from Utilitarian solutions to the optimum saving problem.

Fertility choice isn't of direct concern in Rawls' thought experiment. But under the "present time of entry" interpretation of the original position, it is tempting to ask whether citizens behind the veil could reach a view that is consistent with Generation-Relative Utilitarianism. Dasgupta (1974a) contains a model that could be used to test the idea. When I constructed it I was so unsure Rawls' thought experiment was a suitable basis for developing population ethics, that I didn't go down that route. I still haven't, but would be less hesitant today to try it for size.

A3.2 Inclusive Wealth as Rawlsian Primary Goods

The principles of justice in Rawls (1972) address the basic structure of society. But Rawls needed a way to bind the generations together, and he found exactly the right objects with which to do the binding. Rawlsian "primary goods" are the ingredients of a society's *productive base*. His saving principle addresses the transmission of (inclusive) *wealth*.

If this reads as a bizarre interpretation of what is probably the most intensively scrutinized treatise on political philosophy in the last hundred years, recall the wealth/well-being equivalence theorem (Sect. 3.2). An economy's inclusive wealth is the social worth (or accounting value) of its stock of capital goods (produced capital, human capital, and natural capital). That worth is shaped by the social environment in which people try to realise their projects and purposes. A society where citizens lack opportunities to meet one another freely and develop institutions that protect and promote trust among themselves will not prosper. Which is why the social environment includes various forms of personal liberties and access to opportunities (Rawls' First Principle and part (b) of the Second Principle). An economy's social environment imbues its capital goods with their social worth, large or small depending on its character, implying that it is composed of the economy's *enabling* assets (Sect. 3.2). Other things equal, an economy's stock of capital goods has lower social worth if society is dysfunctional than if it is well-ordered.[117]

The claims of justice are over citizens' ability to create and access their productive base. Inclusive wealth is a measure of the productive base. An economy's inclusive wealth is not the same as social well-being, but the equivalence theorem says it moves in tandem with social well-being, which is why it is the object of interest in both policy analysis and in analyses of sustainable development (Appendix 5).

117 Rawls limited his analysis to economically advanced, well-ordered societies; but I am concerned here with the spirit of his enquiry, not whether his priority rules would be found attractive behind the veil of ignorance in poor societies.

It is however the case that in stating the general conception of justice he was putting forward, Rawls (1972: 62) spoke of the distribution of "income and wealth", an unexpected carelessness in a towering work:

"All social values - liberty and opportunity, income and wealth, and the bases of self respect - are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage."

Wealth is a stock, income is a flow. The two amount to the same object in a stationary economy, because in that world one is proportional to the other. But they are not proportional to each other in any other world. In the former you don't need both; in the latter there is no principle one could appeal to for combining them in an aggregate measure.

Income has no place in Rawls' principles of justice because, being a flow, it is unable to reflect anything about an economy's future prospects and the corresponding prospects of its citizens. To be sure, many of the claims a Rawlsian citizen is entitled to make on her fellow citizens will be in the form of flows (medical care, unemployment benefits, child maintenance allowance, and so on), but they are claims to a share of national wealth. For administrative convenience it may be that she is also taxed on her income, not wealth; but it is (inclusive) wealth that is the real object of interest. The notion of wealth that was sketched in Section 3.2 is far more expansive than the one Rawls had in mind, but it is the inclusive notion that responds to the Rawlsian citizen's needs no matter what kind of society she happens to inhabit. Because inclusive wealth is a measure of an economy's productive base, it is able to reflect her society's future prospects no matter how functional or dysfunctional that society happens to be (Appendix 5). It follows that so long as inclusive wealth is distributed in accordance with the choice made behind the veil of ignorance, it reflects citizens' prospects in the Rawlsian world.

A3.3 Wealth and Materialism

But wealth, even inclusive wealth, smacks of materialism, a worry that lies behind not only the criticism of Rawls' theory by Communitarians (Sandel, 1982), but also the one lodged by Bernard Williams. After giving a favourable term report on *A Theory of Justice*, Williams (1985: 80) wrote:

"The list of primary goods does not plausibly look as if it had been assembled simply from the consideration that they are uniquely necessary for pursuing anything. From that consideration we are not likely to derive more than liberty. It is hard to see, also, how the parties could avoid the reflection (available to them from their knowledge of general social facts) that some of these primary goods, notably money, are more important in some societies than in others."

Rawls included wealth among primary goods (we have already seen why income should not have entered the list), not money, unless the economy in question is monetized. I don't know whether the misconception Williams suffered from about wealth is common, but as it is Williams who had that misconception, I imagine it is not uncommon.

The wealth/well-being equivalence theorem holds in all societies, not just in monetized ones.

For a hunter-gatherer society the list of capital goods would include the amount of territory the tribe commands, the supply of water it can count upon, the extent of game within reach, the quality of their hunting equipment, the infant survival rate, the number of able-bodied men and women, the number of old people (the repositories of tribal knowledge), and so forth. It doesn't matter whether the economy is monetized or whether exchanges are conducted through barter, or social custom, or the dictates of the tribal elder: assets matter to people in all societies. Their worth, whether it is expressed in monetary units or in cowry shells, reflects the contributions they make (at the margin!) to the well-being of tribal members. Whatever else a person needs, he needs capital goods. That is so regardless of whether they are owned and managed collectively or they are privatised.

Behind the Rawlsian veil of ignorance the citizen does not of course know the worth of those capital goods. For one thing there is insufficient knowledge about the social environment that will prevail in the later stages of Rawls' thought-experiment (e.g. whether it will be a property owning democracy). The veil is so thick that the chooser doesn't even possess a sense of his good. But choice behind the veil is only the first stage of political engagement. Citizens know that once the veil lifts gradually and they learn more and more about themselves and the world, they will probably want to establish **the law of limited liability, and** institutions such as banks, charities, cooperatives, and mortgage and insurance companies, at the very least a monetized economy. Choices over matters such as whether a natural monopoly is to be located in the private or public sector are made at later stages of the thought experiment, when the veil has been lifted sufficiently. As the Rawlsian veil is lifted inclusive wealth comes to have a more concrete magnitude. But even behind the veil citizens know they will want primary goods if they are to pursue their projects and purposes, whatever they may happen to be.

I have heard it suggested over the years that Rawlsian choice behind the veil of ignorance is the same as the one to which Harsanyi (1955) appealed for deriving Utilitarianism. The difference between the theories as expressed in the authors' publications, it is often observed, lies in differences in the chooser's reasoned aversion to risk. In fact the difference is far greater. It is to misread Rawls entirely not to recognise the centrality of the temporal structure of the thought experiment he used to identify the principles of justice. The Rawlsian citizen is invited to identify the principles she would choose when they are to guide the allocation of resources in a world where, with the passage of time, she learns more and more about herself and the world. Which is why she is seen as making commitments to herself of varying strengths, the strongest (the principles of justice) being the ones she makes when the veil is thickest. We are thus encouraged to think of a hierarchy of commitments, those made in a given stage being less firm than those in any previous stage. To me this reading puts into perspective the phenomenal opening pair of sentences of his treatise: "Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be abolished if they are unjust."

One reason the Rawlsian citizen *wants* to choose the principles behind a thick veil is that by so doing she is able to tie her hands so as to better exercise her options as and when her ignorance is lifted. The principles of justice serve as a commitment device for future flexibility of choice. Current uncertainty about future circumstances and the fact that many of the choices she makes, or is made on her behalf, are irreversible (e.g. whether to learn to read and write in the early years of one's life; Section 3.3) give value to keeping her options open for as and when occasions arise. The priority rules Rawls thought he had arrived at can be questioned, but that should not detract from the truth that Rawls' primary goods point to a society's productive base, nor that the distribution of primary goods reflects the distribution of inclusive wealth, and thereby the distribution of personal well-being. Choosing to commit oneself to the principles of justice behind the veil of ignorance serves the same purpose (though to be sure it is toward a far greater matter) as purchasing options in contemporary markets for securities. A formal model that illustrates the argument is presented in Appendix 6.[118]

118 Rawls' views on the substance of the original position in political discourse underwent a shift in his later writings (Rawls, 1993), but that doesn't affect my comments on *A Theory of Justice*. I am grateful to Ira Katznelson for an illuminating conversation on Rawls' later views on the content of liberalism.

Appendix 4
Modelling the Biosphere

In order to display the workings of dynamical systems, it helps to simplify human demography by imagining that people are both producers and consumers throughout their life. In Section 10 the biosphere was viewed to be a gigantic renewable natural resource. To illustrate that in formal terms we model human activities in a simple ecosystem. It then helps enormously to regard time as a continuous variable, $t \geq 0$. **(For a more complete study of cupled socio-ecological models, see Brock and Xepapadeus, 2018).**

A4.1 Pure Resources

In the text, we denoted the aggregate measure of capital goods as K. Because the model we construct here includes all three categories of capital goods, we alter the notation and denote the biosphere as S. Now distinguish stocks from flows. Let $S(t)$ be the state of the biosphere at time t. $S(t)$ is a stock, measured in, say, tons of biomass. In this section we study the biosphere (natural capital) as the sole object of interest to the human population. In subsequent sections we combine natural capital with produced capital.

The net output of goods and services produced by the biosphere at t is denoted as $G(S(t))$. $G(S(t))$ is a flow (biomass per unit of time). The most common form of G in the empirical literature on forests and fisheries is quadratic. A generalized form is

$$G(S) = -T + rS(1-S/L), \quad \text{if } S > 0, \quad r,T,L > 0;$$
$$= 0, \quad \text{if } S = 0 \tag{A4.1}$$

In order to ensure that the Earth system can support a biosphere, we assume

$$r > 4T/L \tag{A4.2}$$

(Otherwise $G(S) = 0$ for all S.) Figure 8 displays the G-function in equation (A4.1), in which the parameters of G satisfy condition (A4.2).

We imagine first that there is no human predation. In that situation $S(t)$ changes over time in accordance with the dynamical equation

$$\mathrm{d}S(t)/\mathrm{d}t = -T + rS(t)[1-S(t)/L], \qquad S(t) > 0 \tag{A4.3}$$

Equation (A4.3) has three stationary points (Fig. 8):

$$S = 0;\ S = S_1 = \{r-(r^2-4rT/L)^{1/2}\}L/2r;\ S = S_2 = \{r+(r^2-4rT/L)^{1/2}\}L/2r \tag{A4.4}$$

In Figure 8 $G(S)$ is shown to attain its maximum value at S^{**}. $G(S^{**})$ is the ecosystem's maximum sustainable yield (MSY). From equation (A4.1) and Figure 8 it is apparent that of the three stationary points, $S = 0$ and $S = S_2$ are stable, whereas $S = S_1$ is unstable. We imagine for vividness that before human predation, the system rested at S_2.

The ideas underlying the dynamical system in equation (A4.3) are as follows: (1) S has to be of a minimum size S_1 (read minimal biodiversity) if the biosphere is to flourish; (2) at values of S greater than S_1 (but less than S_2) the stock grows in size, but the growth rate declines at large values of S because the system is constrained by a finite planet; (3) the Earth system is incapable of sustaining a biosphere of size greater than S_2.[119]

Now introduce human predation. Let $N(t)$ be population size at t. Producing the consumption good requires labour and the use of the biosphere's services. Denote by $R(t)$ the rate at which humanity draws on the biosphere at time t. It is conceptually simplest if we think of $R(t)$ as the rate of extraction of biomass at t. We can now express the socio-ecological system as

$$dS(t)/dt = -T + rS(t)(1-S(t)/L) - R(t) \qquad \text{(A4.5)}$$

The purpose behind extracting R is to produce consumption goods. As previously, write global output (GDP) as Q. Using the production structure of equation (4), we have

$$Q(t) = AR(t)^{1-\rho}[N(t)]^{\rho}, \qquad A > 0, \quad 0 < \rho < 1 \qquad \text{(A4.6)}$$

Equation (A4.6) is the dynamic counterpart of equation (4) in the text. In the timeless world of Sections 5 and 8, stocks and flows amount to the same thing. In the text we had denoted the aggregate measure of produced capital and natural capital by K and had assumed aggregate output there was $Q = AK^{1-\rho}N^{\rho}$. But even in the world of Section 9 the biosphere was assumed not to be a dynamic entity. Instead we assumed that the world enjoys a constant supply of the biosphere's goods and services and that the stock of produced capital is constant. That constant stock of produced capital and natural capital (the biosphere) was denoted by an aggregate index K. Matters are different here. The rate at which we appropriate goods and services from the biosphere can be as large as we like, but at the cost of a degraded biosphere; that is, of a reduced K.

To model those possibilities, let $C(t)$ be consumption per capita at date t. As in Section 9, we assume for illustrative purposes (but see Sect. (A4.2)) that what is produced is consumed. It follows from equation (A4.6) that

$$N(t)C(t) = AR(t)^{1-\rho}[N(t)]^{\rho} \qquad \text{(A4.7)}$$

On using equation (A4.7), equation (A4.5) reads

$$dS(t)/dt = -T + rS(t)(1-S(t)/L) - N(t)[C(t)/A]^{1(1-\rho)} \qquad \text{(A4.8)}$$

Equation (A4.8) can be interpreted as the balancing of biomass across time:

$$dS(t)/dt > 0 \text{ if } -T + rS(t)(1-S(t)/L) > N(t)[C(t)/A]^{(1-\rho)} \qquad \text{(A4.9a)}$$

$$dS(t)/dt < 0 \text{ if } -T + rS(t)(1-S(t)/L) < N(t)[C(t)/A]^{(1-\rho)} \qquad \text{(A4.9b)}$$

---

119 A simplified form of $G(S)$ that ecologists use to model a wide variety of ecosystems (fisheries, trees), assumes $T = 0$. In that case equation (A4.3) becomes

$$dS(t)/dt = rS(t)[1-S(t)/L], \qquad r, S > 0 \qquad \text{(F4.1)}$$

Here r is the "intrinsic growth rate" of S (r at small values of S is the percentage rate of growth of S) and S is the ecosystem's carrying capacity for life forms barring humans ($G(S) = 0$). Integrating equation (F4.1) yields

$$S(t) = S(0)Le^{rt}/[L+S(0)(e^{rt}-1)] \qquad \text{(F4.2)}$$

(If $S(0) < L/2$, $S(t)$ assumes the classic sigmoid shape). It follows from equation (F4.2) that $S(t) \rightarrow L$ as $t \rightarrow \infty$. That's the stable stationary point of the system. The stable stationary point of equation (A4.3) that corresponds to L in equation (F4.1) is S_2.

The growth of trees from their seedling stage has been modelled by forestry experts as $S(t)$ in equation (F4.2).

Section 10 collated all too briefly evidence that the biosphere is now below the stock at which it is offers maximum sustainable yield (MSY) and more sharpky that it is experiencing inequality (A4.9b). In keeping with the notation of Sections 9 and 11, Figure 8 **in contrast** draws attention to a stock S^* at which $-T + rS^*(1-S^*/L) = N(t)[C(t)/A]^{(1-\rho)}$. The quartet of variables $\{S^*, C^*, R^*, N^*\}$ denotes a stationary state. The stock is maintained at the level S^* by drawing biomass at each moment at the rate R^*.

To contrast this with the route we followed when conducting numerical exercises in Sections 9 and 11, recall that we stopped all capital goods barring labour on their tracks and denoted the aggregate of those capital goods as $\boldsymbol{K}^*$. The pair $\{C^*, N^*\}$ was obtained in **Section 11** by an optimization exercise in Generation-Relative Utilitarianism, while holding $\boldsymbol{K}^*$ fixed.

A4.2 A Proto-Type Capital Model

We can extend the previous model to include produced capital. To see how, let $M(t)$ be the stock of a non-deteriorating all-purpose commodity that can be consumed or set aside as investment for enhancing future consumption. Suppose output of the all-purpose commodity is given by the function

$$Q = AR^aM^bN^{(1-a-b)}, \qquad A > 0, \quad 0 < a, b, 1-a-b < 1 \tag{A4.10}$$

I have specified the output function in equation (A4.10) in the form of a power function only for convenience, nothing more. A power function is unrealistic because it says no matter how small a value of R you care to choose, output can be as large as you like provided there are enough people and/or enough produced capital to work with. But the Q-function in equation (A4.10) isn't overtly restrictive because we settle our discussion here on stationary states, and that won't require of us to consider extreme substitution possibilities among the factors of production.

The dynamics of the socio-ecological system now reads to include, in addition to equations (A4.5) and (A4.10), an equation that says that each date t the society's intended saving equals its intended investment in the all-purpose consumption good. Using equation (A4.10) we have

$$dM(t)/dt = AR^a(t)M^b(t)N(t)^{(1-a-b)} - N(t)C(t) \tag{A4.11}$$

Stocks of capital goods are state variables in dynamic socio-ecological models, whereas control variables are subject to immediate choice. In the present case the dynamics of the state variables are given by equations (A4.5) and (A4.11).

We now re-introduce the Decision Maker (DM), who is a Total Utilitarian. The Genesis Problem is solved at date $t = 0$. Social well-being is the integral of the discounted flow of utilities from $t = 0$ to ∞. Then

$$V(0) = {}_0\int^{\infty}[N(t)U(C(t))]e^{-\delta t}dt, \qquad \delta > 0 \tag{A4.12}$$

In equation (A4.12), δ can be thought of as either the hazard rate of extinction or the discount rate for time, or a combination of both (Sect. 8.1). We make the same assumptions on the U-function as we did in Section 5.[120]

$C(t)$ and $R(t)$ are control variables. They are subject to DM's choice at each t. That leaves us with population profiles over time. There are two ways to model this. One is to suppose, as would seem natural when DM is the chooser (the Genesis Problem), that $N(t)$ is a choice at each moment t (including the present, $t = 0$). The world pictured there would be, to use the terminology in Arrow and Kurz (1970), *fully controllable*. The other is to suppose that the rate of change in $N(t)$ is a choice, but not $N(t)$. The world pictured there is only partially controllable. In the former case even the present stock of people is a choice for DM, whereas in the latter case the stock of people only in the distant future are subject to choice by DM. I realise that in the latter case Total Utilitarianism is being applied in the world we live in, and that's a long after Genesis, but I want to see how Total Utilitarianism fares in a dynamic model. Because the two cases differ in the extent to which the socio-ecological system is controllable by DM, they point to differences in their long-run optima. We study them in turn.

A4.3 Population Size as Choice

DM's choice variables at each moment t are $C(t)$, $R(t)$, and $N(t)$. The two state variables of the socio-ecological system are $S(t)$ and $M(t)$. As this is the Genesis Problem, it is reasonable to imagine that $S(0) = S_2$ (eq. (A4.1)); fig. 8). So as to give the economy a kick start, we suppose that $M(0) > 0$. The dynamics of the economy are governed by equations (A4.5) and (A4.11).

I avoid technicalities here, as it calls for the use of the variational calculus. Instead I use the findings in Dasgupta (1969; Sect. 3) to sketch the optimum policy. It can be shown that if $\delta < r$ (the latter is the (marginal) productivity of the resource base at S_1 (eq. (A4.1)), the solution of DM's optimization problem converges to a stationary state. If $M(0)$ is large relative to the other factors, $U(C(t)) > 0$ for all t. If $M(0)$ is small relative to other factors, there is a finite initial period when $U(C(t)) < 0$, during which produced capital is accumulated for future prosperity. In due course $U(C(t)) > 0$.

In the long run stationary state (and our assumptions say that there is a unique optimum stationary state) let p and q be the social scarcity prices of produced capital M and natural capital S, respectively. In Section 3 we referred to p and q also as accounting prices. It can be shown that the stationary state to which the Total Utilitarian optimum tends is given by the solution to the equations

$$U_C = p \tag{A4.13a}$$

$$paAR^{a-1}M^bN^{1-a-b} = q \tag{A4.13b}$$

$$U(C) = p[C-(1-b-c)AR^aM^bN^{-(a+b)}] \tag{A4.13c}$$

$$\delta = bAR^aM^{b-1}N^{1-a-b} \tag{A4.13d}$$

$$\delta = G_S(S) \tag{A4.13e}$$

$$G(S) = R \tag{A4.13f}$$

120 The corresponding analysis for Generation-Relative Utilitarianism is harder, especially characterizing the approach path to the optimum stationary state. I have the general formulation, but haven't been able yet to nail the details,

$$AR^aM^bN^{1-a-b} = NC \qquad \text{(A4.13g)}$$

Equations (A4.13a-g) are seven in number and there are seven unknowns to determine: *p*, *q*, *R*, *M*, *N*, *S*, *C*. Given our assumptions, the solution is unique and represents the optimum stationary state. The equations are derivable from the optimization techniques of Pontryagin, and have an intuitive rationale:

Equation (A4.13a) says that marginal well-being equals the accounting price *p* of the all-purpose produced capital.

Equation (A4.13b) says that the social value of the marginal productivity of natural capital equals the accounting price *q* of natural capital.[121]

Equation (A4.13c) is the Sidgwick-Meade Rule. Like equation (3) in the text, it says that average well-being *U* at the optimum stationary state equals the social value of the difference between average consumption and the marginal person's productivity.

It is simple to establish from equation (A4.12) that in a stationary state the social rate of discount is δ. Equation (A4.13d) says that in the optimum stationary state δ equals the social value of the marginal productivity of produced capital.

(A4.13e) says that δ also equals the social value of the marginal regeneration rate of natural capital.

Equations (A4.13.d-e) are fundamental to intergenerational welfare economics. They reflect the balancing act that Total Utilitarianism accomplishes between well-being at one moment and well-being at another.

(A4.13f) says that at each moment the rate at which natural capital is extracted (*R*) equals Nature's regeneration rate *G*(*S*).

(A4.13g) says that aggregate consumption at each date equals aggregate output of the consumption good. In a stationary state net investment is zero.

Denote all economic variables in the optimum stationary state by a dagger (†). Equations (A4.13c) and (A4.13g) tell us that $U^\dagger = U(C^\dagger) > 0$. In Section 3 it was noted that the coin with which we should judge the performance of economies is their (inclusive) wealth. Write wealth as *W*. Our model economy contains two assets: produced capital (*M*) and natural capital (*S*). It could rightly be thought population (*N*) is also a capital asset, but as we are studying the Genesis Problem, *N* has been taken to be a choice variable at all dates, meaning that it is on par with consumption. Wealth in the optimum stationary state is

$$W^\dagger = p^\dagger M^\dagger + q^\dagger S^\dagger \qquad \text{(A4.14)}$$

Conditional on the economy surviving, social well-being (eq. (A4.12)) remains constant over time, as does wealth (eq. (A4.14)). Denote social well-being at the stationary optimum by $V^\dagger$. As the economy is at the optimum, any move away from it by changing the variables would reduce *V* to a

[121] In an exceptional paper, Fenichal and Abbott (2014) used a proto-type of the model here and applied it to fisheries data to estimate the accounting price of reef fish in the Gulf of Mexico.

figure below $V^\dagger$. Simultaneously, it would reduce W to a figure below $W^\dagger$. That's an application of the wealth/well-being equivalence theorem, which was sketched in Section 3.2 and will be developed further in Appendix 5. Social well-being and inclusive wealth are two sides of the same evaluative coin.

A4.4 Birth Rate as Choice

We now suppose that $N(0)$ is not a choice variable for the Decision Maker (DM), nor is the death rate a choice variable. We take it though that the birth rate is a choice variable. Let $x(t)$ be the decision maker's choice of the birth rate at t per unit population (usually expressed by demographers as "per 1000") and ξ be the natural death rate per unit of population. We suppose that $x(t)$ can't exceed some positive number, say λ, which represents the maximum fecundity rate. It is natural to suppose that $\lambda > \xi$. Then the relevant dynamical equation governing population is

$$dN(t)/dt = (x(t) - \xi)N(t), \qquad \lambda \geq x(t) \geq 0 \tag{A4.15}$$

Being a state variable, $N(t)$ is a capital good. Let $v(t)$ be the accounting price of $N(t)$.

DM's problem is to choose the time profiles of $R(t)$, $x(t)$, and $C(t)$ so as to maximize $V(0)$, subject to the dynamical equations (A4.5), (A4.11), and (A4.15). The initial conditions for the optimization problem for DM are $S(0)$, $M(0)$, and $N(0)$. Because DM enjoys only partial control in the present case (unlike the model in Section (A4.3), $N(0)$ here is not a choice variable for DM), the maximum value of $V(0)$ that is attainable here is smaller than the maximum value of $V(0)$ achievable in the model economy of Section (A4.3).

We assume, as in the previous exercise, that $r > \delta$, that $S(0) = S_2$, and that $M(0)$, $N(0) > 0$. It can shown that, conditional on the economy surviving, the optimum policy tends in the long run to a stationary state.[122] The stationary optimum satisfies equations (A4.13a-b) and (A4.13d-g), but not equation (A4.13c). In the latter's place we have two equations:

$$x = \xi \tag{A4.16a}$$

$$\delta = \{U(C) - p[C-(1-b-c)AR^aM^bN^{-(a+b)}]\}/v \tag{A4.16b}$$

Equation (A4.16a) says that in a stationary state the birth rate equals the death rate. Equation (A4.16b) is a generalization of the Sidgwick-Meade Rule. Population is a stock here, which means that to identify the right balance between population at different moments of time, the social discount rate must equal the rate of return on population in its role as a capital good. The equation says that at the optimum stationary state δ equals the social value of the marginal productivity of population number (human capital here; but see Sect. (A4.5)). Unlike equation (A4.13c) population isn't allowed by DM to rise to the point where average well-being equals the difference between average consumption and the marginal productivity of human capital. It is kept larger than the difference because of the discounting that is applied to future well-being. Denote all economic variables in the

122 The transition to the stationary state may involve holding $x(t)$ at one of its two extreme values, λ and 0, for a period of time.

optimum stationary state by a double dagger (††). It follows that the expression for inclusive wealth in the optimum stationary state is

$$W^{\dagger\dagger} = p^{\dagger\dagger}M^{\dagger\dagger} + q^{\dagger\dagger}S^{\dagger\dagger} + v^{\dagger\dagger}N^{\dagger\dagger} \tag{A4.17}$$

In a stationary state net accumulation of produced capital is zero and the size of the population remains constant. The net regeneration rate of natural capital is also zero. Nature, $S^{\dagger\dagger}$, appears as a fixed factor in production in the optimum stationary state. That completes our justification for using the model of Section 9 to global estimates of optimum population (Sect. 11)).[123]

A4.5 Qualifications and Extensions

In Sections A4.3-A4.5 we assumed that resources are managed in a socially optimum manner. Our findings there are meant only to be a benchmark against which to evaluate the world as we know it. We recognised in the text of this essay that the biosphere remains in large measure an open-access resource, and we have seen why and how population can overshoot under these circumstances (Appendix 2). But when population grows large relative to the size of other capital goods, demand for goods and services produced by open-access resources becomes greater than their ability to supply them (Dasgupta, Mitra, and Sorger, 2018). In the context of the biosphere, the ecological footprint rises above 1. In time the productivity of natural capital begins to diminish. In Section 10 we read evidence which says that in recent decades the biosphere has indeed been diminishing in its productivity. Interpreted in terms of Figure 8, the stock of natural capital should be read as now being to the left of S^{**}, even while moving further to the left. Unless collective action is taken by citizens in our model economy, the stock of natural capital would be expected in time to cross S_1, which is a tipping point. That's a point citizens would fervently wish to avoid. (In Appendix 5 we sketch an imperfect economy with those features.) Mathematicians call S_1 a separatrix, meaning that opportunities open to the society when the stock of natural capital falls below S_1 are very different from those that are open to it when the stock of natural capital is above S_1. Above S_1 the biosphere is resilient; if permitted, it is able to regenerate. But it loses resilience once the stock falls below S_1.

Our model economy is stylized, but it displays possibilities of societal catastrophes in a simple manner. The Earth system is vastly more complex. It is a mosaic of dynamical systems of differing spatial reach and speed, and harbours a variety of separatrixes. Steffen et al. (2018), for example, have offered a sketch of the sequence of tipping points that probably lie in wait for humanity if carbon emissions are not abated.

There are however further problems in modelling the biosphere. Even two thousand years ago, when global population was under 250 million and per capita income was only a bit above a dollar a day, it would have been reasonable to treat humanity as a separate entity from the biosphere. Today it is no longer possible to do so. We are much engaged in transforming the biosphere, by both

123 A formal proof of the equivalence between inclusive wealth and social well-being when population change is endogenous was constructed by Arrow, Dasgupta, and Mäler (2003b). The proof is reproduced in Appendix 5.

creating biomass and destroying it. We have to imagine humanity as at once a constituent of the biosphere and an entity separate from it. No doubt that's a stretch, but it is possible to do it without running into contradictions. The way to avoid contradiction while retaining equation (A4.3) is to postulate that the parameters of the regeneration function $G(S)$ are themselves functions of S. For example, we could suppose that the parameter T in equation (A4.1) is a declining function of S, making the biosphere a more fragile entity for human needs even as it declines. Under this reckoning the fact that the global ecological footprint has exceeded 1 for some decades means the G-function has shrunk to the point where we have unleashed the Sixth Extinction. That puts in perspective the proposal by Wilson (2016) that we should leave a vital half of the biosphere alone.

There is deficiency also in the production side of our proto-type model. The dynamical system represented by equations (A4.5), (A4.10)-(A4.11), and (A4.15) does not include the production and dissemination of knowledge, nor does it include health. Both are features of human capital. Our model needs to be augmented by knowledge and health sectors. Two kinds of knowledge may be distinguished. First, there are the knowledge and skills that are embodied in the individual person. Second, there is publicly available knowledge (the arts and sciences), which can be accessed by anyone with the requisite knowledge and skills. Both have been much studied in the economics literature, as has health. Regarding publicly available knowledge, we can imagine that the factor A in the production function in equation (A4.10) is not a parameter but a function of human capital (scientists), produced capital (laboratories), and natural capital (raw materials from Nature). Extending our model in that direction is then a routine exercise.

Regarding the knowledge and skills embodied in a person and the person's state of health, we could combine education and health by an aggregate index h, interpreted as the per capita human capital in the economy. It follows that if N is the population size, hN is the aggregate human capital in the economy. Equation (A4.10) would now read as $Q = AR^aM^b\ (h\,N)^{(1-a-b)}$.

A person's human capital h is itself a function of human capital (teachers), produced capital (schools and teaching material), and natural capital (raw materials from Nature). Extending our model in that direction is also a routine exercise.

Appendix 5

Inclusive Wealth and Social Well-Being

To confirm that inclusive wealth is the basis on which we should both evaluate policies and assess whether the economy is on a sustainable development path, consider an arbitrary date $t \geq 0$. We call the person conducting the exercize the "social evaluator", or "evaluator" for short. The model economy is that of Section A4. The birth rate is a choice variable (Sect. A4.4). Our analysis extends to Generation-Relative Utilitarianism, but as it would involve additional notation, we construct the wealth/well-being equivalence theorem by supposing that the evaluator is a Total Utilitarian. The future consists of all dates $u > t$.

Let $V(t)$ denote social well-being at t. Extending equation (A4.12) to the arbitrary initial date t, we have

$$V(t) = {}_{u=t}\int^{\infty}[N(u)U(C(u))]e^{-\delta(u-t)}\mathrm{d}u, \qquad \delta > 0 \tag{A5.1}$$

The evaluator recognizes that the economy is governed by equations (A4.10)-(A4.11) and (A4.15). We assume she has an understanding of the motivations of the various agents in the economy, which means she can forecast consumption $C(u)$, the birth rate $x(u)$, and the rate at which natural capital is used, $R(u)$; and she can do that for all $u \geq t$. $C(u)$, $x(u)$, and $R(u)$ are "control variables," to borrow the terminology from control theory. The evaluator is thereby able to make a forecast of the pair of functions $\{C(u),N(u)\}$ in equation (A5.1). She needs to make a forecast of the pair for all $u \geq t$ because $\{C(u),N(u)\}$ are factors determining $V(t)$. Below we should how what data she needs and what assumptions she needs to make to do that.

The welfare economics we develop here has a wide reach. It is *not* assumed that $\{C(u),N(u)\}$ maximize $V(t)$. The following analysis includes the possibility that the economy suffers from all manners of institutional distortions and normative weaknesses. The social evaluator's forecast of the economy is the "status quo" around which she studies small changes. Let us see how she can do that.

There are three capital goods: produced capital (M), natural capital (S), population (N). They are the state variables of the dynamical system, namely, equations (A4.10)-(A4.11) and (A4.15). For ease of notation, we write the triplet as $\underline{K}$. Thus $\underline{K} = \{M,S,N\}$. Because the social evaluator's forecast is based on her understanding of the dynamics governing the economy and the motivations of the various economy's agents, she is able to reason with counterfactuals (otherwise her forecast wouldn't be backed by an understanding of the economy). She can make a forecast of the economy's future if the capital goods the economy had inherited at t were to have been other than what they are. Her knowledge and beliefs about the economy enable her therefore to make a forecast of $C(u),N(u)\}$ for every value of $\underline{K}(t)$ she can imagine. We call forecasts based on all counterfactuals a resource allocation mechanism, or RAM.

Definition 1: A *resource allocation mechanism* (RAM) is a many-one mapping from the set of all possible $\underline{K}$s at t into the set of possible pair of functions $\{C(u),N(u)\}$ for all $u \geq t$, satisfying equations (A5.5), (A5.10)-(A5.11) and (A5.15).[124]

$V(t)$ is a function of both the prevailing RAM and the initial stock of capital goods $\underline{K}(t)$. That's because by *Definition* 1, the social evaluator's forecast $\{C(u),N(u)\}$ is a function of the prevailing RAM and the initial stock of capital goods $\underline{K}(t)$. The quality of the economy's social environment, what we have called *enabling assets* (Sect. 3.2) are thus reflected in RAM.

To denote the dependence of $\{C(u),N(u)\}$ on both $\underline{K}(t)$ and the prevailing RAM would make the notation cumbersome. So in what follows we suppress the fact that $\{C(u),N(u)\}$ is also a function of the prevailing RAM. Moreover, in order to keep the account to its bare essentials, we assume also that the prevailing RAM is time autonomous.[125] That means $V(t)$ is not an explicit function of t. That means

$$V(t) = V(\underline{K}(t)) \tag{A5.2}$$

Combining equations (A5.1) and (A5.2) we have

$$V(\underline{K}(t)) = \int_{u=t}^{\infty}[N(u)U(C(u))]e^{-\delta(u-t)}\mathrm{d}u \tag{A5.3}$$

We confirm (Sects. (A5.2)-(A5.3)) that the expression for $V(t)$ on the right-hand side of equation (A5.3) is most suitable for policy evaluation, while the expression on the left-hand side is most suitable for assessing the sustainability of economic performance.

RAM is the most significant item in the reasoning underlying the wealth/well-being equivalence theorem. The social evaluator uncovers RAM from the dynamical equations driving the economy. To illustrate, consider the socio-ecological system governed by equations (A4.5), (A4.10)-(A4.11) and (A4.15) in Appendix 4. The state variables of the system are S (the stock of renewable natural resources), K (the stock of produced capital), and N (population size). The social evaluator would be expected not only to be familiar with the equations, she would also be expected to uncover the behavioural rules that govern the choice of the control variables R (harvest rate), C (per capita consumption), and x (birth rate). For simplicity of exposition, we imagine that consumption is subject to the behavioural rule postulated in the Keynesian "consumption function," which says that aggregate consumption is an increasing function of GDP. (In applied work aggregate consumption is frequently taken to be proportional to GDP.) A possible behavioural rule for harvesting natural capital would be $R(t) = R(Q(t))$, in which the harvest rate is an increasing function of GDP (which in turn is,

124 Defining RAM as a many-one mapping isn't restrictive here because production possibilities in our model define a convex set and individuals are assumed to be identical. If either assumption is dropped RAM should be defined as a many-*many* mapping. As an example, imagine RAM is the mapping from $\underline{K}(t)$s to the set of intertemporal competitive equilibria. Those equilibria cannot be guaranteed to be unique. Note though that if RAM is a many-many mapping, it's definition should be accompanied by a rule that selects from the multiple outcomes.

125 In its general form, the wealth/well-being equivalence theorem includes non-autonomous RAM's. See Dasgupta and Mäler (2000).

of course, a function of the harvest rate itself).[126] And finally, she could appeal to models in economic demography which assume that the birth rate is a declining function of the value of time, which in turn could be regarded to be an increasing function of GDP. In that case $x(t)$ is a decreasing function, $x(Q(t))$.

Introducing these behavioural rules on equations (A4.5), (A4.10)-(A4.10), and (A4.15) gives us, for all $u \geq t$,

$$Q(u) = AR(u)^a M(u)^b N(u)^{(1-a-b)}, \quad A > 0, \quad 0 < a, b, 1\text{-}a\text{-}b < 1 \tag{A5.4}$$

$$dS(u)/du = -T + rS(u)(1-S(u)/L) - R(Q(u)) \tag{A5.5}$$

$$dM(u)/du = AR^a(u)M^b(u)N^{(1-a-b)}(u) - N(u)C(Q(u))/N(u)) \tag{A5.6}$$

$$dN(u)/du = (x(Q(u))/N(u)) - \xi)N(u), \qquad \lambda \geq x(u) \geq 0 \tag{A5.7}$$

Equations (A5.4)-(A5.7) define a fully determined system. The state variables at date u are $S(u)$, $M(u)$, and $N(u)$. The social evaluatior can unravel the economy's RAM at t by using the dynamical equations (A5.4)-(A5.7). Applying the behavioural rules to the economy's dynamics, she is able to make a forecast of its future trajectory.

Definition 2: A RAM is *Ideal* if it maximizes $V(\underline{K}(t))$.

In the theory of dynamic programming, $V(\underline{K}(t))$ under the Ideal RAM is called the Value Function. The Ideal RAM in Section (A4.3) is time autonomous. Recall from Section (A3.2) that the Ideal RAM tends in the long run to the stationary state satisfying equations (A4.13a-b), (A4.13d-g), and (A4.16).

A5.1 Two Types of Change

Let Δ denote a *small change* to the economy. We call the change a perturbation. By "small" we mean small relative to the economy's size, say in relation to GDP. We are interested here in a small change to $\underline{K}(t)$.[127] We write the perturbation to $\underline{K}(t)$ as $\Delta\underline{K}(t)$. $\Delta\underline{K}(t)$ should be interpreted as a reassignment of capital goods among the economy's activities. The altered stock, $\underline{K}(t)+\Delta\underline{K}(t)$, gives rise to a perturbed future, which the social evaluator is able to forecast. Her task is to study the perturbation's contribution to social well-being.

We write the perturbation to $V(t)$ as $\Delta V(\underline{K}(t))$. In the calculations that follow, I assume that the economy is at a value of $\underline{K}(t)$ at which $V(\underline{K}(t))$ is differentiable. This is certainly a false assumption to make if $\underline{K}(t)$ happens to be a separatrix of the dynamical system being tracked by RAM. But the risk we are going wrong is nil because in familiar socio-ecological systems, such as the one represented by equations (A5.4)-(A5.7), separatrices are contained in a manifold of a smaller dimension than that of space of $\underline{K}(t)$s. If, on the other hand, by pure fluke the economy is at a separatrix at t, the social evaluator will have to read from the dynamical system as to which stability regime the socio-economy

[126] The circularity in the causal chain is not a cause for worry, for we are looking at equilibrum harvest rates. National carbon emission rates, for example, have been found to be proportional to GDP.

[127] Below we shown how the wealth/well-being equivalence theorem, which we prove presently, should be restated if the change being considered is not small.

will move into at t.

From equation (A5.2) we have

$$\Delta V(t) = \Delta V(\underline{K}(t)) = V(\underline{K}(t)+\Delta\underline{K}(t)) - V(\underline{K}(t)), \tag{A5.8}$$

which on using equation (A5.1) yields

$$\Delta V(\underline{K}(t)) = {}_{u=t}\int^{\infty}\{\Delta[N(u)U(C(u))]\}e^{-\delta(u-t)}\mathrm{d}u \tag{A5.9}$$

The perturbed integral in equation (A5.9) can be decomposed to yield

$$\Delta V(\underline{K}(t)) = {}_{u=t}\int^{\infty}[U(C(u))\Delta N(u) + N(u)\Delta U(C(u))]e^{-\delta(u-t)}\mathrm{d}u \tag{A5.10}$$

As the change is small and V is differentiable, equation (A5.10) can be expressed as

$$\Delta V(\underline{K}(t)) = {}_{u=t}\int^{\infty}[U(C(u))\Delta N(u) + N(u)U_C(C(u))\Delta C(u)]e^{-\delta(u-t)}\mathrm{d}u \tag{A5.11}$$

But

$$\Delta V(\underline{K}(t)) = [\partial V(\underline{K}(t))/\partial M(t)]\Delta M(t) + [\partial V(\underline{K}(t))/\partial N(t)]\Delta N(t) + [\partial V(\underline{K}(t))/\partial S(t)]\Delta S(t) \tag{A5.12}$$

Equations (A5.11) and (A5.12) are equivalent ways of expressing $\Delta V(\underline{K}(t))$.

Define

$$\partial V(\underline{K}(t))/\partial M(t) = p''(t) \tag{A5.13a}$$

$$\partial V(\underline{K}(t))/\partial S(t) = q''(t) \tag{A5.13b}$$

$$\partial V(\underline{K}(t))/\partial N(t) = v''(t) \tag{A5.13c}$$

$p''(t)$, $q''(t)$, and $v''(t)$ are accounting prices of $M(t)$, $S(t)$, and $N(t)$, respectively. They reflect the marginal contribution of the respective capital goods to social well-being under the prevailing RAM.

Using equations (A5.13a-c) in equation (A5.12),

$$\Delta V(\underline{K}(t)) = p''(t)\Delta M(t) + q''(t)\Delta S(t) + v''(t)\Delta N(t) \tag{A5.14}$$

Define

$$W(t) = W(\underline{K}(t)) = p''(t)M(t) + q''(t)S(t) + v''(t)N(t) \tag{A5.15}$$

$W(t)$ is the worth of the economy's capital goods when valued at their accounting prices. That's inclusive wealth. Equation (A5.15) says that the change to inclusive wealth owing to the perturbation to $\underline{K}(t)$ is

$$\Delta W(\underline{K}(t)) = p''(t)\Delta M(t) + q''(t)\Delta S(t) + v''(t)\Delta N(t) \tag{A5.16}$$

Combining equations (A5.14) and (A5.16), we have

$$\Delta V(\underline{K}(t)) = \Delta W(\underline{K}(t)) \tag{A5.17}$$

Equation (A5.17) gives us the

Wealth/Well-Being Equivalence Theorem: A perturbation to an economy that increases social well-being raises inclusive wealth. Similarly, a perturbation that decreases social well-being lowers inclusive wealth.

We may now distinguish two types of perturbations to $\underline{K}(t)$ and use the wealth/well-being equivalence theorem to show that each has a favoured method of analysis.

A5.2 Policy Analysis

Imagine that a small policy change is proposed at t. We interpret the proposal to be an investment project. If accepted, the project would require a reallocation of capital goods among various activities at t (e.g. deploying drills for cutting earth from other uses to the project). The future trajectory of the economy will be slightly different from what it would be if the project were not to be undertaken. The social evaluator's task is to evaluate the project. This is the arena of social cost-benefit analysis. She recommends the project if $\Delta V(\underline{K}(t)) > 0$ (unless a project variant were found to be even better) but rejects it if $\Delta V(\underline{K}(t)) < 0$.

For notational ease we now write the base of the natural logarithm (e) interchangeably with "*exp*" (for "exponential"). Thus for any number y, we write e^y and $exp(y)$ interchangeably. From equation (A4.15) of Section A4,

$$N(u) = N(t)exp\{{}_t\int^u[x(\tau]d\tau\}, \qquad (A5.18)$$

which means

$$\Delta N(u) = \Delta\{N(t)exp\{{}_t\int^u[x(\tau]d\tau\} = N(t)(\Delta x(t))exp\{{}_t\int^u[x(\tau]d\tau\} \qquad (A5.19)$$

Equations (A4.4), (A5.17)-(A5.19) imply

$$\Delta V(\underline{K}(t)) = {}_{u=t}\int^{\infty}\{[N(t)(\Delta x(t))exp[{}_t\int^u[x(\tau]d\tau]U(C(u)) + N(t)exp\{{}_t\int^u[x(\tau]d\tau\}U_C(C(u))\Delta C(u)\}e^{-\delta(u-t)}du \qquad (A5.20)$$

Equation (A5.20) is the social cost-benefit rule that should be applied for evaluating the project. Large projects can be evaluated by integrating the expression on the right hand side of equation (A5.20) over changes in C and N that would be occasioned by the project.

Ideally the social evaluator will continue her search for investment projects until $\underline{K}(t)$ has been so reallocated among various activities that no further project could enhance social well-being. Imagine then that the economy is so well governed that it enjoys the Ideal RAM. Using equation (A5.6) we express the optimality condition as

$${}_{u=t}\int^{\infty}\{[N(t)(\Delta x(t))exp[{}_t\int^u[x(\tau]d\tau]U(C(u)) + N(t)exp[{}_t\int^u[x(\tau]d\tau]U_C(C(u))\Delta C(u)\}e^{-\delta(u-t)}du = 0 \qquad (A5.21)$$

Equation (A5.21) is the Sidgwick-Meade Rule in a fully dynamic setting.

Equation (A5.21) is the present-value rule (Sect. 3.2). It says the evaluator should use two social discount factors - one with which to evaluate the perturbation to the birth rate at t (for which the discount factor for date u is $N(t)U(C(u))exp[-\delta(u-t)+{}_t\int^u[x(\tau]d\tau])$, the other to evaluate the perturbation caused by the project to consumption (the discount rate at date u for which is $N(t)exp\{-\delta(u-t)+{}_t\int^u[x(\tau]d\tau\}U_C(C(u)))$. The corresponding social discount rates are the percentage rates of change of the two discount factors. They are related by the optimality conditions of the Ideal Ram.

Assume the prevailing RAM is Ideal. Equations (A4.25a-c) and (A5.13a-c) say that $p^{\dagger\dagger}(t) = p''(t)$, $q^{\dagger\dagger}(t) = q''(t)$, and $v^{\dagger\dagger}(t) = v''(t)$.

A5.3 Sustainability Analysis

Consider a short interval of time Δt that begins at t. We write the change in $\underline{K}(t)$ over the interval as $\Delta\underline{K}(t)$. Equation (A5.15) then says

$$\Delta W(\underline{K}(t)) = p''(t)[dM(t)/dt]\Delta M(t) + q''(t)[dN(t)/dt]\Delta N(t) + v''(t)[dS(t)/dt]\Delta S(t) \qquad (A5.22)$$

Equations (A5.16) and (A5.22) say that social well-being increases in the short interval of time [t,

$t+\Delta t$] if and only if inclusive wealth increases in that same short interval of time. Equation (A5.22) is the criterion the social evaluator should use to assess the economy's performance over the period [t, $t+\Delta t$]. She can extend her assessment to any interval of time by integrating the small perturbations over the the interval.

A5.4 Inclusive Wealth and the Substitutibility Among Capital Goods

An economy's inclusive wealth is a weighted sum of the stocks of all the capital goods it possesses (eq. (A5.15)). The weights are their respective accounting prices. The index is a linear function of the stocks. That feature may give the impression (e.g. Daly et al., 2007) that the wealth/well-being equivalence theorem is built implicitly on the hypothesis that the various capital stocks are perfect substitutes for one another in production.

The impression is mistaken. Nowhere in the derivation of the wealth/well-being equivalence theorem was there any mention of the substitutibility of one capital good for another. But as equations (A5.13a-c) make clear, accounting prices are themselves functions of the stocks of capital goods. The extent to which various capital goods substitute for one another in production are reflected in the structure of accounting prices. And there may be little-to-no substitution possibilities between key forms of natural capital and produced capital, or for that matter any other form of capital. Suppose, to use the socio-ecological system of equations (A5.4)-(A5.7) as an example, the resource base is near its tipping point T. Crossing it would be very costly in terms of social well-being, because output in due course would perforce decline to zero. It follows from equation (A5.13c) that the accounting price of natural capital, $v''(t)$, would be so large than any further dimunition of its stock would lower inclusive wealth dramatically. The index would signal the precariousness of the situation. This and several other common misconceptions of the wealth/well-being equivalence theorem are discussed in Arrow et al. (2007).

A5.5 GDP and the Short Run

We have confirmed that economic evaluation involves wealth comparisons, over time and in the choice of policies. Economic progress should be read as growth in inclusive wealth. Nevertheless, in national economic accounts the measure most commonly taken to correspond to social well-being is gross domestic product (GDP), which is the market value of the flow of final goods and services in a country in a given year. The index is a measure of economic activity. In any given year if Z is aggregate private consumption (i.e. total consumer spending), I is gross investment in produced capital, G is the sum of government expenditures, and Ex and Im are the market value of exports and imports, respectively, then

$$\text{GDP} = Z + I + G + Ex - Im \qquad \text{(A5.23)}$$

In a market economy GDP is also the sum of domestic wages, salaries, profits, interests, rents, and government income net of taxes (the output has to reach *somebody's* hands!). That means GDP and gross domestic income are two sides of the same coin.

If the economy is closed to trade with the outside world, $Ex = Im = 0$, and equation (A5.23) reduces to

$$\text{GDP} = Z + I + G \qquad \text{(A5.24)}$$

Equation (A5.24) is a reasonable approximation for economies that trade little and is exact for the world economy, which has no outside world to trade with.

GDP shouldn't be used to evaluate long run prospects. Among other things, GDP ignores the depreciation of capital assets. The measure was created and designed for a different purpose from economic evaluation. Estimating the magnitude of economic activity became necessary in the Great Depression of the 1930s, when some 25 per cent of working-age people in Europe and the US were recorded as being unemployed and a corresponding proportion of factories and resources lay idle (Kuznets, 1941). In the years following the Second World War, when reconstruction of Europe and the Far East and economic development in what were previously European colonies became a matter for economic policy, GDP assumed its role as the measure of economic progress. I don't know why that transfer of purpose from the short to the long run came about.[128]

But even as you go beyond GDP, you find yourself returning to it. The index remains essential in short-run macroeconomics. It allows economists to estimate the gap between the economy's potential output and actual output and is useful too for studying household and corporate behaviour. As public goods and services are typically supplied by government (local or national), the government requires funds to finance them. If the resources are to be obtained from taxes, there has to be sufficient income in the economy *to* tax. So Finance Ministers are drawn to GDP for revenue. As a criterion for evaluating short run economic policy, GDP has served admirably, but ignoring capital depreciation is indefensible in economic evaluation involving the long run.

Here is a sectoral illustration of what goes wrong. Repetto et al. (1989) and Vincent et al. (1997) estimated the decline in forest cover in Indonesia and Malaysia, respectively. They found that when the decline is included, national accounts look different: net domestic saving rates turn out to have been some 20-30 per cent lower than recorded saving rates. In their work on the depreciation of natural resources in Costa Rica, Solorzano et al. (1991) found that the depreciation of three resources — forests, soil, and fisheries — amounted to about 10 per cent of GDP and over a third of domestic saving.

That GDP can mislead badly when it is put to use in sustainability and policy analyses is now well known, which is why the measure has a hard time these days among the general public and non-governmental organizations. Hardly a month goes by before another publication bearing the title, "Beyond GDP", makes an appearance. Nevertheless, GDP growth is likely to remain a measure of

128 Lewis (1988) offers an account of the historical roots of modern development theory, but doesn't question that GDP is the object of normative interest. Coyle (2014) contains a history of the evolution of the ways in which GDP has been measured. She explains why seemingly small changes in the way GDP is measured can and have brought about large changes in estimates.

economic progress in official eyes. Here are two reasons why:

GDP and Global Competition

GDP is the market value of final goods and services. Those goods and services can be deployed so as to gain advantage in the international sphere. Never mind that a country may be enjoying a large GDP by depleting its natural resources, ruining the environment, even running rough-shod over indigenous populations for access to land and minerals; GDP can be (and is) used by governments as a strategic weapon in a world where nations compete against one another for economic and political influence. Not only does a nation's status in the world rise if it enjoys high rates of growth of GDP, a large GDP enables a nation to tilt the terms of trade with the rest of the world to its advantage. History is replete with examples that demonstrate the strategic advantages of GDP growth.

GDP and Employment

There is another systemic problem in modern industrial societies that makes GDP pivotal. Economists have yet to discover ways to manage the macroeconomy in which GDP is de-linked from recorded employment. That GDP needs to keep rising if employment is not to decline is a view that appears to be shared by economists and decision makers, be they Keynesians or otherwise. Politicians and media commentators become deeply anxious when spending on consumer goods shows signs of stalling. We are encouraged to think that to consume is to contribute to the social good. It is more than an irony that short run macroeconomic reasoning is wholly at odds with accumulation of inclusive wealth.

Appendix 6

Valuing Freedom of Choice

It was shown in Section 3.2 that Capabilitarianism is an expression of Utilitarianism. The argument drew on the commonality between choosing capabilities and purchasing options in the market for securities. Such choices reflect a desire for flexibility in the face of uncertainty about what the future holds. Freedom to choose cannot discriminate among alternative choice sets unless there is a machinery for valuing the objects of choice. Utilitarianism provides that machinery.

In Appendix 3 Rawlsian primary goods were interpreted as objects that best serve to protect and promote an individual's well-being no matter what the state of the world turns out to be. To commit oneself behind the veil of ignorance to a fair allocation of primary goods is to promote choice once the veil is lifted. The wealth/well-being equivalence theorem tells us that Rawlsian primary goods should be regarded as the ingredients of inclusive wealth. In this Appendix I construct a formal model to illustrate the value of freedom, using expected well-being as the criterion of choice.[129]

Imagine that people are to choose their individual life plans. We call a set of life plans an opportunity set. Consider a person who is faced with a *set* of opportunity sets. He may choose a life plan from any of them, meaning that he compares life plans from the union of opportunity sets available to him. He could of course choose a life plan that appears to him to be the best. We are interested in situations where he will not want to do that. So we study situations where he chooses his life plan from the available set of opportunity sets of life plans in a two-stage setting. In the first stage the individual chooses (rationally!) an opportunity set from those available to him. In the second stage he chooses a life plan from the opportunity set he had selected in the first stage. The reason the person may not necessarily want to commit himself in the first stage to a life plan is that he knows he will acquire information in the second stage about the value (to him) of the various life plans available to him. He knows life plans can go awry if chosen too early; they can prove to be bad bets. Choosing an opportunity set is a way of keeping his options open. For the moment we imagine that the available set of opportunity sets is exogenously given. Once we study that situation we can extend the model to include the more interesting case where an investment or binding commitment made before the first stage determines the available set of opportunity sets in the first stage.

Consider then someone deliberating over alternative sets of life plans. Without loss of generality we suppose that there are two opportunity sets available to him. He may choose at most one. The sets are indexed by a and b and are labelled as Y_a and Y_b. The sets need not be disjoint, there may be common elements among them, but cases of interest are those where neither set is a subset of the other. Let y_a, a life plan, be a generic element of Y_a. Similarly, let y_b, also a life plan, be a generic

---

129 Kreps (1979) calls the value of freedom a preference for flexibility of choice. He notes that the analysis below remains valid if expected well-being is replaced by state-dependent well-beings, but observes that informed justification of that desire would require a lot more than axioms on the way opportunity sets are to be ordered.

element of Y_b.[130]

That future contingencies are uncertain means that the worth of a life plan to the individual is uncertain. So we let the random variable $\tilde{\varepsilon}$ reflect that uncertainty and imagine that the person is to choose an opportunity set before observing the realization of $\tilde{\varepsilon}$. That's the first stage. For simplicity, assume that after he has chosen an opportunity set, the uncertainty resolves itself (the true value of $\tilde{\varepsilon}$ is revealed) and the individual then proceeds to select a life plan from the opportunity set he had chosen. That's the second stage.

Consider Y_a. Let $U(y_a,\varepsilon)$ be the person's well-being if he were to choose y_a from Y_a and ε is the realization of $\tilde{\varepsilon}$. In the second stage the value of $\tilde{\varepsilon}$ is realized. Denote by $y_a*(\varepsilon)$ the person's best life plan in Y_a should ε be realized. $y_a*(\varepsilon)$ is the best life plan conditional on ε being the realization of $\tilde{\varepsilon}$.[131] So he values Y_a in terms of his uncertain well-being under the optimum life-plan function $y_a*(\varepsilon)$. For concreteness, we imagine that choice under uncertainty involves maximizing expected well-being.[132] In that case the value he would attach to Y_a is $E[U(y_a*(\varepsilon), \tilde{\varepsilon})]$, where E is the expectation operator. In contrast, if he were to choose a life plan from Y_a at the first stage, he would choose a life plan in ignorance of the true value of $\tilde{\varepsilon}$. We denote the optimum choice in that case by y_a*. Thus y_a* maximizes $E[U(y_a, \tilde{\varepsilon})]$ on Y_a. Notice that y_a* is not conditioned on ε. Where $y_a*(\varepsilon)$ is life plan as a function of ε, y_a* is a life plan, period.

The key point to note now is that

$$E[U(y_a*(\varepsilon), \tilde{\varepsilon})] > E[U(y_a*, \tilde{\varepsilon})] \qquad \text{(A6.1)}$$

Inequality (A6.1) says that by keeping his options open, the individual raises his expected well-being. The difference between the two figures, $E[U(y_a(\varepsilon),\tilde{\varepsilon})]$-$E[U(y_a*,\tilde{\varepsilon})]$, is the value to him of keeping his options open.

The inequality holds regardless of the individual's attitude to risk. To confirm, note that risks are not being compared in the example. The risk remains the same whether the person keeps his options open or chooses not to. The inequality rests on the fact that the individual can make a more discriminating choice if he waits to receive relevant information before choosing his life plan.

Write $G(a) = E[U(y_a(\varepsilon),\tilde{\varepsilon})]$. $G(a)$ is the value the individual attaches to the opportunity set Y_a. An identical reasoning yields $G(b)$, which is the value the person attaches to the opportunity set Y_b. In contrast to Capabilitarianism he is able to rank Y_a and Y_b no matter what life plans they contain. That means they are comparable by the person even if neither is a subset of the other. Y_a is worth more to

130 This very stylized version can be extended in any number of ways: The sets can be imagined to evolve over time (later additions being constrained by earlier choices; Appendix 3); the choices are made by others at the earliest stages of one's life (hopefully on the person's behalf!); and so on. The analysis in the text holds.

131 Formally, $y_a*(\varepsilon)$ is any element in Y_a that maximizes $U(y_a,\varepsilon)$. In many contexts $y_a*(\varepsilon)$ would be unique. If it isn't, the individual would be indifferent between and may as well choose any one of them.

132 It may be that the probabilities in the exercise are entirely subjective.

the person than Y_b if and only if $G(a) > G(b)$. An abiding attraction of the broad class of Utilitarian theories is that it traces the value of freedom of choice to the objects over which that freedom is to be exercised.

Waiting can be costly. So the cost of waiting until the second stage has to be compared to the benefit of keeping options open. From inequality (A6.1) we conclude that if the cost is Π, it is worth keeping options open if $E[U(y_a*(\varepsilon), \tilde{\varepsilon})]-\Pi > E[U(y_a*, \tilde{\varepsilon})]$, but not otherwise.

Investments create opportunities. We now extend the model to a three-stage setting. In stage-0 the individual chooses whether to make an investment (or a choice is made on his behalf) that will enhance his opportunity sets.[133] There follow stages 1 and 2 as above. It is now a trivial matter to extend the line of reasoning we have followed to include the decision to invest, followed by choice of an opportunity set, which in turn is followed by choice of a life plan from the opportunity set that was chosen in the first stage. In Rawls' theory choosing the principles of justice are commitments citizens would rationally make behind the veil of ignorance. The veil is the setting of stage-0.

133 The economics of human capital, for example, studies situations where the investment is in education or health.

References

Arrhenius, G. (2000), "An Impossibility Theorem for Welfarist Axiologists," *Economics & Philosophy*, 16(2), 247-266.

Arrhenius, G. (2005), "Superiority in Value," *Philosophical Studies*, 123(1/2), 97-114.

Arrhenius, G., J. Ryberg, and T. Tännsjö (2017), "The Repugnant Conclusion," *Stanford Encyclopedia of Philosophy*, Online (revised 16 January 2017).

Arrow, K.J. (1965), *Aspects of the Theory of Risk-Bearing* (Helsinki: Yrjö Jahnssonin säätiö).

Arrow, K.J. (1973), "Rawls' Principle of Just Savings," *Swedish Journal of Economics*, 75, 4(4), 323-335.

Arrow, K.J. (1981), "Optimal and Voluntary Income Distribution," in S. Rosefielde, ed., *Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Berson* (Cambridge: Cambridge University Press).

Arrow, K.J. (1995), "A Note On Flexibility and Freedom," in K. Basu, P. Pattanaik, and K. Suzumura, eds., *Choice, Welfare, and Development* (Oxford: Clarendon Press).

Arrow, K.J. (1999), "Discounting, Morality, and Gaming," in P.R. Portney and J.P. Weyant, eds., *Discounting and Intergenerational Equity* (Washington, DC: Resources for the Future).

Arrow, K.J. (2000), "Observations on Social Capital," in P. Dasgupta and I. Serageldin, eds., *Social Capital: A Multifaceted Perspective* (Washington, DC: World Bank).

Arrow, K.J. and P.S. Dasgupta (2009), "Conspicuous Consumption, Inconspicuous Leisure," *Economic Journal*, 119(541), F497-F516.

Arrow, K.J., P. Dasgupta, and K.-G. Mäler (2003a), "Evaluating Projects and Assessing Sustainable Development in Imperfect Economies," *Environmental & Resource Economics*, 26(4), 647-685.

Arrow, K.J., P. Dasgupta, and K.-G. Mäler (2003b), "The Genuine Savings Criterion and the Value of Population," *Economic Theory*, 21(2), 217-225.

Arrow, K.J., P. Dasgupta, L.H. Goulder, G. Daily, P. Ehrlich, G.M. Heal, S. Levin, K.-G. Maler, S. Schneider, D. Starrett, and B. Walker (2004), "Are We Consuming Too Much?" *Journal of Economic Perspectives*, 18(3), 147-172.

Arrow, K.J., P. Dasgupta, L.H. Goulder, G. Daily, P. Ehrlich, G.M. Heal, S. Levin, K.-G. Maler, S. Schneider, D. Starrett, and B. Walker (2007), "Consumption, Investment, and Future Well-Being: Reply to Daly et al.," *Conservation Biology*, 21(5), 1363-1365.

Arrow, K.J., P. Dasgupta, L.H. Goulder, K.J. Mumford, and K. Oleson (2012), "Sustainability and the Measurement of Wealth," *Environment and Development Economics*, 17(3), 317-355.

Arrow, K.J., P. Dasgupta, L.H. Goulder, K.J. Mumford, and K. Oleson (2013), "Sustainability and the Measurement of Wealth: Further Reflections," *Environment and Development Economics*, 18(4), 504-516.

Arrow, K.J. and A. Fisher (1974), "Preservation, Uncertainty and Irreversibility," *Quarterly Journal of Economics*, 88(3), 312-19.

Arrow, K.J. and M. Kurz (1970), *Public Investment, the Rate of Return and Optimal Fiscal Policy* (Baltimore: Johns Hopkins University Press).

Arrow, K.J. and M. Priebsch (2014), "Bliss, Catastrophe, and Rational Policy," *Environmental and Resource Economics*, 58(4), 491-509.

d'Aspremont, C. and L. Gevers (1977), "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies*, 44(2), 199-209.

Atkinson, A.B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory*, 2(3), 244-263.

Baland, J.-M. and J.-P. Platteau (1996), *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* (Oxford: Clarendon Press).

Balasubramanian, R. (2008), "Community Tanks vs Private Wells: Coping Strategies and Sustainability Issues in South India," in R. Ghate, N.S. Jodha, and P. Mukhopadhyay (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press).

Barbier, E.B. (2011), Scarcity and Frontiers: How Economies Have Developed Trhough Natural Resource Exploitation (Cambridge: Cambridge University Press).

Barrett, S. (2003), *Environment and Statecraft: The Strategy of Environmental Treaty Making* (Oxford: Oxford University Press).

Barrett, S. (2012), "Credible Commitments. Focal Points, and Tipping: The Strategy of Climate Treaty Design," in R.W. Hahn and A. Ulph, eds., *Climate Change and Common Sense* (Oxford: Oxford University Press).

Barrett, S. and A. Dannenberg (2012), "Climate Negotiations Under Scientific Uncertainty," *Proceedings of the National Academy of Science*, 109(43), 17372-17376.

Barrett, S. and A. Dannenberg (2014), "Sensitivity of Collective Action to Uncertainty About Climate Tipping Points," *Nature Climate Change*, 4, 36-39.

Barrett, S., A. Dasgupta, P. Dasgupta, N. Adger, J. Anderies, J. von den Bergh, C. Bledsoe, J. Bogaarts, S. Carpenter, F.S. Chapin, A.-S. Crepin, G. Daily, P. Ehrlich, C. Folke, N. Krautsky, E. Lambine, S.A. Levin, K.-G. Maler, R. Naylor, K. Nyborg, S. Polasky, M. Scheffer, J. Shogren, P.S. Jorgensen, B. Walker, and J. Wilen (2018), "Shifting Fertility and Consumption Norms to Safeguard the Global Commons," Discussion Paper, Beijer Institute of Ecological Economics, Stockholm.

Barry, B. (1990), "Introduction," in *Political Argument* (Berkeley, CA: University of California Press), a re-issue of Barry (1965) with a new introduction.

Beach, T. and S. Luzzader-Beach (2018), "Out of the Soil: Soil (Dark Matter Biodiversity) and Societal "Collapses" from Mesoamerica to the Mesopotamia and Beyond," in P. Dasgupta, P.H. Raven, and A. McIvor, eds., *Biological Extinction: New Perspectives* (Cambridge: Cambridge University Press), forthcoming 2019.

Beard, S. (2018), "Why Fairness May Require the Use of a Pure Time Discount Rate," Unpublished Notes, Cenre for the Study of Existential Risk, University of Cambridge.

Behrens, K.G. (2012), "Moral Obligations Towards Future Generations in African Thought," *Journal of Global Ethics*, 8(2-3), 179-191.

Blackorby, C. and D. Donaldson (1982), "Ratio-Scale and Translation-Scale Full Interpersonal Comparability Without Domain Restrictions: Admissible Social-Evaluation Functions," *International Economic Review*, 23(2), 249- 268.

Blackorby, C. and D. Donaldson (1984), "Social Criteria for Evaluating Population Change," *Journal of Public Economics*, 25(1-2), 13-34.

Blackorby, C., W. Bossert, and D. Donaldson (1997), "Critical-Level Utilitarianism and the Population-Ethics Dilemma," *Economics and Philosophy*, 13(2), 197-230.

Bledsoe, C. (1994), "Children are Like Bamboo Trees: Potentiality and Reproduction in Sub-Saharan Africa," in K. Lindahl-Kiessling and H. Landberg, eds., *Population, Economics Development and the Environment* (Oxford: Oxford University Press).

Bohn, H. and C. Stewart (2015), "Calculation of Population Externality," *American Economic Journal: Economic Policy*, 7(2), 61-87.

Bongaarts, J. and J. Casterline (2013), "Fertility Transition: Is Sub-Saharan Africa Different?," *Population an Development Review*, 39(1), 153-168.

Boonin, D. (2014), *The Non-Identity Problem and the Ethics of Future People* (New York: Oxford University Press).

Boserup, E. (1965), *The Conditions of Agricultural Growth: The Economics of Agrarian Change Under Population Pressure* (London: Allen & Unwin).

Bostrom, N. (2002), "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Journal of Evolution and Technology*, 9(1).

Bourdieu, P. (1984), *Distinction: A Social Critique of the Judgement of Taste* (London: Routledge and Kegan Paul).

Brander, J.A. and M.S. Taylor (1998), "The Simple Economics of Easter Island: A Ricardo-Malthus Model of Renewable Resource Use," *American Economic Review*, 88(1), 119-138.

Brock, W.A. and A. Xepapadeus (2018), "Modeling Coupled Climate, Ecosystems, and Economic Systems," in P. Dasgupta, S. Pattanayak, and V.K. Smith, eds, Handbook of Environmental Economics, Vol. IV (Amsterdam: Elsevier).

Brook, B.W., N.S. Sodhi, and C.J.A. Bradshaw (2008), "Synergies Among Extincton Drivers Under Global Change," *Trends in Ecology and Evolution*, 23(8), 453-460.

Broome, J. (1992), *Counting the Cost of Global Warming* (London: White Horse Press).

Broome, J. (1996), "The Welfare Economics of Population," *Oxford Economic Papers*, 48(2), 177-93.

Broome, J. (1999), *Ethics Out of Economics* (Cambridge: Cambridge University Press).

Broome, J. (2004), *Weighing Lives* (Oxford: Oxford University Press).

Broome, J. (2012), *Climate Matters: Ethics in a Warming World* (New York: W.W. Norton).

Butzer, K.W. (2012), "Collapse: Environment and Society," *Proceedings of the National*

Academy of Science, 109(10), 3632-3639.

Caldwell, J.C. (1981), "The Mechanisms of Demographic Change in Historical Perspective," *Population Studies*, 35(1), 5-27.

Caldwell, J.C. (1982), *The Theory of Fertility Decline* (New York: Academic Press).

Cardinale, B.J., J.E. Duffy, A. Gonzalez, D.U. Hooper, C. Perrings, P. Venail, A. Narwani, G.M. Mace, D. Tilman, D.A. Wardle, A.P. Kinzig, G.C. Daily, M. Loreau, J.B. Grace, A. Larigauderie, D.S. Srivastave, and S. Naeem (2012), "Biodiversity Loss and Its Impact on Humanity," *Nature*, 486: 59-67.

Carlson, E. (2007), "Higher Values and Non-Archimedean Additivity," *Theoria*, 73(1), 3-27.

Casey, J. (2009), *After Lives: A Guide to Heaven, Hell, and Purgatory* (Oxford: Oxford University Press).

Cavendish, W. (2000), "Empirical Regularities in the Poverty-Environment Relationships of Rural Households: Evidence from Zimbabwe," *World Development*, 28(11), 1979-2003.

Ceballos, G., A.H. Ehrlich, and P.R. Ehrlich (2015), *The Annhilation of Nature: Human Exhaustion of Birds and Mammals* (Baltimore, MD: The Johns Hopkins University Press).

Chichilnisky, G. and G.M. Heal (1998), "Economic Returns from the Biosphere," *Nature*, 391, 629-630.

Cline, W.R. (1992), *The Economics of Global Warming* (Washington, DC: Institute for International Economics).

Cochrane, S.H. and S.M. Farid (1989), "Fertility in Sub-Saharan Africa: Analysis and Explanation," World Bank Discussion Paper No. 43, World Bank, Washington DC.

Cohen, G.A. (1989), "On the Currency of Egalitarian Justice," *Ethics*, 99(4), 906-944.

Cohen, J.E. (1995a), "Population Growth and Earth's Human Carrying Capacity," *Science*, 269(21 July), 341-346.

Cohen, J.E. (1995b), *How Many People Can the Earth Support?* (New York: W.W. Norton).

Colbert, E. (2013), "The Lost World: Parts I and II," *The New Yorker* (16 December and 23&30 December Issues).

Cowen, T. (1989), "Normative Population Theory," *Social Choice and Welfare*, 6(1), 33-44.

Cowie, C. (2016), "Does the Repugnant Conclusion Have Any Probative Force?", *Philosophical Studies*, online publication: DOI 10.1007/s 11098-016-0844-7.

Coyle, D. (2014), *GDP: A Brief but Affectionate History* (Princeton, NJ: Princeton University Press).

Crist, E., C. Mora, and R. Engelman (2017), "The Interaction of Human Population, Food Production, and Biodiversity Protection, *Science*, 36(6334), 260-264.

Crutzen, P.J. and E.F. Stoermer (2000), "The Anthropocene," *Global Change Newsletter*, 41, 17-18.

Daily G, ed. (1997), *Nature's Services: Societal Dependence on Natural Ecosystems* (Washington, D.C.: Island Press).

Daly, H.E., B. Czech, D.L. Trauger, W.E. Rees, M. Grover, T. Dobson, and S.C. Trombulak (2007), “Are We Consuming Too Much – For What?,” *Conservation Biology*, 21. DOI:10.1111j.1523-1739.2007.00779.x.

Daly, M.C., A.J. Oswald, D. Wilson, and S. Wu (2011), "Dark Contrasts: The Paradox of High Rates of Suicide in Happy Places," *Journal of Economic Behaviour & Organization*, 80(3), 435-442.

Dasgupta, A. and P. Dasgupta (2017), "Socially Embedded Preferences, Environmental Externalities, and Reproductive Rights," *Population and Development Review*, 43(3), 405-441.

Dasgupta, A. and P. Dasgupta (2018), "Population Overshoot," in K. Bykvist and T. Campbell, eds. *Oxford Handbook of Population Ethics* (Oxford: Oxford University Press), forthcoming 2019.

Dasgupta, P. (1969), "On the Concept of Optimum Population," *Review of Economic Studies*, 36(3), 295-318.

Dasgupta, P. (1974a), "On Optimum Population Size," in A. Mitra, ed., *Economic Theory and Planning: Essays in Honour of A.K. Dasgupta* (Delhi: Oxford University Press).

Dasgupta, P. (1974b), "On Some Alternative Criteria for Justice Between Generations," *Journal of Public Economics*, 3(4), 405-423.

Dasgupta, P. (1982), *The Control of Resources* (Cambridge, MA: Harvard University Press).

Dasgupta, P. (1989), "Population Size and the Quality of Life," *Proceedings of the Aristotelian Society*, *Supplementary Volumes*, 63, 23-40.

Dasgupta, P. (1993), *An Inquiry into Well-Being and Destitution* (Oxford: Clarendon Press).

Dasgupta, P. (1994), "Savings and Fertility: Ethical Issues," *Philosophy & Public Affairs*, (Spring), 99-127.

Dasgupta, P. (1995), "Population, Poverty, and the Local Environment," *Scientific American,* 272(2), 40-45.

Dasgupta, P. (2004), *Human Well-Being and the Natural Environment* (Oxford: Oxford University Press), 2nd ed.

Dasgupta, P. (2005a), "Regarding Optimum Population," *Journal of Political Philosophy*, 13(4), 414-442. Reprinted in J. Fishkin and R. Goodin, eds., *Population and Political Theory - Philosophy, Politics and Society, 8* (Chichester: Wiley-Blackwell), 2010.

Dasgupta, P. (2005b), "Three Conceptions of Intergenerational Justice," in H. Lillehammer and D.H. Mellor, eds., *Ramsey's Legacy* (Oxford: Clarendon Press).

Dasgupta, P. (2007a), *Economics: A Very Short Introduction* (Oxford: Oxford University Press).

Dasgupta, P. (2007b), "Comments on the Stern Review's Economics of Climate Change," *National Institute Economic Review*, No. 199, 4-7.

Dasgupta, P. (2008), "Discounting Climate Change," *Journal of Risk and Uncertainty*, 37(2-3), 141-169.

Dasgupta, P. (2010), "The Place of Nature in Economic Development," in D. Rodrik and M. Rosenzweig, eds, *Handbook of Development Economics, Vol. V* (Amsterdam: North Holland).

Dasgupta, P. (2014), "Measuring the Wealth of Nations," in G. Rausser, M. Hanneman, and D. Zilberman, eds., *Annual Review of Resource Economics* (Palo Alto: CA), 17-32.

Dasgupta, P. and P. Ehrlich (2013), "Pervasive Externalities at the Population, Consumption, and Environment Nexus," *Science*, 340 (April 19), 324-328.

Dasgupta, P. and G.M. Heal (1979), *Economic Theory and Exhaustible Resources* (Cambridge: Cambridge University Press).

Dasgupta, P. and K.-G. Mäler (2000), "Net National Product, Wealth, and Social Well-Being," *Environment and Development Economics*, 5(1), 69-93.

Dasgupta, P., S. Marglin, and A. Sen (1972), *Guidelines for Project Evaluation* (New York: United Nations).

Dasgupta, P., T. Mitra, and G. Sorger (2018), "Harvesting the Commons," *Environmental and Resource Economics*, https://doi.org/10.1007/s10640-018-0221-4, 25 January.

Dasgupta, P., P.H. Raven, and A. McIvor, eds. (2018), *Biological Extinctions* (Cambridge: Cambridge University Press), forthcoming, 2019.

Dasgupta, P. and I. Serageldin, eds., (2000), *Social Capital: A Multifaceted Perspective* (Washington, DC: World Bank).

Dasgupta, P., D. Southerton, A. Ulph, and D. Ulph (2016), "Consumer Behaviour with Environmental and Social Externalities: Implications for Analysis and Policy," *Environmental and Resource Economics*, 65(2), 191-226.

Diamond, J. (2005), *Collapse: How Societies Choose to Fail or Survive* (London: Allen Lane).

Diamond, P.A. (1965), "The Evaluation of Infinite Utility Streams," *Econometrica*, 33(1), 170-177.

Diener, E., J.F. Helliwell, and D. Kahneman, eds. (2010), *International Differences in Well-Being* (Oxford: Oxford University Press).

Douglas, M. and B. Isherwood (1979), *The World of Goods: Towards an Anthropology of Consumption* (New York: Basic Books).

Douglas, M. and S. Ney (1998), *Missing Persons: A Critique of Personhood in the Social Sciences* (Berkeley: University of California Press).

Downey, S.S., W. Randall Haas Jr., and S.J. Shennan (2016), "European Neolithic Societies Showed Early Warning of Population Collapse," *Proceedings of the National Acaddemy of Science*, 113(35), 9751-9756.

Dworkin, R. (1981a), "What is Equality? Part 1: Equality of Welfare," *Philosophy & Public Affairs*, 10(3), 185-246.

Dworkin, R. (1981b), "What is Equality? Part 2: Equality of Resources," *Philosophy & Public Affairs*, 10(4), 283-345.

Edgeworth, F.Y. (1881), *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London: Kegan Paul).

Ehrlich, P.R. and A.H. Ehrlich (1981), *Extinction: The Causes and Consequences of the Disappearance of Species* (New York, NY: Random House).

Ehrlich, P.R. and A.H. Ehrlich (2008), *The Dominant Animal: Human Evolution and the Environment* (Washington DC: Island Press).

Ehrlich, P.R. and J.P. Holdren (1971), "Impact of Population Growth," *Science*, 171 (3977), 1212-1217.

English, J. (1977), "Justice Between Generations," *Philosophical Studies*, 31(2), 91-104.

Enke, S. (1966), "The Economic Aspects of Slowing Population Growth," *Economic Journal*, 76(1), 44-56.

Feeny, D., F. Berkes, B.J. McCay, and J.M. Acheson (1990), "The Tragedy of the Commons: Twenty-two Years Later", *Human Ecology*, 18(1), 1-19.

Feldman, F. (1995), "Justice, Desert, and the Repugnant Conclusion," *Utilitas*, 7(2), 189-206.

Fenichel, E.P. and J.K. Abbott (2014), “Natural Capital: From Metaphor to Measurement,” *Journal of the Association of Environmental and Resource Economics*, 1(1/2), 1-27.

Fortes, M. (1978), "Parenthood, Marriage and Fertility in West Africa," *Journal of Development Studies*, 14 (Special Issue on Population and Development), 121-149.

Freedom House (2017), *Freedom in the World 2017*. https://freedomhouse.org/report/freedom-world/freedom-world-2017.

Freeman III, A.M. (2002), *The Measurement of Environmental and Resource Values: Theory and Methods* (Washington, DC: Resources for the Future), 2nd Ed.

Galor, O. (2011), *Unified Growth Theory* (Princeton, NJ: Princeton University Press).

Ghate,R., N.S. Jodha, and P. Mukhopadhyay, eds. (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press).

Goody, J. (1976), *Production and Reproduction* (Cambridge: Cambridge University Press).

Gordon, H.S. (1954), "The Economic Theory of a Common Property Resource: The Fishery," *Journal of Political Economy*, 62(2), 124-142.

Gosseries, A. (2010), "On Future Generations' Future Rights," in J. Fishkin and R. Goodin, eds., *Population and Political Theory - Philosophy, Politics and Society, 8* (Chichester: Wiley-Blackwell).

Gottlieb, M. (1945), "The Theory of Optimum Population for a Closed Economy", *Journal of*

Political Economy, 53(3), 289-316.

Griffin, J. (1986), *Well-Being: Its meaning, measurement, and moral importance* (Oxford: Oxford University Press).

Grootaert, C. and T. van Bastelaer, eds. (2002), *The Role of Social Capital in Development: An Empirical Assessment* (Cambridge, UK: Cambridge University Press).

Guyer, J.L. (1994), "Lineal Identities and Lateral Networks: the Logic of Polyandrous Motherhood," in C. Bledoe and G. Pison, eds., *Nupitality in Sub-Saharan frica: Contemporary Anthropological and Demographic Perspectives* (Oxford: Clarendon Press).

Haberl, H., K.-H. Erb, F. Krasmann, V. Gaube, A. Bondeau, C. Plutzer, S. Gingrich, W. Lucht, and M. Fisher-Kowalski (2007), "Quantifying and Mapping the Human Appropriation of Net primary Production in Earth's Terrestrial Ecosystems," *Proceedings of the National Academy of Science*, 104(31), 12942-12947.

Hamilton, K. and M. Clemens (1999), “Genuine Saving Rates in Developing Countries,” *World Bank Economic Review*, 13(2), 333-356.

Hammond, P.J. (1976), "Equity, Arrow' Conditions and Rawls' Difference Principle," *Econometrica*, 44(3), 793-804.

Haque, A.K.E., M.N. Murty, and P. Shyamsundar, eds. (2011), *Environmental Valuation in South Asia* (Cambridge: Cambridge University Press).

Hardin, G. (1968), "The Tragedy of the Commons," *Science*, 162, 1243-1248.

Harford, J.D. (1997), "Stock Pollution, Child-Bearing Externalities, and the Social Rate of Discount," *Journal of Environmental Economics and Management*," 33(1), 95-105.

Harford, J.D. (1998), "The Ultimate Externality," *American Economic Review*, 88(1), 260-265.

Harrod, R.F. (1948), *Towards a Dynamic Economy* (London: McMillan).

Harsanyi, J. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Welfare," *Journal of Political Economy*, 63(4), 309-321.

Hartwick, J.M. (1977), "Intergenerational Equity and the Investing of Rents from Exhaustibel Resources," *American Economic Review*, 67(5), 972-974.

Haveman, R. and B. Wolf (1995), "The Determinants of Children's Attainments: A Review of Methods and Findings," *Journal of Economic Literature*, 33(4), 1829-1878.

Hector, A., B. Schmid, C. Beierkuhnlein, M.C. Cladeira, M. Diemer, P.G. Dimitrakopoulos, J.A. Finn, H. Freitas, P.S. Giller, J. Good, R. Harris, P. Högberg, K. Huss-Danell, J. Joshi, A. Jumpponen, C. Körner, P.W. Leadley, M. Loreau, A. Minns, C.P.H. Mulder, G.O'Donovan, S.J. Otway, J.S. Pereira, A. Prinz, D.J. Read, M. Scherer-Lorenzen, E.-D. Schultze, A.-S.D. Siamantziouras, E.M. Spehn, A.C. Terry, A.Y. Troumbis, F.I. Woodward, S. Yachi, and J.H. Lawton (1999), "Plant Diversity and Productivity Experiments in European Grasslands", *Science*, 286, 1123-1127.

Helliwell, J.F., R. Layard, and J.D. Sachs, eds. (2013), *World Happiness Report 2013* (New York: Earth Institute).

Henry, C. (1974), "Investment Decisions under Uncertainty: the Irreversibility Effect," *American Economic Review*, 64, 1006-1012.

Heyd, D. (1992), *Genethics: The Morality of Procreation* (Berkeley, CA: University of Calfornia Press).

Hooper, D.U., F.S. Chapin III, J.J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J.H. Lawton, D.M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Setala, A.J. Symstad, J. Vandermeer, and D.A. Wardle (2005), "Effects of Biodiversity on Ecosystem Functioning: A Consensus on Current Knowledge," *Ecological Monographs*, 75(1), 3-35.

Huemer, M. (2008), "In Defense of Repugnance," *Mind*, 117(468), 899-933.

Hurka, T. (1983), "Value and Population Size," *Ethics*, 93(4), 497-507.

Irwin, E.G., S. Gopalakrishnan, and A. Randall (2016), "Welfare, Wealth, and Sustainability," *Annual Review of Resource Economics*, 8, 77-98.

Jodha, N.S. (1986), "Common Property Resources and the Rural Poor", *Economic and Political Weekly*, 21, 1169-81.

Jodha, N.S. (2001), *Living on the Edge: Sustaining Agriculture and Community Resources in Fragile Environments* (Delhi: Oxford University Press).

Jones, C.I. and P.M. Romer (2010), "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Review: Macroeconomics*, 2(1), 224-245.

Kavka, G.S. (1982), "The Paradox of Future Generations," *Philosophy & Public Affairs*, 11(2), 93-112.

Kitzes, J., M. Wakernagel, J. Loh, A Peller, S. Goldfinger, D. Cheung, and K. Tea (2008), "Shrink and Share: Humanity's Present and Future Ecological Footprint," *Philosophical Transactions of the Royal Society, B*, 363(25 July), 467-475.

Kolm, S.-Ch. (1969), "The Optimal Production of Social Justice," in J. Margolis and H. Guitton, eds., *Public Economics* (London: Macmillan).

Koopmans, T.C. (1960), "Stationary Ordinal Utility and Impatience," *Econometrica*, 28(2), 287-309.

Koopmans, T.C. (1965), "On the Concept of Optimal Economic Growth," *Pontificiae Academiae Scientiarum Scripta Varia*, 28. Reprinted in T.C. Koopmans (1966), *The Econometric Approach to Development Planning* (Amsterdam: North Holland).

Koopmans, T.C. (1967), "Objectives, Constraints, and Outcomes in Optimal Growth Models," *Econometrica*, 35(1), 1-15.

Koopmans, T.C. (1972), "Representation of Preference Orderings over Time," in C.B. McGuire and R. Radner, eds., *Decision and Organization* (Amsterdam: North Holland).

Koopmans, T.C. (1977), "Concepts of Optimality and Their Uses," *American Economic Review*, 67(3), 261-274.

Kreps, D. (1979), "A Representation Theorem for Preference for Flexibility," *Econometrica*, 47(3), 565-577.

Kuznets, S. (1941), *National Income and Its Composition, 1919-1938* (New York: National Bureau of Economic Research).

Landes, D. (1998), *The Wealth and Poverty of Nations* (New York: W.W. Norton).

Lee, R. (2000), "A Cross-Cultural Perspective on Intergenerational Transfers and the Economic Life Cycle," in A. Mason and G. Tapines, eds., *Sharing the Wealth: Demographic Change and Economic Transfers Between Generations* (Oxford: Oxford University Press).

Lee, R. (2007), "Demographic Change, Welfare, and Intergenerational Transfers: A Global Overview," in J. Veron, S. Pennec, and J. Legare, eds., *Age, Generations and the Social Contract: The Demographic Challenge Facing the Welfare State* (Amsterdam: Springer).

Lewis, W.A. (1988), "The Roots of Development Theory," in H. Chenery and T.N. Srinivasan, eds., *Handbook of Development Economics, Vol. 1* (Amsterdam: North Holland).

Lind, R. ed. (1982), *Discounting for Time and Risk in Energy Policy* (Baltimore: Johns Hopkins University Press).

Little, I.M.D. and J.A. Mirrlees (1968), *Manual of Industrial Project Analysis in Developing Countries: Social Cost Benefit Analysis* (Paris: OECD).

Little, I.M.D. and J.A. Mirrlees (1974), *Project Appraisal and Planning for Developing Countries* (London: Heinemann).

Lomborg, B., ed, (2014), *How Much Have Global Problems Cost the World? A Scoreboard from 1900 to 2050* (Cambridge: Cambridge University Press).

Loreau, M., S. Naeem, P. Inchauti, J. Bengtsson, J.P. Grime, A. Hector, D.U. Hooper, M.A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D.A. Wardle (2001), "Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges," *Science*, 294, 804-808.

Lovejoy, T.E. and C. Nobre (2018), "Amazon Tipping Point," Editorial, *Science Advances*, 4:eaat2340.

Luce, R.D. and H. Raiffa (1957), *Games and Decisions* (New York: John Wiley).

Maddison, A. (2001), *The World Economy: A Millennial Perspective* (Paris: OECD).

Managi, S. and P. Kumar (2018), *Inclusive Wealth Report 2018: Measuring Progress Toward Sustainability* (New York: Routledge).

Marothia, D.K., ed. (2002), *Institutionalizing Common Pool Resources* (New Delhi: Concept Publishing Company).

Maskin, E.S. (1978), "A Theorem on Utilitarianism," *Review of Economic Studies*, 45(1), 93-96.

McMahan, J.A. (1981), "Problems of Population Theory," *Ethics*, 92(1), 96-127.

MEA - Millennium Ecosystem Assessment - eds., R. Hassan, R. Scholes, and N. Ash - (2005a), *Ecosystems and Human Well-Being, I: Current State and Trends* (Washington, DC: Island Press).

MEA - Millennium Ecosystem Assessment - eds., S.R. Carpenter, P.L. Pingali, E.M. Bennet, and M.B. Zurek (2005b), *Ecosystems and Human Well-Being, II: Scenarios* (Washington, DC: Island Press).

MEA - Millennium Ecosystem Assessment - eds., K. Chopra, R. Leemans, P. Kumar, and H. Simmons - (2005c), *Ecosystems and Human Well-Being, III: Policy Responses* (Washington, DC: Island Press).

MEA - Millennium Ecosystem Assessment - eds., D. Capistrano, C. Samper K., M.J. Lee, and C. Randsepp-Hearne - (2005d), *Ecosystems and Human Well-Being, IV: Multiscale Assessments* (Washington, DC: Island Press).

Meade, J.E. (1955), *Trade and Welfare* (Oxford: Oxford University Press).

Micklethwait, J. and A. Wooldridge (2000), *A Future Perfect: The Challenge and Promise of Globalization* (New York: Random House).

Mirrlees, J.A. (1967), "Optimum Growth When Technology is Changing," *Review of Economic Studies*, 34(1), 95-124.

Mokyr, J. (2002), *The Gifts of Athena: Historical Origins of the Knowledge Economy* (Princeton, NJ: Princeton University Press).

Mokyr, J. (2016), *The Culture of Growth: The Origins of the Modern Economy* (Princeton, NJ: Princeton University Press).

Mukhopadhyay, P. (2008), "Heterogeneity, Commons, and Privatization: Agrarian Institutional Change in Goa", in R. Ghate, N.S. Jodha, and P. Mukhopadhyay, eds., (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press)..

Nagel, T. (1976), "Moral Luck," *Proceedings of the Aristotelian Society*, *Supplementary Volumes*, 50, 137-151. Republished in *Mortal Questions* (Cambridge: Cambridge University Press), 1979.

Nagel, T. (1979a), *Mortal Questions* (Cambridge: Cambridge University Press).

Nagel, T. (1979b), "Death," in *Mortal Questions* (Cambridge: Cambridge University Press). Originally published in *Nous*, 1970, 4(1), 73-80.

Nagel, T. (1986), *The View From Nowhere* (New York: Oxford Univesity Press). Chapter IX republished as "Autonomy and Deontology," in S. Scheffler, ed., *Consequentialism and Its Critics* (Oxford: Oxford University Press), 1988.

Narveson, J. (1967), "Utilitarianism and New Generations," *Mind*, 76(1), 62-72.

Narveson, J. (1973), "Moral Problems of Population," *Monist*, 57(1), 62-86.

Norberg, J. (2016), *Progress: Ten Reasons to Look Forward to the Future* (London: One World).

Nordhaus, W.D. (1994), *Managing the Global Commons: The Economics of Climate Change* (Cambridge, MA: MIT Press).

Nordhaus, W.D. (2007), "A Review of the Stern Review on the Economics of Climate

Change," *Journal of Economic Literature*, 45(3), 686-702.

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press).

Ostrom, E. and T.K. Ahn, eds. (2003), *Foundations of Social Capital* (Cheltenham: Edward Elgar).

Ostrom, E., T. Dietz, N. Dolsak, P.C. Stern, S. Stonich, and E.U. Weber, eds., (2002), *The Drama of the Commons* (Washington, DC: National Academy Press).

Parfit, D. (1976), "On Doing the Best for Our Children," in M. Bayles, ed., *Ethics and Population* (Cambridge, MA: Schenkman). Reprinted in J.S. Fishkin and R.E. Goodin, eds., *Population and Political Theory*, Philosophy, Politics and Society 8 (Chichester: Wiley-Blackwell), 2010.

Parfit, D. (1982), "Future Generations: Further Problems," *Philosophy & Public Affairs*, 11(2), 113-172.

Parfit, D. (1984), *Reasons and Persons* (Oxford: Oxford University Press).

Parfit, D. (2016), "Can We Avoid the Repugnant Conclusion?" *Theoria*, 82(2), 110-127.

Perrings, C. (2014), *Our Common Heritage: Biodiversity Change, Ecosystem Services, and Human Well-Being* (Cambridge: Cambridge University Press).

Phelps, E.S. and R. Pollak (1968), "Second-Best National Savings and Game Equilibrium Growth," *Review of Economic Studies*, 35(2), 185-199.

Pimm, S.L., C.N. Jenkins, R. Abell, T.M. Brooks, J.L. Gittleman, L.N. Joppa, P.H. Raven, C.M. Roberts, and J.O. Se (2014), "The Biodiversity of Species and Their Rates of Extinction, Distribution, and Protection," *Science*, 344(6187), 1246752 (1-10).

Pinker, S. (2018), *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress* (New York: Allen Lane).

PNAS (2012), *Special Feature: Critical Perspectives on Historical Collapse*, PNAS, 109(10), 3632-3681.

Power, M.E., D. Tilman, J.A. Estes, B.A. Menge, W.J. Bond, L.S. Mills, G. Daily, J.C. Castilla, J. Lubchenco, and R.T. Paine (1996), "Challenges in the Quest for Keystones," *Bioscience*, 46(8), 609-620.

Pratt, J.A. (1964), "Risk Aversion in the Small and in the Large," *Econometrica*, 32(1-2), 122-136.

Prest, A.R. and R. Turvey (1965), "Cost-Benefit Analysis: A Survey," *Economic Journal*, 75(4), 683-735.

Putnam, R.D., with R. Leonardi and R.Y. Nanetti (1993), *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton, NJ: Princeton University Press).

Raiffa, H. (1968), *Decision Analyis: Introductory Lectures on Choice Under Uncertainty* (Reading, MA: Adddison-Wesley).

Räikkä, J. (2002), "The Repugnant Conclusion and the Welfare of Actual People," *Theoria*, 68(2), 162-169.

Ramsey, F.P. (1928), "A Mathematical Theory of Saving," *Economic Journal*, 38(4), 543-559.

Ramsey, F.P. (1931), "Epilogue" in *The Foundations of Mathematics and Other Logical Essays*, R.B. Braithwaite, ed. (London: Routledge and Kegan Paul), pp. 287-292.

Rawls, J. (1972), *A Theory of Justice* (Oxford: Oxford University Press).

Rawls, J. (1993), *Political Liberalism* (New York: Columbia University Press).

Rees, M.J. (2003), *Our Final Century* (London: Random House).

Rees, W.E. (2006), "Ecological Footprints and Biocapacity: Essential Elements in Sustainability Assessment," in J. Dewulf and H.V. Langenhove, eds., *Renewable-Based Technology: Sustainability Assessment* (Chchester, UK: John Wiley & Sons).

Rees, W.A. and M. Wackernagel (1994), "Ecological Footprints and Appropriated Carrying Capacity: Measuring the Natural Capital Requirements of the Human Economy," in A.M. Jansson *et al*., *Investing in Natural Capital: The Ecological Economics Appropriate for Sustainability* (Washington, DC: Island Press).

Repetto, R., W. Magrath, M. Wells, C. Beer, and F. Rossini (1989), *Wasting Assets: Natural Resources and the National Income Accounts* (Washington DC: World Resources Institute).

Ricketts, T.H., G.C. Daily, P.R. Ehrlich, and C.D. Michener (2004), "Economic Value of Tropical Forest to Coffee Production," *Proceedings of the National Academy of Science*, 101(34), 12579-12582.

Ridley, M. (2010), *The Rational Optimist: How Prosperity Evolves* (London: 4th Estate).

Sandel, M.J. (1982), *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press).

Scheffer, M. (2016), "Anticipating Societal Collapse: Hints From the Stone Age," *Proceedings of the National Academy of Science*, 113(35).

Scheffer, M., S.R. Carpenter, and T.M. Linton et al. (2012), "Anticipating Critical Transitions," *Science*, 338(19 October), 344-348.

Scheffler, S. (1982), *The Rejection of Consequentialism* (Oxford: Clarendon Press).

Scheffler, S. (1985), "Agent-Centred Restrictions, Rationality, and the Virtues," *Mind*, 94(375), 409-419. Reprinted in S. Scheffler, ed., *Consequentialism and Its Critics* (Oxford: Oxford University Press), 1988.

Scheffler, S. (2013), *Death & the Afterlife* (Oxford: Oxford University Press).

Schell, J. (1982), *The Fate of the Earth* (New York: Avon).

Schultz, T.W. (1961), "Investment in Human Capital," *American Economic Review*, 51(1), 1-17.

Seabright, P. (1989), "Population Size and the Quality of Life, II: Creating Persons," *Proceedings of the Aristotalian Society, Supplementary Volumes*, 63, 41-54.

Sen, A. (1970), *Collective Choice and Social Welfare* (San Francisco: Holden Day).

Sen, A. (1985), *Commodities and Capabilities* (Amsterdam: North Holland).

Sen, A. (1992), *Inequality Reexamined* (Oxford: Clarendon Press).

Sen, A. (1999), *Development as Freedom* (Oxford: Oxford University Press).

Sen, A. (2009), *The Idea of Justice* (Cambridge, MA: Havard University Press).

Sher, I. (2018), "Evaluating Allocations of Freedom," *Economic Journal*, 128(612), F65-F94.

Shyamsundar, P. (2008), "Decentralization, Devolution, and Collective Action," in R. Ghate, N.S. Jodha, and P. Mukhopadhyay, eds. (2008), *Promise, Trust and Evolution: Managing the Commons of South Asia* (Oxford: Oxford University Press).

Sidgwick, H. (1907), *The Methods of Ethics* (London: MacMillan), 7th Edition.

Sikora, R.I. (1978), "Is It Wrong to Prevent the Existence of Future Generations?" in R.I. Sikora and B. Barry, eds., *Obligations to Future Generations* (Philadelphia: Temple University Press).

Sikora, R.I. (1981), "Classical Utilitarianism and Parfit's Repugnant Conclusion: A Reply to McMahan," *Ethics*, 92(1), 128-133.

Sodhi, N.S., B.W. Brook, and C.J.A. Bradshaw (2009), "Causes and Consequences of Species Extinctions," in S.A. Levin et al., eds., *The Prineton Guide to Ecology* (Princeton, NJ: Princeton University Press).

Solorzano, R., R. de Camino, R. Woodward, J. Tosi, V. Watson, A. Vasquez, C. Villalobos, J. Jimenez, R. Repetto, and W. Cruz (1991), *Accounts Overdue: Natural Resource Depreciation in Costa Rica* (Washington, DC: World Resources Institute).

Solow, R.M. (1974a), "The Economics of Resources and the Resources of Economics", *American Economic Review*, 64(Papers & Proceedings), 1-21.

Solow, R.M. (1974b), "Intergenerational Equity and Exhaustible Resources," *Review of Economic Studies*, 41 (Symposium Issue), 29-45.

Solow, R.M. (2000), "Notes on Social Capital and Economic Performance," in P. Dasgupta and I. Serageldin, eds., *Social Capital: A Multifaceted Perspective* (Washington, DC: World Bank).

Somanathan, E. (1991), "Deforestation, Property Rights and Incentives in Central Himalaya," *Economic and Political Weekly*, 26 (Special Issue: January 26) PE37-46.

Steffen, W., J. Rockström, K. Richardson, T.M. Lenton, C. Folke, D. Liverman, C.P. Summerhayes, A.D. Barnosky, S.E. Cornell, M. Crucifix, J.F. Donges, I. Fetzer, S.J. Lade, M. Scheffer, R. Winklemann, and H.J. Schellnhuber (2018), "Trajectories of the Earth System in the Anthropocene," *Proceedings of the National Academy of Science*, August 6, https://doi.org/10.1073/pnas.1810141115.

Sterba, J.P. (1987), "Explaining Asymmetry: A Problem for Parfit," *Philosophy and Public Affairs*, 16(2), 188-192.

Stern, N.H. (2006), *The Stern Review of the Economics of Climate Change* (Cambridge: Cambridge University Press).

Suppes, P. (1987), "Maximizing Freedom of Decision: An Axiomatic Analysis," in G. Feiwel, ed., *Arrow and the Foundation of the Theory of Economic Policy* (London: McMillan).

Tännsjö, T. (2002), "Why We Ought to Accaept the Repugnant Conclusion," *Utilitas*, 14(3), 339-359.

Temkin, L.S. (1987), "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs*, 16(2), 138-187.

Thomson, J.T., D.H. Feeny, and R.J. Oakerson (1986), "Institutional Dynamics: the Evolution and Dissolution of Common Property Resource Management," in National Research Council, *Proceedings of Conference on Common Property Resource Management* (Washongton, DC: US National Academy of Sciences).

Tilman, D. (1982), *Resource Competition and Community Structure* (Princeton, NJ: Princeton University Press).

Tilman, D. (1997), "Biodiversity and Ecosystem Functioning," in G. Daily, ed., *Nature's Services: Societal Dependence on Natural Ecosystems* (Washington, DC: Island Press).

Tilman, D. and J.A. Downing (1994), "Biodiversity and Stability in Grasslands," *Nature*, 367 (27 January), 363-367.

Tilman, D. and C.L. Lehman (1997), "Habitat Destruction and Species Extinctions," in D. Tilman and P. Kareiva, eds., *Spatial Ecology: the role of space in population dynamics and interspecific interactions* (Princeton, NJ: Princeton University Press).

Tomlinson, J. (2018), "Nigerian Briefing: How Engineers Can Help Steer a Sustainable Economy," *Proceedings of the Institution of Civil Engineers - Energy*. https://doi.org/10.1680/jener.17.00024.

Turner, B.L. and A.M.S. Ali (1996), "Induced Intensification: Agricultural Change in Bangladesh with Implications for Malthus and Boserup," *Proceedings of the National Academy of Sciences*, 93(25), 12984-14991.

UNDP (1990), *Human Development Report* (New York: United Nations)

UNFPA (1995), *Programme of Action of the International Conference on Population and Development* (New York: United Nations Population Fund).

UNPD (2015), *World Population Prospects: the 2015 Revision* (New York: United Nations Population Division).

UNU-IHDP/UNEP - (prepared by A. Duraiappah et al.) - (2012), *Inclusive Wealth Report 2012: Measuring Progress Toward Sustainability* (Cambridge: Cambridge University Press).

UNU-IHDP/UNEP - (prepared by A. Duraiappah et al.) - (2014), *Inclusive Wealth Report 2012: Measuring Progress Toward Sustainability* (Cambridge: Cambridge University Press).

Veblen, T. (1925, [1899]). *The Theory of the Leisure Class: An Economic Study of Institutions*. Originally published in 1899. Reprinted (London: George Allen & Unwin), 1925.

Vincent, J.R., R.M. Ali, and Associates (1997), *Environment and Development in a Resource-Rich Economy: Malaysia Under the New Economic Policy* (Cambridge, MA: Harvard Institute for International Development).

Vosen, P. (2016), "Anthropocene Pinned Down to Post War Period," *Science*, 353(6302), 852-853.

Walker, B., A. Kinzig, and J. Langridge (1999), "Plant Attribute Diversity, Resilience, and Ecosystem Function: The Nature and Significance of Dominant and Minor Species," *Ecosystems*, 2(2), 95-113.

Waters, C.N., J. Zalasiewicz, C. Summerhayes, A.D. Barnosky, C. Poirier, A. Galuszka, A. Cerreta, M. Edgeworth, E.C. Ellis, C. Jeandel, R. Leinfelder, J.R. McNeill, W. Steffen, J. Syritski, D. Vidas, M. Wagreich, M. Williams, A. Zhisheng, J. Grineveld, E. Odada, N. Oreskes, and A.P. Wolfe (2016), "The Anthropocene is Functionally and Stratigraphically Distinct from the Holocene," *Science*, 351(6269), aad2622-(1-10).

Weisbrod, B.A. (1964), "Collective Consumption Services of Individual Consumption Goods," *Quarterly Journal of Economics*, 77(1), 71-77.

Williams, B.A.O. (1976), "Moral Luck," *Proceedings of the Aristotelian Society, Supplementary Volumes*, 50, 115-135.

Williams, B.A.O. (1985), *Ethics and the Limits of Philosophy* (London: Fontana).

Williams, B.A.O. (1993), *Shame and Necessity* (Berkeley, CA: University of California Press).

Wilson, E.O. (1992), *The Diversity of Life* (Cambridge, MA: Harvard University Press).

Wilson, E.O. (2016), *Half-Earth: Our Planet's Fight for Life* (New York: Liveright Publishing Corporation).

World Bank (2016), *World Development Indicators* (Washington DC: World Bank).

Worm, B., E.B. Barbier, N. Beaumont, D.E Duffy, C. Folke, B.S. Halpern, J.B.C. Jackson, H.K. Lotze, F. Micheli, S.R. Palumbi, E. Sala, K.A. Selkoe, J.J. Stachowicz, and R. Watson (2006), "Impacts of Biodiversity Loss on Ocean Ecosystem Services," *Science*, 314, 787-790.

Wrigley, E.A. (2004), *Poverty, Progress, and Population* (Cambridge: Cambridge University Press).

WWF - World Wildlife Fund - (2008), *Living Planet Report 2008* (Gland: WWF International).

Yaari, M. (1965), "Uncertain Lifetime, Life Insurance, and the Theory of Consumer," *Review of Economic Studies*, 32(2), 137-158.

Yamaguchi, R. (2018), "Wealth and Population Growth Under Dynamic Average Utilitarianism," *Environment and Development Economics*, 23(1), 1-18.

Yamaguchi, R., M. Sato, and K. Ueta (2016), "Measuring Regional Wealth and Assessing Sustainable Development: an application to a disaster-torn region of Japan," *Social Indicators Research*, 129)(1), 365-389.