

Intergenerational Cooperation and Distributive Justice

JOSEPH HEATH
University of Toronto
Toronto, ON
Canada M5S 1A1

Kevin Sauvé has recently argued in this journal that David Gauthier's conception of 'morals by agreement' is inimical to the development of long-term productive investment and sustainable levels of resource exploitation.¹ According to Sauvé, this is because society is confronted with an intergenerational interaction problem whose strategic equilibrium is suboptimal (a 'Prisoner's Dilemma'). However, unlike the 'contemporaneous Prisoner's Dilemma' that Gauthier analyzes, the intergenerational version cannot be solved by an appeal to constrained maximization. As a result, Sauvé claims, Gauthier cannot effectively address the question of intergenerational justice.

The portion of Sauvé's argument that concerns me is the following:

Gauthier solves the contemporaneous Prisoner's Dilemma by ensuring that each person will cooperate only if all others cooperate, and indeed his conception of morality is aimed at ensuring that all individuals incur the costs as well as the benefits of social cooperation. But the contemporaneous solution cannot be applied to the Intergenerational Dilemma: if each generation will save for the next only if the previous generations have also saved, none will ever save. (170)

Sauvé's conclusion here is not so much incorrect as it is an artifact of the manner in which he chooses to model the intergenerational dilemma. Roughly speaking, he follows Gauthier in representing the problem as

1 'Gauthier, Property Rights, and Future Generations,' *Canadian Journal of Philosophy* 25 (1995) 163-76

169

a series of discrete one-shot games, rather than as a repeated game. This creates a variety of problems, the most conspicuous of which is that it implicitly eliminates the *generational overlap* that featured prominently in Gauthier's informal account.² In Sections I and II, I outline a more realistic model of intergenerational interaction, and show that with overlap, a wide range of cooperative savings and investment arrangements can be sustained as strategic equilibria. Furthermore, this can be done using only standard rational choice theory, and thus without appealing to the more controversial notion of constrained maximization.

Sauvé argues further that even if intergenerational cooperation could be secured within Gauthier's general framework, the only investment policies that would be excluded are those that lead to total resource depletion (168). Since any social arrangement that generates a cooperative surplus is superior to the state of nature, future generations will continue to cooperate even if they inherit the results of extremely inefficient investment policies. In Section III, I would like to show that this claim is false. In a repeated game model of intergenerational interaction, inefficient investment policies can be eliminated, as Gauthier suggests, by having each generation's policy determined by an intergenerational bargaining procedure. However, I will show that the bargaining problem this presents is considerably more complex than Gauthier anticipates.

I Cooperation

I would like to begin by presenting a formal model of intergenerational cooperation in which the productive capacity of the society is assumed to be constant. This model is intended to show how cooperation in the production of public goods can be secured across generations, without yet introducing the more complicated issue of how investment and consumption decisions will affect the productive capacity of the society over time.

In standard rational choice theory, a temporally extended interaction problem with no known terminus is modelled as an infinitely repeated

² See David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press 1986), 299. In this respect, the implied model is similar to the ones that have become standard in the discussion of Rawls's conception of intergenerational distributive justice. See Partha Dasgupta, 'On Some Alternative Criteria for Justice Between Generations,' *Journal of Public Economics* 3 (1974) 405-23, and Kenneth J. Arrow, 'Rawls's Principle of Just Saving,' *Swedish Journal of Economics* 75 (1973) 323-35.

game. The basic unit of play in a repeated game is a 'stage game,' which is just a standard simultaneous-move strategic game. The repeated game consists of an infinite sequence of stages, which are time periods in which players select moves in a stage game, then receive their payoffs for this game. Because players interact in a series of such games, they are able to adopt strategies that make their present actions contingent upon the behavior of other players in past play. Thus a strategy for any player is a set specifying an action for each stage, for each possible history of play. This means that in order to calculate the value of any strategy, each player must take into consideration not only the immediate payoffs to be gained from actions in a given stage game, but also the impact this choice of action will have on payoffs in future games, insofar as it affects the beliefs and actions of other players.

As usual, players in such a game select the strategy that maximizes their expected utility, given the strategies of the others.³ An equilibrium in a repeated game is a strategy profile in which no player could benefit by unilaterally deviating at any single stage.⁴ As is well known, a model of this type allows players to sustain a variety of outcomes that would be out of equilibrium in the component stage games. Consider a repeated game in which, at every stage, each of n players must simultaneously either ask a referee to give her \$3, or else ask a referee to give every member of the group \$2. Call the first strategy defect, and the second cooperate. This is a standard multiplayer Prisoner's Dilemma (PD). When x players cooperate and the remainder defect, each player who defects receives $\$(2x+3)$, while each one who cooperates receives $\$(2x)$. So while defection is a dominant strategy in each stage game, it can result in a dramatically suboptimal outcome. For a game with 100 players, universal defection gives each player \$3 and universal cooperation results in each receiving \$200.

³ Normally, a discount factor is introduced that reduces the present value of future payoffs. This is required in order to be able to rank strategies (since the simple sum of every infinite series of positive payoffs is the same). However, because the model that I am developing here has players 'living' for only a finite period, I have taken the liberty of dropping the discount factor. Also, it should be noted that the introduction of a discount factor enables one to use a model of the type I am presenting here not only to represent infinitely repeated games, but also finitely repeated games in which there is some uncertainty as to when the game will end. Thus the possibility that there may not be an infinite number of future generations need not affect the results of this model. See, e.g. Eric Rasmusen, *Games and Information* (Cambridge, MA: Blackwell 1989), 108-10.

⁴ See Drew Fudenberg and Jean Tirole, *Game Theory* (Cambridge, MA: The MIT Press 1991), 110.

However, full cooperation can be achieved in this game if each player adopts the strategy: 'cooperate in every game, until someone defects, then defect in every subsequent game.' With such a strategy (called a 'trigger strategy'), the benefit of defecting in any particular stage is easily outweighed by the loss of utility in the following stages, and so all players cooperate. It should be noted that the threat of reversion to the stage-game Nash equilibrium in response to unilateral defection does not require that agents be capable of making commitments. The trigger strategy responds to unilateral defection by *shifting expectations* away from cooperation to defection. The threat is credible, simply because repeated play of the stage-game Nash equilibrium is necessarily an equilibrium strategy of the repeated game. The fact that the punishment seems rather drastic is irrelevant, since it will never have to be carried out among rational agents, simply because none will ever defect.

This mechanism can sustain an infinite number of similar equilibria in this game. In fact, any outcome $o = (u_1, u_2, u_3, \dots, u_n)$ such that for every player y , $3 \leq u_y \leq 2n + 1$, can be sustained through some combination of strategies.⁵ For instance, consider a strategy profile that specified: 'everyone cooperates, except that player y defects unilaterally every third stage. If anyone else ever defects, then everyone defects in every subsequent game.' No player would defect from this strategy, since even though one player is given the opportunity to free-ride periodically, the exploited players still do better with the partially cooperative outcome than with universal defection.

An intergenerational version of a game like this is easy to construct. Suppose that each player in the repeated PD outlined above 'lives' for only t stages. For simplicity, assume that the size of each cohort is the same, so that after each stage, n/t players 'die' (exit the game), while the same number are born. We can call the number of stages a player has been in the game her 'age.' To make this more concrete, let us suppose there are 80 players, each of whom will live for 8 stages, where each stage represents one decade.⁶ The stage game will remain as above, so that in each stage, players have the option of cooperating or defecting.

5 This is a weak version of the so-called folk theorem. For non-technical discussion, see Kasmusen, *Games and Information*, 121-9.

6 The assumption that players know with certainty how long they will live is introduced for simplicity only, and has no significant impact on the results of this model. A more realistic account would assign players a certain probability of dying at each stage, and then have them discount future payoffs in accordance with life-expectancy.

In this game, the Pareto-optimal outcome is still full cooperation (providing \$160 per player per stage), but it is no longer a strategic equilibrium. This is because the Nash reversion threat is no longer effective against players of age 8. Since they know that their choice in stage 8 will have no impact on their future payoffs (since they have none), they will simply play their dominant strategy, which is defection. However, the Nash reversion threat is sufficient to ensure cooperation among players of ages 7 and less. Consider the following strategy: 'cooperate in stages 1-7 and defect in stage 8, unless someone of age 1-7 defects, then defect in all stages.' This offers a player of age 7 who cooperates at stage 7 an expected payoff of \$140 for that stage, followed by \$143 at stage 8. Defection, on the other hand, would give \$141 at stage 7, followed by \$3 at stage 8. Since the former payoff sequence easily outweighs the latter, it is possible to sustain cooperation among players aged 1-7, even while those of age 8 free-ride.

Note, however, that a wide range of other cooperative arrangements are possible. In fact, any outcome in which players receive lifetime earnings somewhere between \$24 and \$1123 is sustainable. For example, the group might adopt an enlightened labor policy in which players aged 1-2 and 7-8 could defect with impunity. For convenience, let us call players aged 1-2 'children,' aged 3-6 'adults,' and aged 7-8 'seniors.' We can call the strategy profile in which only adults cooperate *arrangement* Γ . Since cooperation among adults would ensure each player lifetime earnings of \$652, and any adult who defected would (at very least) forfeit his pension of \$160, there would be no incentive for anyone to deviate from this arrangement.

Consider the following interpretation of this game. Take a simple economy in which players have the option of either working in the collective corn field, or going off and picking berries. For some reason, corn consumption cannot be policed, so the entire crop is simply divided up equally among all members of the society. In every stage, the corn field is able to produce a return of 200 bushels for each player who works it (e.g. when the field is worked by 40 players, it produces 8000 bushels, resulting in a distribution of 100 bushels to each player). Berry-picking on the other hand produces 3 baskets per stage for the berry-picker, and none for the other members of society. Assume all members of the society are indifferent between a bushel of corn and a basket of berries.

Although berry-picking is a dominant strategy in every stage game (since each player receives a share of corn regardless of whether she works the field), it is possible to sustain a variety of mixed economies. For instance, arrangement Γ , in which all players berry-picked as children and seniors but worked the corn field as adults, would result in a lifetime income of 12 baskets of berries and 800 bushels of corn. This arrangement would be sustained by a Nash reversion threat: if any adult

refuses to work the corn field, everyone reverts to the berry-picking state of nature.

II Investment

The corn-berry model outlined above does not raise any questions about investment, because the returns on cooperative activity are assumed to be constant. In order to introduce this dimension, consider a situation in which the size of the cooperative surplus is determined jointly by the number of players who cooperate and the level of inherited investment. (For convenience, I will lump together resource conservation and productive investment under the single heading of 'investment.')

In order to do this, we must add two little complications to the game.

First, let us suppose that in order to sustain the corn field, a certain portion of the corn produced in each stage must be set aside as seed for the next crop. If one bushel of corn can be produced using A bushels of corn as input, along with L units of labor (these units represent the work of one player for one stage), the basic agricultural technology of this society can be represented by a *production function*:

$$F(A,L) = 1.^9 \quad (1)$$

Let us suppose that when $L = 1$, $A = 1/50$. This means that in one stage, one worker is able to produce an output equivalent to 50 times the amount of seed corn he is given. The *base productivity of labor* can then be represented:

$$p = \frac{1}{AL} = 50$$

Now let us introduce one further complication. Assume that agricultural labor is subject to diminishing marginal returns, so that each worker's productivity declines as the number of bushels she must plant increases. We can select some factor α (where $0 < \alpha < 1$) to specify the rate at which productivity declines as the number of units worked increases, and represent each worker's *marginal product* on the y^{th} bushel worked as:

$$p_y = (\alpha)^{y-1} p. \quad (2)$$

To see how (2) works, assume that $\alpha = 9/10$. A worker will then be able to produce 50 bushels of output for the first bushel of seed she

works, 45 bushels for the second, 40.5 bushels for the third, and so on. (The reason for introducing this complication is so that advantages continue to be realized from increasing the number of players who cooperate in working the corn field, while some upward bound is placed on the advantages to be realized through savings.)

Once this is done, it is possible to produce an overall production function for the society. Taking s to be the total seed corn saved, l the total labor units contributed (i.e. the number of workers cooperating in each stage), and p the base productivity of labor, total output is equal to:

$$l \left[\sum_{y=1}^{s/l} (\alpha)^{y-1} (p) \right]. \quad (3)$$

Equation (3) basically takes the total amount of seed corn, divides it up among the workers, and adds up the amount each worker will be able to produce (given diminishing returns). In this way, 10 bushels of seed worked by 5 players would produce 475 bushels of output, whereas the same amount worked by 10 players would produce 500.

Under arrangement Γ , in which 40 players work the corn field per stage, 200 bushels of seed corn will produce a gross output of 8200 bushels:

$$40 \left[\sum_{y=1}^5 (\alpha)^{y-1} (50) \right] = 40(50+45+40.5+36.5+33) = 8200.$$

If 200 of these bushels are saved, and 8000 of them distributed among the members of the society for consumption, then this productive arrangement can, in principle, be sustained indefinitely. With these figures the level of production and consumption in this model would be equivalent to the level in the simple corn-berry economy presented at the end of the previous section. The difference here is that the productive capacity of the society is not a 'built-in' feature of the model, but can be controlled by the players, insofar as they are able to decide for themselves how much grain should be consumed, and how much should be set aside for seed (i.e. invested). This allows us to address the question that is important from the standpoint of intergenerational distributive justice, viz. what sort of investment arrangements players will be able to sustain as strategic equilibria.

In order to model this aspect of the problem, we must first specify how savings decisions are to be made, and how they fit into the game. For simplicity, let us suppose that each player makes an individual decision

64

as to how much she should save. In each stage, players first choose whether to work the corn field or pick berries. They then receive their payoffs, in grain and/or berries, and decide how many bushels of corn to save. A strategy for a stage game therefore consists in a choice of action, followed by a decision on how much to save. This allows players to develop strategies for the repeated game that are conditional upon both the working and savings decisions of other players.

In such a situation, the strongly dominant strategy for any stage game will be to save nothing. Since all corn produced is distributed out equally, saving 1 bushel of corn in one stage will result in a maximum payoff (given the base productivity of labor) of 5/8ths of a bushel in the stage immediately following. However, a variety of savings strategies that are dominated in the stage game can be sustained in the repeated game. Suppose all players adopt the following strategy: 'all adults must work in the corn field and save 5 bushels of corn, but seniors and children are exempt from work and savings. If any adult defects, all players subsequently refuse to work or save.' This arrangement is obviously in equilibrium, since it provides each adult member of society with a net income of 97.5 bushels of corn per adult state, along with the anticipation of a pension of 102.5 bushels of corn plus 3 baskets of berries per senior stage. Any player contemplating defection (either refusing to work or failing to save), can look forward to consumption of only 3 baskets of berries in every subsequent stage — a significantly inferior payoff.

Let us call this equilibrium strategy profile *arrangement Γ* . Again, it is only one out of an infinite set of sustainable social arrangements. I have chosen it for discussion because it has the attractive feature of maintaining a constant level of production and investment across generations, and because the manner in which it distributes the burden of labor and savings bears some resemblance to the arrangements that are current in our society.

III The Rate of Investment

Having outlined this model, it is now possible to address some of the interesting questions that arise concerning the specification of an adequate intergenerational conception of distributive justice. The model shows that compliance can be secured for a very wide range of investment schemes. The question now is how a particular scheme should be selected.

Let us refer to the component of the strategy that specifies who must work the corn field as the 'labor policy' and the component that specifies who must save, and how much they must save, as the 'investment policy.' An equilibrium will therefore consist of a strategy profile that

contains some combination of labor and investment policies. We can refer to a 'depletionary investment policy' as one that will, in the foreseeable future, reduce the size of the cooperative surplus (e.g. the productivity of the corn field) to less than the state of nature.

The first thing to notice is that any strategy profile that contains a depletionary investment policy will be out of equilibrium, *no matter how far off in the future the full effects of the depletion are felt*. Consider the following corn investment policy. Under arrangement Γ , each adult was required to save roughly 5 per cent of her gross income. Suppose we now lower this to 3 per cent. Such a policy would lower the gross output of the corn field from 8200 bushels to below 240 in exactly 13 stages. At a gross output level of 240, the corn field yields 3 bushels of corn for each player. Since berry-picking ensures each player an equivalent 3 baskets of berries, 240 marks the indifference point between cooperative farming and the state of nature (this output is the result of a total savings of 4.8 bushels). This means that 13 stages after a 3 per cent savings rate was levied on working adults, there would be so little seed corn remaining that everyone would be better off picking berries.

The interesting feature of this policy is that because seed supplies are depleted only on the unlucky 13th stage, every player living at the time such a policy was adopted (along with their immediate children) could continue to enjoy a cooperative surplus. It might therefore be concluded that all currently living players could be motivated to accept this policy. This would be a mistake. Let us refer to the players aged 8 at the time such a policy was adopted as generation x , those aged 7 generation $x+1$, etc. Once a 3 per cent savings policy was adopted, the 6th next generation born (generation $x+14$) would no longer be able to count on receiving any cooperative surplus at age 7. Because of this, the Nash reversion threat would no longer be effective against this generation at age 6 and so it would defect. But then generation $x+15$, which could not count on any surplus at ages 6 or 7, could not be motivated to cooperate at age 5. But then anticipating this defection by generation $x+15$, generation $x+14$ could no longer be motivated to cooperate at age 5, and generation $x+13$ could no longer be motivated to cooperate at age 6. Iterate this a few times and the entire cooperative equilibrium unravels all the way back to generation x . Immediately upon adoption of a depletionary investment policy, universal defection by all living generations becomes the sole equilibrium.

The reason for this is quite obvious. A depletionary investment policy transforms the infinitely repeated multiplayer PD into a *finitely repeated* PD. This makes the trigger strategy ineffective in the last round, and therefore ineffective in the second-last round, and by backward induction, ineffective in all rounds. Once it is known when cooperation will end, it is possible to start the backward induction argument that unravels

the cooperative equilibrium. In cases where the investment policy maintains production above the level supplied by the state of nature, there is no foreseeable end to cooperation, and so no way to begin an inference of this type. This is a counterintuitive but very well-known result in game theory. The important point here is simply that transforming the repeated game into one with overlapping generations of players does nothing to change it.⁷

Note that this 'unraveling' does not eliminate inefficient equilibria. If the savings rate were dropped to 4 per cent, production would drop each year until it eventually leveled out at 2000 bushels. But similarly, the group could dramatically increase production by opting to save 10 bushels per stage (this would bump output up from 8200 to over 13,000). There is, however, an upward bound on output imposed by the decreasing marginal productivity of labor. A gross savings of 2925 bushels under arrangement Γ (73.125 per adult) maximizes both corn production and consumption. At this rate, one more bushel saved at stage t would produce exactly one more unit of output in stage $t+1$.⁸

This means that the set of investment policies that can be sustained in equilibrium provides the rational choice theorist with something of an embarrassment of riches. The shaded region of Figure 1 shows the equilibrium investment levels available under arrangement Γ (with no saving beyond the margin⁹). Lines A-C represent some of the specific investment policies that the players might adopt.

So how is an investment policy to be chosen? Gauthier follows Rawls in arguing that the investment policy selected should be the outcome of a hypothetical intergenerational bargain. The idea here is to introduce a device that will allow agents to pick one particular equilibrium from the available set.¹⁰

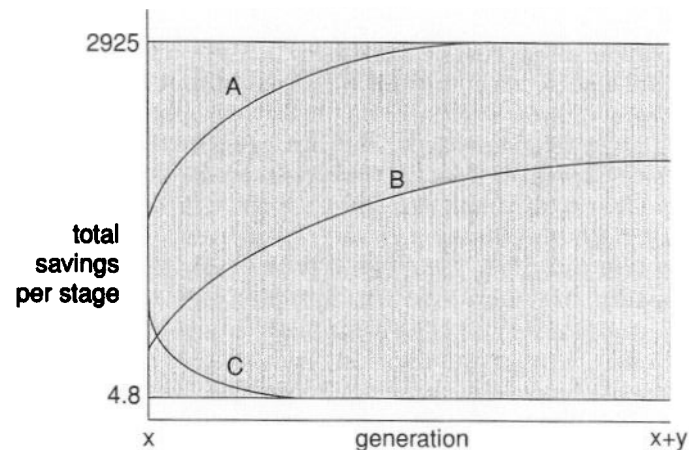


Figure 1

First, it should be noticed that *every* generation has an effective threat that it can wield against all previous generations. By simply refusing to cooperate, any generation could, through the mechanism outlined above, terminate cooperation in all previous generations. This means that any investment policy adopted must be one that each generation anticipates will be acceptable to all future generations. Naturally, the *minimal criterion* for the acceptability of an investment policy is that it not be depletionary. But there is nothing that stops us from imposing additional constraints on what policies are to count as acceptable.¹¹ Because each generation is threatening the previous with a suboptimal *equilibrium* rather than a suboptimal *action*, and since game theory lacks an intentional equilibrium-selection mechanism, the instrumental conception of rationality does not impose any constraints on the reasonableness or credibility of the threat, nor does it dictate the criteria that determine what is to count as acceptable. This means that the theorist is free, quite literally, to make something up. This is just what Gauthier does, by

7 For further discussion, see Fudenberg and Tirole, *Game Theory*, 171. It is worth noting that depletionary investment policies are excluded, not because existing generations have some obligation to posterity, but through a simple combination of their self-interest and the assumption that future generations will act rationally. Otherwise put, the elimination of depletionary investment policies is a deductive consequence of the assumption that all individuals maximize expected utility. Because of this, Parfit's non-identity problem does not arise. See Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press 1984).

8 Naturally, introducing some form of time-discounting would lower this.

9 I am not entirely confident about this restriction. Although it seems quite reasonable, our intuitions about such things are notoriously suspect in strategic contexts.

10 For an interpretation of the social contract as an equilibrium-selection device, see Ken Binmore, *Playing Fair* (Cambridge, MA: The MIT Press 1994), 334-40.

11 Furthermore, there is no in principle limit on the number of other issues that could be introduced into these negotiations. For instance, each generation could also make its cooperation dependent upon an acceptable population policy. This would resolve the intergenerational PD that Derek Parfit thought would lead to overpopulation. See *Reasons and Persons*, 383.

96

suggesting that the rate of investment should be determined by the principle of minimax relative concession.¹²

Because each generation wields an effective threat against previous generations, it is also in a position to make demands upon them. Bargaining theory therefore suggests itself as a way of specifying an arrangement that will mediate these demands in a manner that is acceptable to all. This idea exactly parallels Gauthier's approach to cooperation in the one-shot PD. In this situation, players should be willing to accept any arrangement that gives them even a tiny portion of the cooperative surplus. This creates an allocation-of-surplus problem that Gauthier argues can be resolved through bargaining. Thus his definition of constrained maximization specifies that agents of this type will only be willing to participate in cooperative arrangements when their expected utility approaches what they would expect from the cooperative outcome determined by minimax relative concession.¹³

Since the model of intergenerational cooperation I have been developing uses only standard game theory, the use of a bargaining model to select an investment policy can be introduced as a *refinement* on the solution concept that I have been using, rather than as a new type of rational agency.¹⁴ Thus the results presented here do not rely upon any of the more controversial elements of Gauthier's program. (Nor are they incompatible. Gauthier's constrained maximizers act exactly like stand-

ard rational choice agents *except* when they are faced with suboptimal strategic equilibria. Since all of the Pareto-optimal outcomes in this model are also strategic equilibria, whether or not agents possess a special cooperative choice disposition is irrelevant, since it would not be manifested in this context.)

By threatening to refuse all cooperation, each generation is in a position to place demands on the previous generations. Gauthier sets up the problem as follows. For simplicity, let us say that each generation makes such a demand only on the immediately older generation.¹⁵ Each younger generation will then demand that the older one adopt an investment policy that will maximize productive output (under arrangement Γ , this is 2925 gross). At the same time, each older generation will demand an investment rate that will produce the minimum feasible output (under arrangement Γ , this is 4.8 bushels), and thereby maximize their consumption. As Gauthier observes (correctly) a utilitarian solution to this problem would prescribe the maximum investment level, and (incorrectly) a Rawlsian solution would prescribe a minimum investment level.¹⁶ The former corresponds to line A in Figure 1, while the latter corresponds to line C. He suggests instead that the minimax relative concession principle would have agents split the difference (as shown by line B). In our model, this would result in a gross savings of 1460 bushels, or a rate of 36.5 bushels per adult per stage.

The basic idea here is certainly correct, although the details of the proposal are not. First, Gauthier tacitly employs a model in which there is no intergenerational overlap. In a model with overlap, each generation will actually have a different maximum and minimum claim at each stage. Similarly, the manner in which the investment policy distributes the burdens of savings across age groups will significantly affect these claims. For instance, under arrangement Γ , no living generation could benefit from a minimum savings rate. Since seniors are not required to save, they have no incentive to lower the rate. And since adults who could benefit from a reduction in the investment rate would

12 Although in *Morals by Agreement*, Gauthier tries to suggest that his bargaining solution follows directly from the instrumental model. He has since retracted this claim, acknowledging that his bargaining solution lacks microfoundations. See his 'Uniting Separate Persons,' in David Gauthier and Robert Sugden, eds., *Rationality, Justice and the Social Contract* (Hemel Hempstead: Harvester Wheatsheaf 1993).

13 Gauthier, *Morals by Agreement*, 167. Thus Sauvé directly begs the question against Gauthier when he claims that: 'The only cases in which new generations will find it irrational to continue the social contract are cases where the benefits of continued cooperation no longer outweigh the benefits available in the state of nature' (168). In Gauthier's terms, while *broadly compliant* agents would reject only depletionary policies, *narrowly compliant* agents would reject all policies that are unfair (Gauthier, *Morals by Agreement*, 178). Similarly, Sauvé's claim that consumption will 'ratchet down' as the social contract is renegotiated across generations, slowing but not stopping resource depletion, is incorrect. Once the cooperative equilibrium is selected, it determines the rational strategy for all players in all generations, so there will be no renegotiation.

14 Elsewhere, I have argued that Gauthier's introduction of constrained maximization as a special choice disposition is superfluous across the board, because the 'contemporaneous' choice problem that he considers should also be modelled as a repeated game. See Joseph Heath, 'A Multi-Stage Game Model of *Morals by Agreement*,' *Dialogue* 35 (1996) 529-52.

15 In Gauthier's actual model, each generation makes a claim upon all previous generations. In my corn-berry model this complication is not necessary because the production level can be maximized in a single stage.

16 Gauthier, *Morals by Agreement*, 304-5. The claim about Rawls is incorrect because the 'zero-investment' policy results only in models with no generational overlap, and occurs only if one ignores Rawls's stipulation that the difference principle not be applied to intergenerational contracts.

also thereby lose a portion of their pension, they have no incentive to lower it either.¹⁷

This means that the maximum and minimum demands will be very difficult to calculate, and will depend a lot upon specific aspects of the associated investment and labor policies (not to mention the technologies available in the society). This means that the overall bargaining problem will be vastly more complicated than the one Gauthier envisioned.¹⁸ Nevertheless, there is no problem with the *principle* of his approach. And he is perfectly justified in claiming that, within his framework, the only sustainable investment policy is one that is *fair to all future generations*.

IV Conclusion

This suffices to show that Sauvé's two arguments against Gauthier fail: first, there is no problem in securing stable investment across generations under the assumption of mutual unconcern, and second, it is not only depletionary policies that are eliminated under Gauthier's scheme. However, even if Gauthier's model does 'work' as he claims, there may be some doubt as to the value of a model based upon the assumption of mutual unconcern, since it is so obviously counterfactual. In response to such doubts, I would argue that because the increasing complexity of modern societies has led to the gradual detachment of economic and political interactions from the sorts of close personal relationships that generate strong affective ties, mutual unconcern is *becoming* an increasingly realistic assumption.

If all production and investment were strictly intrafamilial, then provision for one's immediate heirs would naturally provide the primary impetus for agents to forego consumption. However, in an advanced industrial economy, in which most investment and production occurs

on a very large scale with savings that have been 'pooled' together by banks, and in which most retirement benefits are provided by governments or large investment firms, the transfer of wealth among generations has largely been *collectivized*. This means that most agents are no longer in a position to provide for their children or parents directly, but do so only under much larger cooperative schemes (e.g. children support their parents by paying taxes that the government uses to pay out interest on retirement savings bonds). This fact, combined with the *anonymity* of large-scale economic relations, has resulted, I believe, in an increasing dilution of mutual concern.

The well-being of most individuals depends not so much on their wealth as it does upon their income. But their income is determined primarily by the overall performance of the economy, which in turn depends upon the level of inherited investment. Since the savings decisions of individual agents have little appreciable impact on this performance, an obvious collective action problem arises. Naturally, where parents are still in a position to provide for their children directly, as in the case of educational benefits, affective ties are sometimes able to sustain relatively high levels of investment. But the alarmingly low savings rates in many Western nations (particularly the United States), suggests that a significant free rider problem has arisen. Not only are most individuals counting upon others to make the productive investments that will produce their retirement income, they are also (insofar as they care) counting upon others to make the investments that will sustain their children's income levels.

The fact that such collective action problems exist suggests that the assumption of mutual unconcern is realistic in the current social context. In countries like Canada and the United States, where these problems have largely gone unchecked, it also suggests that government policies with regard to consumption and investment are being made without regard for basic principles of intergenerational distributive justice ('pay-as-you-go' financing of major social programs, such as health care and government pensions, is the most egregious example¹⁹). Gauthier's

17 Those who are troubled by the fact that the non-utilitarian savings policies create Pareto-inefficient outcomes might like to rig up a model so that all minimum and maximum claims coincide.

18 Furthermore, my production function assumes that the returns on investment are enjoyed in the very next stage. A more realistic model, in which payoffs may only arise several turns later, gives rise to the possibility of investment policies in which short-term decreases in production can create a greater long-run average. There is, however, no justification whatsoever for Sauvé's claim that a rational choice model of this type entails a preference for short-term rather than long-term investments.

19 For discussion see Laurence J. Kotlikoff, *Generational Accounting* (New York: The Free Press 1992). It is worth noting that the current pressure from some members of the 'baby boom' generation to defect from the Canada Pension Plan is not a result of their own contribution level being too high. On the contrary, it is because their own contribution level is so low. Their underfunding of the plan amounts to an attempt to effect a massive transfer of wealth from members of younger generations to themselves. However, the scale of this proposed redistribution is so large that many boomers have come to anticipate the defection of these generations. It is this *anticipated* defection that supplies them with the *current* motivation to defect from

analysis is appealing because it suggests that this problem can be addressed, not only through emotional appeals and ethical suasion, but through the institutionalization of effective intergenerational bargaining procedures.²⁰

Received: September, 1995

Revised: May, 1996

the plan. This would appear to be an example of the 'unraveling' that can occur when savings arrangements are adopted that are predictably unacceptable to future generations.

²⁰ I would like to thank Gustaf Arrhenius, Kevin Sauvé, Margaret Schabas, and three anonymous *CJP* referees for commenting on this paper, and the Social Sciences and Humanities Research Council of Canada for financial assistance during the period in which it was written.