

INTERGENERATIONAL JUSTICE AND SAVING

Clark Wolf

If current and proximate generations were to destroy the resources of the earth, then later generations would be much worse off. Would such destruction violate any requirements of justice? Many people have recently argued that the answer to this question is not obvious. First, current destruction might change the identities of the people who exist in future generations: Different people will exist if we adopt more destructive policies than would have existed otherwise. Members of the deprived generations, then, could not complain that they had been made worse off by our destructive treatment of the earth's resources since they would not have been better off—they would not be at all—had we behaved differently. Many people regard this so-called non-identity problem as surprisingly important.¹ But a second reason why some people regard it as unclear whether current destruction of the earth's resources would be unjust to future generations is that we have no widely accepted account of intergenerational justice, and most of the literature on the problem of intergenerational justice has been negative. Not only does no

widely accepted theory exist, we do not even have a substantial set of contender theories, and no agreement about what an account of intergenerational justice would look like. While this chapter will not remedy all of these gaps, it will examine a promising way to look at the problem of intergenerational justice.

I will assume that one crucial part of any theory of intergenerational justice is an account of whether and how much the present generation ought to save resources for the future rather than destroying or consuming them. The problem of just saving is not the only problem that would be treated in a full theory of intergenerational justice. But within the broader subject of intergenerational justice, the saving problem is crucial and has received scant treatment from philosophers. The problem is not just a philosophical exercise: Practical proposals to mitigate pollution, to preserve wild lands, to protect biodiversity, and to conserve resources are often framed in terms of intergenerational justice, and of our current obligations to save resources for future generations. The thought is that we have an obligation to leave later generations an earth capable of supporting their needs, and that we would violate this obligation if we were to use too much of our exhaustible resources, or if we were to destroy the earth's great ecosystems.

By "save," I do not necessarily mean consciously setting aside goods or funds for later use by future persons, or using less oil or coal so that later generations will have it. In the sense relevant here, "saving" applies to whatever resources come to be at the disposal of future persons, whether we consciously set them aside for them or not. We might save in this sense because we are unable to consume resources quickly enough to use them all up, and not because we have any concern for the welfare of future persons. The problem of saving is to some extent independent of an analysis of the institutions used to accomplish it. It might turn out, as some have argued, that free markets will adequately save for future generations, and that we need not concern ourselves with their predicament.² Or it might turn out that markets will do this badly or unreliably. If so, then we may find it appropriate to protect resources or ecosystems or institutions from the market. Such protection may be necessary to insure that future persons will have adequate resources to live autonomous lives, or to exercise fundamental human capacities, or whatever it is that justice is understood to require.

I will ignore many important issues that have occupied recent discussions of intergenerational justice, including the non-identity problem, problems

of population choice, the paradoxes of intertemporal harm, and questions about the moral standing of future persons and groups. My reason for ignoring these issues is not that I regard them as trivial; I have focused on them in other recent works.³ But in recent literature on intergenerational justice, an almost exclusive focus on these other problems combined with recent disaffection for distributive justice in general has led to an inexcusable neglect of the problem of just saving. It was not always so: The savings problem attracted the attention of many important economists and philosophers shortly after the publication of John Rawls's *Theory of Justice* in 1971. Three years after the appearance of *Theory*, Kenneth Arrow, Robert Solow, and Partha Dasgupta each published seminal papers on the problem of intergenerational distributive justice and the problem of just saving.⁴ But their conclusions were mostly negative, and work on the issue has hardly moved forward since they abandoned it.

This chapter will revive a decades-old discussion of the problem of just saving. I will argue that Arrow, Solow, Dasgupta, and Rawls himself were all led astray by a serious fault in the models they employed in thinking about the problem of intergenerational distributive justice and savings, and that work on this problem has been held back by a simple mistake. One of my aims is to explain where earlier treatments of the savings problem went wrong. But my aims are not primarily critical and my conclusions are not negative: Once we have an understanding of the central problem with the model that has been used to frame the issue of intergenerational distribution, it becomes clear that the model can be improved. The improved model, incorporating the fact of generational overlap, provides a convenient framework for understanding and evaluating alternative conceptions of intergenerational justice. Once it becomes clear that work on the problem has been hampered by faulty modeling and that it is not inherently intractable, the problem of intergenerational justice should take a prominent place in the theory of justice.

1. A SIMPLE MODEL OF INTERGENERATIONAL SAVING AND PRESENT CONSUMPTION

I begin with a simple economic model of intergenerational production, distribution, and saving. With only small variations, this model is the same

as the one used by Kenneth Arrow and Robert Solow, but it is most clearly discussed by Partha Dasgupta, whose presentation I will follow here.⁵ The model is simple, but it is tremendously useful for framing an understanding of the problem of just savings and intergenerational distribution. As I will argue in Section 4, this useful model is also deceptive. Its implicit and explicit use has led to a persistent and thorough misunderstanding of the problem of intergenerational justice and just savings.

Dasgupta asks that we consider a society that possesses a single, non-deteriorating, homogeneous commodity K , the stock of which at time t is K_t . Assume that time is discrete, and takes on the non-negative integer values. Assume that population is constant from one generation to the next. Each generation lives for only one period, and a new equinumerous generation arrives as soon as they die. With these assumptions in place, we can normalize to one individual per period and consider the consumption and savings rates for each generation separately. We can also describe the stream of consumption and savings from one generation to the next. For example, suppose that generation t consumes C_t of the commodity, where $K_t \geq C_t \geq 0$. Then $(K_t - C_t)$ is carried on to the next period, where it becomes $\lambda(K_t - C_t)$. Here λ represents the rate at which the saved portion appreciates as it is carried over to the next generation. This gives us a basic accumulation equation:⁶

$$\begin{aligned}
 K_{t+1} &= \lambda(K_t - C_t) \\
 K_t \geq C_t \geq 0 & \quad \text{for } t = 0, 1, 2, 3, \dots \\
 K_0 & \text{ given.}
 \end{aligned}
 \tag{1}$$

An intergenerational allocation or consumption stream is series of allocations $(C_t) = (C_0, C_1, C_2, \dots, C_n, \dots)$, with C_n being the amount consumed by generation n . We can refer to the set F of all sequences (C_t) that satisfy Equation 1. This set contains all *feasible* intertemporal allocation sequences. If we assume, as does Dasgupta, that $\lambda > 1$, then whatever is saved by one generation grows for the benefit of the following generation. But of course there are cases for which this assumption is not justified. If $\lambda = 1$, then the amount of K left over for the next generation does not grow. In this case, the rate at which the resource is depleted is exactly the rate at which it is consumed by the early generations, so in this special case, $K_{t+1} = (K_t - C_t)$ and (C_t) will be a decreasing sequence as long as there is some consumption at all. And in the

worst case, if $\lambda < 1$, then the resource decays over time and the first generation may need to consume it quickly before it spoils. In what follows, I will accept Dasgupta's assumption that saved resources appreciate in value—that $\lambda > 1$.

We can represent the savings rate S_t —the rate at which any generation t saves for the future—as follows:

$$S_t = (K_t - C_t)/K_t \quad \text{for } t = 0, 1, 2, \dots \quad (2)$$

In this model, each generation chooses how much to save for the benefit of later generations. If they save more, later generations will have more to consume. If they save less, then future generations will be poorer. But in this model, any saving by a generation implies that there is less to consume now. If consumption is an indicator or a partial measure of well-being, then any saving at all is worse for those who save. This feature of the model will be important in what follows.

Is there a minimum saving rate required by justice, such that saving less would violate obligations to future generations? If the first generation gratuitously consumes all of the available resource, the following generations will be left with nothing at all. On the other hand, if earlier generations save too much, they might needlessly impoverish themselves for the sake of the future. A theory of intergenerational justice would, among other things, provide a principled method for choosing an appropriate rate of saving.

2. SAVING FOR THE FUTURE AND INTERGENERATIONAL JUSTICE: SUSTAINABILITY

When is an intergenerational distribution unjust? We might begin by weeding out those intergenerational patterns of consumption and savings that seem most undesirable. Dasgupta himself suggests that we should eliminate those schemes that tend toward zero consumption, and that it would “not be very just” for early generations to leave nothing for later ones.⁷ He suggests that justice requires that intergenerational consumption and savings should at least be sustainable over time. But the possibility of sustainable production and consumption will obviously depend on the sizes of λ and K_t . Suppose we define “sustainability” as the condition that $K_{t+1} = K_t$. This means that the amount of K available to later generations is exactly the same as the amount

available to earlier generations. In order to obtain this condition, we must have $K_t = \lambda(K_t - C_t)$. Solving for C_t , we get:

$$C_t = (K_t - K_t/\lambda) \quad (\text{For } \lambda > 0) \quad (3)$$

So $(K_{t+1} - K_t/\lambda)$ is the steady state consumption package for any period. As long as each generation's consumption meets the formally weaker condition that

$$C_t \leq (K_t - K_t/\lambda) \quad (4)$$

the consumption stream (C_t) will be nondecreasing. If we put this in terms of necessary savings instead of permissible consumption, we find that the steady state savings ratio for the economy is given by $S = 1/\lambda$. Savings plans that consistently consume more than the sustainable rate and consistently save at a rate lower than $1/\lambda$ will tend toward zero consumption in the long run.

Sustainability in the sense described above might seem reasonable as a minimal requirement of intergenerational justice. We find something similar to it in Sir John Hicks's discussion of income in *Value and Capital*:

The purpose of income calculations in practical affairs is to give people an indication of the amount they can consume without impoverishing themselves. Following out this idea, it would seem that we ought to define a man's income as the maximum value that he can consume during a week, and still be expected to be as well off at the end of the week as he was at the beginning. Thus when a person saves, he plans to be better off in the future, when he lives beyond his income, he plans to be worse off. Remembering that the practical purpose of income is to serve as a guide for prudent conduct, I think it is fairly clear that this is what the central meaning must be.⁸

Hicks's aim was to describe income in terms of prudent individual saving, but one could think of an account of intergenerational justice as a prescription for understanding generational income in terms of prudent saving and investment for future generations. In extreme circumstances, it may be necessary for an individual to consume more than income can sustain over time, but such overconsumption comes at the price of future poverty. In the individual case, this price is paid by the same person who consumed too much too early, but in the intergenerational case the cost of early overconsumption may be passed on to later generations. Although self-interest may

dictate prudent consumption over a lifetime, it will not always dictate intergenerationally prudent consumption.

3. RAWLS ON SAVING AND INTERGENERATIONAL JUSTICE

Sustainability in the sense described may be a minimal requirement. But perhaps justice requires more, in some circumstances. Another way to approach the problem of intergenerational justice and savings would be to consider which feasible intertemporal distribution—which allocation sequence (C_n) , member of the set F of feasible intertemporal allocations—persons would choose from an original position from which they were uncertain to which generation they will belong. According to Rawls's well-known theory, the principles of justice are those that free and rational persons would choose, under appropriate circumstances, those identified in the description of an "original position," as principles to govern the basic social and political institutions of society. The choice is rendered impartial by a veil of ignorance that blinds them to any knowledge of themselves that would make it possible for them to tailor the choice of principles arbitrarily to favor themselves. Parties to this original position choice will not know their race, position in society, religious commitments, abilities, or anything else that could introduce partial bias into the choice of principles. The restrictions of this original position choice force parties to be impartial between their interests and the interests of others.

Rawls argues that parties to the original position would choose two principles to govern the basic institutions of society. The first asserts that each person is to have an equal right to the most extensive total system of basic liberties compatible with a similar system for all. The second principle specifies which kinds of inequality are tolerable, and is in two parts: social and economic inequalities are to be arranged so that they are both (1) attached to offices and positions open to all under conditions of fair equality of opportunity, and (2) tolerable only when they work to the greatest benefit of the least advantaged members of society. This last requirement, the Difference Principle, is surely the most controversial element of Rawls's theory, and has attracted wide critical discussion. Rawls calls the conception of justice embodied in these two principles the "conception of justice as fairness," and his

earlier work was devoted to the defense of this conception of justice against perfectionist, utilitarian, and libertarian alternatives.⁹

But what does justice require in the intergenerational case? Rawls is much less specific in what he says about intergenerational justice, and what he does say is confusing. Rawls's veil of ignorance does extend to generational membership, so the parties to the original position do not know in which generation, or at what stage of social and economic development they will live. This ignorance ensures that their choice will not be partial in favor of earlier generations over later ones or vice versa. In the first edition of *Theory*, Rawls wrote:

The parties do not know to which generation they belong or, what comes to the same thing, the stage of civilization of their society. They have no way of telling whether it is poor or relatively wealthy, largely agricultural or already industrialized, and so on. 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 on the assumption that all other generations are to save at the same rates. That is, they are to consider their willingness to save at any given phase of civilization with the understanding that the rates they propose are to regulate the whole span of accumulation. In effect, then, they must choose a just savings principle that assigns an appropriate rate of accumulation to each level of advance.¹⁰

In 1971 Rawls stipulated that the motive to save was that parties to the original position choice were assumed to care for their descendants. This was unsatisfactory for many reasons that I can only mention and briefly identify here. First, it is otherwise unmotivated and ad hoc. In the original position, the choice of principles is supposed to be based on self-interest and mutual disinterest, not altruistic concern for others. And Rawls specifies that parties to the original position choice are "mutually disinterested." The interests of contemporaries are represented in the original position by the extended self-interest of the parties to the choice behind the veil of ignorance, but Rawls's ad hoc assumption implies that the interests of future generations are represented in quite a different way. Second, different people have different degrees of concern for their descendants, and there may be no uniquely appropriate level of concern. Because of this, Rawls's ad hoc stipulation renders it difficult and perhaps impossible to specify what principle of savings free rational agents in the original position would choose. Third, some

free and rational people will have no descendants. It would seem to follow that parties to the original position should understand themselves to represent some individuals who have an interest in saving for future generations and others who have no such interest. This plurality of interests will lead parties to incompletely represent the interests of later generations. Finally, Rawls uses questionable language to describe the motivational assumption of concern for descendants. He wrote that parties to the original position choice should understand themselves as representing "heads of households," and that they should imagine themselves to be "fathers" deciding "how much they should set aside for their sons and grandsons by noting what they would believe themselves entitled to claim of their fathers and grandfathers."¹¹ As critics have pointed out, this formulation is sexist and begs many questions about the structure of families. Given ample evidence that the interests of different family members diverge, the assumption that the interests of the "head of the household" accurately represent the interests of families is clearly unacceptable. Since it is usually assumed that the head of a household is a man, it has been argued that Rawls's model might lead to the systematic neglect of the interests of women, the elderly, and children.¹²

In more recent work, Rawls has revised his account of the choice of principles for intergenerational savings. While he still regards the appropriate principle of savings as the one that would be chosen from the original position, he has recently argued that parties to the choice should understand themselves to be choosing that principle they would want earlier generations to have adopted:

[S]ince society is a system of cooperation between generations over time, a principle for savings is required. Rather than imagine a (hypothetical and nonhistorical) direct agreement between all generations, the parties can be required to agree to a savings principle subject to the further condition that they must want all previous generations to have followed it. Thus 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.¹³

This change is an improvement since, appropriately interpreted, it can avoid all of the problems listed above. Most important, it recovers concern for fu-

ture generations as a function of the extended self-interest of parties to the original position choice.

But what principles would such parties choose? Rawls says very little to identify what a just intertemporal distribution would look like. He does say that he regards the difference principle as inappropriate for the intergenerational case. He adds that the purpose of saving is not just to make future generations more affluent, but to secure justice. Thus savings "may stop once just (or decent) basic institutions have been established. At this point real saving (net additions to real capital of all kinds) may fall to zero; and existing stock only needs to be maintained, or replaced, and nonrenewable resources carefully husbanded for future use as appropriate."¹⁴ According to Rawls, parties to the original position will choose a principle of intergenerational justice that requires savings at stages when too little wealth exists to secure just basic institutions, but maintains existing resources once justice is secure.

Although the changes since the first edition of *A Theory of Justice* are improvements, Rawls's more recent comments about intergenerational justice raise new problems. His remarks are vague; the new account does not raise the issue of changing population size. Perplexingly, Rawls says nothing specific about the rate at which earlier generations should save for the benefit of later ones. Most seriously, such saving appears contrary to other Rawlsian commitments: Rawls argues that it is only poorer generations that have an obligation to save—generations that are too poor to secure just institutions for themselves, but whose members hope that their saving will enable their descendants to be better off. But on Rawls's view their saving implies a sacrifice on their part, and the beneficiaries of this sacrifice will be future generations who will be better off as a result. In other contexts, Rawls insists that it is inappropriate for those who are worse off to make sacrifices for the benefit of those who are better off. Here he recommends such sacrifice as a requirement of intergenerational justice. This is inconsistent with Rawls's insistence that parties to the original position choice would employ maximin reasoning in the choice of principles of justice.

In what follows I will consider a super-Rawlsian theory of intergenerational justice: the intertemporal application of the difference principle. This theory is, as one commentator put it, "*plus Rawlsien que le Rawls*."¹⁵ But as I will argue, it has been discarded for the wrong reasons.

4. HOW TO BE MORE RAWLSIAN THAN RAWLS

Rawls's suggestion is that we should choose that savings principle that we would wish the preceding generation to have adopted if we were ignorant of our own generational membership. One might worry that this guideline would promote too much saving: If we are only concerned with the saving rate of our predecessors, we would naturally prefer that they save as much as possible. The more that earlier generations have saved, the better off our own generation will be, since the alternative to saving, in this model, is consumption. Higher savings rates imply costs for those who save, but those costs are not obviously counted on Rawls's scheme. They might be implicit in the concern that descendants have for the welfare of their predecessors—the concern of children for their parents, for example. From a generation-blind original position, it is not difficult to imagine that parties choosing principles of justice would take such concern into account even between generations that are remote. But a move such as this would re-introduce many of the problems that plagued Rawls's earlier (1971) account of the motivation to save.

We should instead, then, compare savings schemes that cover the "entire span of accumulation," taking into account the interests of individuals in each generation. We should choose the savings scheme that we would want our predecessors to have adopted regardless of our generational membership. But does a unique saving scheme exist that satisfies this vague condition? Sustainability might be considered a minimal requirement for intergenerational justice, and it is plausible to think that parties to the original position choice might begin by restricting themselves to savings schemes that are sustainable. Surely we would wish previous generations to have saved at least at the sustainable rate. But if so, then they would be left with a choice among sustainable schemes: some of the remaining schemes would be more demanding than others. How would a free rational agent choose among more and less demanding principles of intergenerational saving without knowing which generation would be the agent's?

One might expect that such an agent would begin by comparing the problem of intergenerational justice to the problem of intragenerational justice. If, as Rawls suggests, such agents would find the difference principle

attractive in the intragenerational case, then they would consider the Intergenerational Difference Principle (IDP) as a solution to the problem of just savings. The IDP holds that just saving should maximize benefits to those who are worst off, regardless of their generational membership.

Most commentators have discarded the difference principle as a requirement on intergenerational distribution, and Rawls rejects the intergenerational application of the difference principle. But we might wonder why he does so: If we accept the argument for the difference principle as a requirement of justice, uniquely acceptable to free and equal contractors choosing from behind a veil of ignorance, then it might be natural to suppose that such contractors would choose the difference principle as an intergenerational requirement as well. And given the general argument for the difference principle and given Rawls's account of "reasonableness" and "fairness," it may be difficult to see what other principle would do as a requirement of intergenerational distributive justice, especially if we accept these other features of Rawls's account. Such saving is inconsistent with Rawls's (dubious) claim that parties to the original position must use maximin reasoning in their choice of principles. Reasonable persons, says Rawls, will be willing to comply only if the terms of the savings scheme are fair. Rawls has insisted elsewhere that fairness requires, among other things, that a system of cooperation must not improve the situation of those who are better off at cost to those who are worst off. It is difficult to see why this should not apply intergenerationally if it is accepted in the intragenerational case. Rawls should either explain why this conception of fairness does not apply in the intergenerational case, or he should accept it as a general principle for distributive justice. I will refer to the view that the difference principle should apply intergenerationally as the "super-Rawlsian" view.

Rawls explains his own reasons for rejecting the difference principle as a requirement in the intergenerational case: He argues that the IDP would inappropriately prohibit saving, since earlier generations would necessarily be worse off than the later generations who would benefit from savings. The model described in Section 1 makes it easy to see why one might come to this conclusion: Given an initial stock K_t , the first generation would have an obligation to save at least K_t/λ . This is the amount of saving required at the sustainable rate, since saving less would increase their consumption at cost

to generation $t+1$, which would then be worse off than generation t . But if any generation t saves at a rate greater than the steady-state replacement rate, then the following generations will be better off at the expense of generation t . Such savings will involve sacrifices for the first generation that saves, and this first generation will not benefit from savings of any earlier generation. Their uncompensated sacrifice will then benefit those who are better off—clearly prohibited by the difference principle. So such saving is prohibited as unjust to the first generation and savings can never get started.

So it has seemed to most of those who have considered the problem. Most are in agreement that an intergenerational difference principle would allow no positive saving to benefit the future. Robert Solow writes:

... the max-min criterion is ... at the mercy of the initial conditions. If the initial capital stock is very small, no more will be accumulated and the standard of living will remain low forever. This result follows from the principle itself. Capital could be accumulated and consumption increased subsequently, but only at the cost of a lower standard of living for earlier generations. It is part of Rawls' general argument for the max-min criterion that we should regard earlier and later generations as facing each other contemporaneously when the social contract is being drawn up. But then it is hardly surprising that the preferred strategy refuses to make some people poorer than others in order to make the others richer, just because the first group can be given the essentially arbitrary label of "earlier."¹⁶

Kenneth Arrow concurs:

It is pretty obvious that [a straightforward transposition of the Rawls' maximin criterion to the intertemporal context] would lead to zero savings in every generation, for there is no way to compensate the first generation for any savings they might do, and they would be worse off than any of their successors.¹⁷

And in a (1994) article, Dasgupta writes:

We may conclude that an economy that was poor to begin with would remain in poverty if it were at the mercy of the intergenerational maxi-min principle.¹⁸

And Rawls makes the same claim in both editions of *Theory*. In the second (1999) edition, he writes:

[W]hen the difference principle is applied to the question of savings over generations, it entails either no savings at all or not enough savings to improve social circumstances sufficiently so that all the equal liberties can be effectively exercised.¹⁹

As I will show in the next section, Solow, Arrow, Dasgupta, and Rawls are all wrong about this. They were led to a false conclusion by the limitations of the model described in Section 1 above, or, in Rawls's case, by assumptions that mirror those limitations. It may, as Arrow puts it, seem "pretty obvious" that an intergenerational difference principle would imply no savings, but it is not true.

5. INTERGENERATIONAL SAVINGS:

FROM A REPEATED GAME TO A MULTISTAGE GAME

The model described in Section 1 is employed, with some small variations, by all of the economists quoted above at the end of Section 4. I have argued that a similar model is implicit in Rawls's remarks about intergenerational justice and savings and is associated with his rejection of the difference principle in the intergenerational case. But the model is flawed: It incorporates an implicit assumption that generations do not interact or overlap. Like mayflies and paulo verde beetles, all members of the Rawlsian population disappear after consuming their share, and the next generation comes on stage all at once. In his earlier papers, Dasgupta claimed that this assumption introduces no special problems, but he was wrong: When the model is adjusted to accommodate generational overlap, it becomes clear that saving can indeed benefit the first generations to engage in it, and that self-interested motives to save can promote positive intergenerational saving. It can even be shown that intergenerational saving is sometimes achievable as a strategic equilibrium among narrowly rational self-interested agents: The adoption of an appropriate intergenerational saving plan may be strongly Pareto superior to sustainability and beneficial to literally everyone involved. It follows that the IDP will permit or even require positive savings rates in a wide variety of circumstances, and that it has been rejected for the wrong reasons. In refining the model described above, we transform it from a repeated game in which each stage is independent from any other, to a "stage game" in which each

“play” reflects the choices of different individuals whose ages and interests are different.²⁰

Consider a revision on the model described in Section 1. Suppose that we begin with a population of one-hundred people, and suppose that each player “lives” for 10 periods of the game. In periods 1–2 they are “children,” in periods 3–8 they are “adults,” and in periods 9–10 they are “seniors.” After period 10, players leave the game and do not come back. No longer can we normalize to one person per period as in Section 1, since people of different ages have different interests and motives. For now, suppose that population is constant over time, with equal numbers of people at each period, and an equal-sized new population of “children” appearing in each period as the stage 10 “seniors” die off and leave the game for good. In this assumption of constant population, the enriched model still diverges from our own situation in one respect. It will be important to investigate the behavior of the model when this assumption is relaxed.

In one version of the game, each individual must make a decision to “cooperate” in a joint savings plan, or to “defect” by not saving at all. In this version, saving for the future will be effective only if enough people choose to cooperate in each stage. In another version, saving is a group decision and the choice of a saving scheme involves the choice of a level at which each individual’s consumption will be “taxed.” In either case, saving for the future is a public good: Whatever is saved for the next period will go into a public “pot,” where it grows and becomes available for the subsequent generation. As in the earlier model, the saved resources for period t ($K_t - C_t$), will grow at the given appreciation rate, λ , to become $\lambda(K_t - C_t)$ in the next period. Savings are divided into equal shares and distributed to participants at the beginning of each period. If nothing at all is saved, then no resources will be left for consumption in subsequent periods. Call this unfortunate baseline the “state of nature.”

First consider whether the IDP permits or requires sustainable savings, or whether it dooms us to the “state of nature”—the outcome in which no one saves at all. The “state of nature” would leave later generations—those who arrive at age one after the first play—worse off than they would have been if previous generations had saved at a sustainable rate. Eventually, unsustainable saving rates will leave nothing at all for later generations. These

later generations, the “worst off group,” would have been better off if earlier generations had saved at the sustainable rate. Sustainable savings would benefit almost everyone—everyone except the age 10 seniors who leave the game after the first play. These seniors may not have a legitimate “complaint,” however, since at the sustainable rate their share, during the period in which they enjoy it, is not less than the share of others, even though their lifetime share might be slightly less than it would otherwise have been. Since the worst-off group is better off under a sustainable savings plan than it would be in the “state of nature,” it follows that the IDP requires at least sustainable savings.

But if each individual has a choice to cooperate in the savings scheme or not, and if we imagine participants to be mutually disinterested, then it is quite plausible to think that period 10 seniors will not save. They have nothing to gain and will be unable to consume whatever they save for the enjoyment of others in the future. They might save if, as Rawls first stipulated, they care about the welfare of members of the following next generations. But such concern for the welfare of future generations is not universal, and is not ordinarily thought to be a requirement of rationality.

Given the choice to cooperate or not, is it rational for others (those ages 1–9) to cooperate in sustainable saving? Compare sustainable savings to the state of nature: If all “children and adults” were to save at the sustainable rate, all would be better off over all. But in the case described, saving is a public good. It will be rational to cooperate in the production of such a good only if participants can coordinate their choices, only if those who save can be assured that a sufficient number of others will also save, and only if free-riding can be prevented. Each individual saves in “isolation,” but saving is effective only if a sufficient number of others cooperate in saving as well. So younger members might rationally choose not to save even when cooperation in a saving scheme would be mutually advantageous. Amartya Sen calls this the “isolation paradox,” but it is really just a special case of the more general problem of public goods.²¹ In such circumstances, saving might be achieved only by coercive taxation or through voluntary agreement that solves the public goods problem. For example, it would be rational for all children and adults to sign a conditionally binding agreement to save if and only if all other children and adults will save as well.²² In this way, saving might be chosen as the object of a voluntary agreement, even if contractors were not in

a Rawlsian original position. This should not lead us to conclude that saving for the future, in the sense of the problem described here, will be adequately accomplished by voluntary contract and market exchange. But if such voluntary exchanges can sometimes achieve sustainable saving for the future, this is a hopeful result.

Is the Super-Rawlsian theory trapped by the initial circumstances so that "an economy that was poor to begin with would remain in poverty" if it were governed by the IDP?²³ Or does the IDP sometimes permit or even require saving at a rate greater than the sustainable rate? It can be shown that under appropriate circumstances, the IDP does require positive saving and accumulation and is not "at the mercy of initial conditions" as Solow claims.²⁴ Suppose that the "saved resource" is a crop like wheat, and that the rate of saving is measured as the proportion of this crop that is planted (in a commons) rather than consumed. The single period payoff for this game is an individual's portion of the resources available at that stage. The difference principle will prohibit savings by the first generation only if such saving makes the worst-off individuals worse off than they would otherwise have been. The first generation to save will consume less during that period than they would have if they had instead consumed everything and saved nothing. It does not follow that they will be worse off tout court.

To see that this is so, we can make the example more specific: Suppose that the initial community includes one-hundred people [$N_1 = 100$], with ten people of each age.²⁵ The initial stock given this community is 5000 bushels of wheat [$K_1 = 5000$]. Each bushel of wheat planted generates 50 harvestable bushels for the next period, so the next-period return on savings is fifty times what was saved [so the growth rate, $\lambda = 50$]. If each person saves one bushel, then total savings will be 100 bushels [$K_1 - C_1 = 100$]. In the first period, they will have 4900 [$C_1 = 4900$] to distribute, or 49 bushels for consumption each. This is the steady state savings ratio, since planting 100 bushels will yield the same 5000 bushel stock in period two.

But perhaps this savings scheme would violate the difference principle: As noted earlier, the ten people of age 10 would be made worse off without compensation. They would consume one fewer bushel of wheat, but would gain no benefits from savings. These age-10 seniors would be the "worst off" group, and they would be disadvantaged for the benefit of younger genera-

tions who would be better off. According to the difference principle, a savings scheme must not come at cost to those who are worst off. This problem is easily solved by relieving the oldest participants, all seniors, of the obligation to save. Only individuals of ages 1–8 are required to contribute to the savings plan. For sustainable consumption, these younger participants will still need to save only a little more than one bushel per period if their aim is intergenerationally sustainable consumption and production.

The younger generations might find it advantageous to save much more than this. If they save at 4 percent per period, then those of ages 1–8 will save two bushels in the first period. Under such a savings plan, those who are age 8 at the inception of the plan (and who therefore contribute only in the first year) would increase their income for the next two periods by over 100 bushels at a cost of two bushels in period 8. Those who are age 1 at the plan's inception would increase their lifetime income by more than 350,000 bushels. Such a savings scheme would be Pareto superior as compared to the sustainable level, since these benefits would be achieved at overall cost to no one.

The older generations would be least advantaged by the savings scheme, since they will enjoy its benefits for the fewest number of periods. But it does not immediately follow from this that such savings violates the difference principle. For even those who are age 9 when the savings scheme is put in place may still benefit quite a lot: Their final period consumption would be eighty bushels instead of forty-nine (their allocation at steady state savings) even though they never contribute a thing. The only generation that would not benefit overall would be the eldest—those who leave the game at age 10 after the first play. In such a case, the IDP would require a first-period tax to benefit these period-10 seniors. After the first period of play, the problem does not arise again, since those who are age-10 seniors in the second period of play (and who were age-9 seniors in the first period) will already enjoy the benefits of cooperative saving. With side payments to the eldest seniors in the first period, saving could be made strongly Pareto superior to sustainability—literally beneficial to everyone involved.

Where positive savings is strongly Pareto superior to sustainability, saving benefits everyone. So it benefits those who are worst off and is favored over sustainability by the IDP. It follows that the IDP is consistent with, and sometimes even requires, positive saving. And it is easy to see why we would

choose such an intergenerational savings scheme from an original position like the one Rawls describes. Even if we reject Rawls's questionable argument for maximin reasoning and for the difference principle, mutually beneficial positive saving would be preferable, in some circumstances, to the "state of nature" at which no saving occurs and often preferable to merely sustainable saving. It is worthwhile to make explicit some of the assumptions that make this result possible: First, the possibility of productive saving depends, in part, on the rate of intergenerational appreciation, λ in the model above. Where λ is relatively large, saving will be more advantageous than when it is smaller. But as long as $\lambda > 1$, it will still be possible to describe mutually advantageous saving. Second, in the model described here, positive saving is mutually beneficial in part because population is constant. Where population size grows from period to period, more must be saved if later generations are to enjoy per capita benefits comparable to earlier ones. But where λ is larger than the rate of population growth, positive per capita accumulation over time will still be possible.

Positive rates of savings will not always disadvantage the first generation to adopt a savings scheme. It is also clear that the IDP can permit and will sometimes even require substantial intergenerational savings. The difference principle may be an inappropriate principle for intergenerational distributive justice, but not for the reasons given by Solow, Arrow, Dasgupta, and Rawls.

6. JUST SAVINGS IN THE REAL WORLD

I do not intend the above argument as a general defense of the IDP, the "Super-Rawlsian Theory" of intergenerational justice and saving. From the fact that the adoption of the IDP is consistent with increasing affluence and positive savings over time, it does not follow that it is superior to relevant alternative principles, since there exists a wide variety of other principles that are also consistent with positive savings. It is not clear, for example, that agents in an intergenerational original position would choose the IDP over an intergenerational "sufficiency" principle that minimized the number of persons (or generations) who would fall below a threshold of consumption, or capability, or primary goods provision. And the IDP is susceptible to many of the serious objections that have been raised against Rawls's use of the difference

principle in other contexts. Perhaps most important is the so-called "black hole problem": If some persons (or generations) are horribly disadvantaged, the IDP would recommend that all resources be devoted to them, even if the benefits to the worst-offs would be minimal, and resources could alternatively be employed in ways that would provide extravagant benefits for those who are only slightly less disadvantaged. This objection is much more troubling than the standard utilitarian concern that the difference principle would sometimes forbid trade-offs that would, on balance, maximize well-being.

The models employed here are still remote from practical choices we face, and it is difficult to draw from them any simple moral about how we should structure our own saving for the future. We do not choose a savings rate from an original position from which we are ignorant of our own generational membership. In our case, saving for the future does not involve investing a given resource in order to allow it to grow, and we do not redistribute the proceeds of savings at the beginning of each new period. The models described above do not include any accounting of the costs of such redistribution or the nature of redistributive institutions, and they will not tell us whether market institutions are likely to accomplish adequate saving for the future. Future generations are sure to be much more numerous than the present generation: Even if population growth slows and stabilizes in the next century, as some predict, we have good reason to predict that population size will more than double before it finally becomes stable. In all of these respects, our choices are quite different from the choices faced by those in the hypothetical models discussed above.

But even if we do not accept the IDP, we can glean benefits from this discussion. First, savings can benefit those generations that engage in it even while it provides benefits for the future. Since saving is appropriately modeled as a multi-stage game and not as a repeated one, saving for the future can be rational and universally beneficial. Parties to an original position choice like the one Rawls describes would surely choose to save at least at a sustainable rate, but might rationally choose a much higher rate of saving. And under appropriate circumstances, where saving is mutually advantageous, positive saving for future generations can be the object of a voluntary agreement among those who save. A second benefit is equally important: Portions of the model described in Section 1 can still be used to evaluate intergenerational

principles. It is still useful to compare principles of intergenerational justice according to the intertemporal allocation streams they imply. And while intergenerational saving is more complex than the original version of the model allows, the analytic strategy implicit in that model can still provide a crucial context for discussions of intergenerational justice.

NOTES

1. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1982), 351–79. See Clark Wolf, “Do Future Persons Presently Have Alternate Possible Identities?” in *The Non-Identity Problem*, eds. Melinda Roberts and David Wasserman (New York: Springer, 2009), 93–114.
2. David Gauthier, *Morals By Agreement* (Oxford: Clarendon Press, 1986) 268–305; Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), 182; and Clark Wolf, “Markets, Justice, and the Interests of Future Generations” *Ethics and the Environment*, 1(1996): 153–75.
3. Clark Wolf, “Property Rights, Lockean Provisos, and the Interests of Future Generations,” *Ethics*, 105 (1995): 791–818, “Social Choice and Normative Population Theory,” *Philosophical Studies*, 81 (1996): 263–82, “Markets, Justice, and the Interests of Future Generations,” *Ethics and the Environment*, 1 (1996): 153–75; “Person-Affecting Utilitarianism and Population Policy, or, Sissy Jupe’s Theory of Social Choice,” in *Contingent Future Persons*, eds. N. Fotion and J. Heller (Dordrecht, Holland: Kluwer, 1997); “Population,” in *The Blackwell Companion to Environmental Philosophy*, ed. Dale Jamieson (Cambridge: Blackwell Publishers, 2000); K. Wellman and R. Frey, “Intergenerational Justice,” in *The Blackwell Companion to Practical Ethics* (Cambridge: Blackwell Publishers, 2001); “Intergenerational Justice, Human Needs, and Climate Policy,” in *Intergenerational Justice*, eds. Axel Gosseries and Lukas Meyer (New York: Oxford University Press, 2009), 349–78.
4. John Rawls, *A Theory of Justice* (Cambridge, Mass: Harvard University Press, 1971); Kenneth Arrow, “Rawls’ Principle of Just Savings,” *Swedish Journal of Economics*, 75 (1973): 232–35; Robert Solow, “Intergenerational Equity and Exhaustible Resources,” *Review of Economic Studies*, Symposium Issue (1974): 29–45; Partha Dasgupta, “On Some Alternative Criteria for Justice Between Generations,” *Journal of Public Economics*, 3 (1974): 405–23; Partha Dasgupta, “On Some Problems Arising from Professor Rawls’ Conception of Distributive Justice,” *Theory and Decision*, 4 (1974): 325–44; and Partha Dasgupta, “Savings and Fertility: Ethical Issues,” *Philosophy and Public Affairs*, 23 (1994): 99–127.
5. Partha Dasgupta, “On Some Alternative Criteria for Justice Between Generations,” 407–409.
6. Partha Dasgupta, “On some Alternative Criteria for Justice Between Generations,” p. 415.
7. Partha Dasgupta, “On some Alternative Criteria for Justice Between Generations,” p. 415.
8. John Hicks, *Value and Capital* (Oxford: Clarendon Press, 1946), 172.
9. John Rawls, *A Theory of Justice*, Section I.
10. *Ibid.*, 287.
11. *Ibid.*, 289.
12. Susan Moller Okin, *Justice, Gender, and the Family* (New York: Basic Books, 1989).
13. John Rawls, *Political Liberalism* (New York: Columbia University Press, 1992), 274.
14. John Rawls, *The Law of Peoples* (Cambridge, England: Harvard University Press, 1999), 107; and *A Theory of Justice*, 1st ed., 287.

15. Solow, "Intergenerational Equity and Exhaustible Resources," 30.
16. *Ibid.*, 33-34.
17. Arrow, "Rawls' Principle of Just Savings," 325.
18. Dasgupta, "Savings and Fertility: Ethical Issues," 105.
19. John Rawls, *A Theory of Justice: Second Edition* (Cambridge, Mass: Harvard University Press, 1999), 254.
20. See Joseph Heath, "Intergenerational Cooperation and Distributive Justice," *Canadian Journal of Philosophy*, 27 (1997): 361-76 for a similar model.
21. Amartya Sen, "Isolation, Assurance, and the Social Rate of Discount," *Resources, Values, and Development* (Cambridge: Harvard University Press, 1997), 135-46.
22. David Schmitz, *The Limits of Government* (Boulder, Colo.: Westview Press, 1991).
23. Dasgupta, "Savings and Fertility: Ethical Issues," 105.
24. Solow, "Intergenerational Equity and Exhaustible Resources," 33.
25. My model here owes much to Heath (1997), though Heath's model is more detailed.

ESSAYS ON PHILOSOPHY, POLITICS & ECONOMICS

Integration & Common Research Projects

Edited by

Christi Favor, Gerald Gaus,

and Julian Lamont

Stanford Economics and Finance,
An Imprint of Stanford University Press
Stanford, California

2010