

Dynamic Free Riding with Irreversible Investments[†]

By MARCO BATTAGLINI, SALVATORE NUNNARI, AND THOMAS R. PALFREY*

We study the Markov equilibria of a model of free riding in which n infinitely lived agents choose between private consumption and irreversible contributions to a durable public good. We show that the set of equilibrium steady states converges to a unique point as depreciation converges to zero. For any level of depreciation, moreover, the highest steady state converges to the efficient level as agents become increasingly patient. These results are in contrast to the case with reversible investments, where a continuum of inefficient equilibrium steady states exists for any level of depreciation, discount factor, and size of population. (JEL D11, H41)

The most significant kinds of free rider problems are characterized by two key features. First, they are dynamic. Public goods, for example, are often durable: it takes time to accumulate them and they depreciate slowly, projecting their benefits for many years. Similarly, environmental problems depend on variables that slowly evolve over time like capital goods. In these examples, what matters for the agents in the economy is the stock of the individual contributions accumulated over time. Second, they are characterized by irreversibility. Major investments in public goods—such as bridges, roads, and military equipment—are not easily transformed into other forms of consumption; similarly, the effects of pollution are difficult or impossible to reverse. Although there is a large literature dedicated to free rider problems, surprisingly little is known about dynamic problems with irreversibility. How large are the inefficiencies in dynamic environments with irreversibility? How do the distortions with irreversibility compare to the distortions in static models and in dynamic models with reversibility? To date, these questions have been posed only for very specific environments, and never fully answered.

In this paper, we present a simple model of free riding to address these questions. In the model, n infinitely lived agents allocate their income between private consumption and contributions to a public good in every period. The public good is

*Battaglini: Department of Economics, Princeton University, Fisher Hall, Princeton NJ 08544 (e-mail: mbattagl@princeton.edu); Nunnari: Department of Political Science, Columbia University, Mail Code 3320, New York, NY 10027 (e-mail: snunnari@columbia.edu); Palfrey: Division of the Humanities and Social Sciences, California Institute of Technology, Mail Code 228-77, Pasadena, CA 91125 (e-mail: trp@hss.caltech.edu). Battaglini gratefully acknowledges financial support from the Alfred P. Sloan Foundation. Palfrey gratefully acknowledges financial support from NSF (SES-0962802). We are grateful to seminar participants at the Einaudi Institute for Economics and Finance, Essex, London School of Economics, Ohio State, Princeton, Toulouse, Warwick, Yale, and at the 2012 Southwest Economic Theory Conference. Steve Matthews provided very helpful detailed comments on an earlier draft. Juan Ortner provided excellent research assistance. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†]Go to <http://dx.doi.org/10.1257/aer.104.9.2858> to visit the article page for additional materials and author disclosure statement(s).

durable and depreciates at a rate d . We study the properties of the Markov equilibria in which the public good is irreversible. We present three main results. First, we show that, for any group size and rate of depreciation, the steady state of the best Markov equilibrium converges to the efficient steady state as the agents' discount factor converges to one. Second, as depreciation converges to zero, the set of steady states converges to a unique point. It follows that, when the discount factor is high and depreciation is low, all equilibrium steady states are close to efficient. Third, however, convergence to the steady state is inefficiently slow. From these results we draw two general conclusions on dynamic free rider games with irreversibility: (i) contrary to what happens with the more widely studied case with reversible investments, multiplicity of steady-state equilibria is not an issue when depreciation is small; (ii) the problem with dynamic free riding with irreversibility is not so much an inefficient steady state, but an inefficiently slow accumulation path.

These results are related to two strands of literature. First, they are related to the large literature on dynamic public good games with reversibility. As in our work, this literature has studied the inefficiencies arising in the Markov equilibria of dynamic public goods games. Contrary to our work, however, this literature has focused exclusively on environments with reversible investments.¹ In contrast, our approach is general enough to allow sharp comparisons of equilibrium public good investments with and without reversibility. The comparisons we obtain show clearly that the irreversibility of investments is very important in modeling dynamic free rider problems. With reversibility, for any group size, discount factor, or rate of depreciation, there is always a continuum of equilibrium steady states, most of which are very inefficient. Irreversibility, therefore, has a *beneficial* effect on welfare because it eliminates the many inefficient equilibria that arise in economies with reversibility.

The second strand of literature to which our paper is related is the research on monotone contribution games, notably Lockwood and Thomas (2002) and Matthews (2013).² In these papers, the techniques used to fully characterize subgame perfect equilibria require that players have a dominant strategy of zero contribution in the stage game and the depreciation rate is exactly zero. This literature has been focused on the comparison of the most efficient subgame perfect equilibria in economies with reversibility (when the actions can increase and decrease) and with irreversibility (when the actions cannot decrease), and it arrives at a conclusion that is essentially the opposite of ours. While we find that irreversibility is beneficial, its main finding is that, when the discount factor is sufficiently high, irreversibility always induces less efficient allocations because it limits the effectiveness of trigger strategies in punishing deviations. That conclusion, however, critically depends on the assumption that players have a dominant strategy in the stage game of zero contribution,

¹Classic contributions include Levhari and Mirman (1980) and Fershtman and Nitzan (1991). More recent works include Dockner and Long (1993); Dutta and Radner (2004); Rubio and Casino (2002); Battaglini and Coate (2007); and Besley and Persson (2011). The last two papers have studied dynamic public good games in which investments are chosen in a collective decision-making process.

²A number of important papers in the monotone games literature are less directly related. These papers require additional assumptions which make their environments hard to compare to ours. Gale (2001) studies games in which agents care only about the limit contributions as $t \rightarrow \infty$. Admati and Perry (1991); Compte and Jehiel (2004); and Marx and Matthews (2000) consider environments in which the benefit of the contribution occurs at the end of the game if a threshold is reached and in which players receive either partial or no benefit from interim contributions. The first two of these papers, moreover, assume that players contribute sequentially, one at a time, so the game form is different.

and that the rate of depreciation of the state variable is exactly zero. As we prove in Section IV, when depreciation is positive, even if arbitrarily small, subgame perfect equilibria are not sufficiently restrictive to allow a clear comparison of economies with and without irreversibility.

As a methodological contribution, the paper develops a novel approach to characterize the Markov equilibria that may have more general applicability in the analysis of stochastic games with discrete time. The idea is to construct pure strategies that induce an objective function with a flat top: the flat region makes the players indifferent between different rates of accumulation. This provides additional freedom in choosing the players' reaction functions that is essential in proving existence of a pure strategy Markov equilibrium.

I. The Model

Consider an economy with n agents. There are two goods: a private good x and a public good g . The level of consumption of the private good by agent i in period t is x_t^i , the level of the public good in period t is g_t . An allocation is an infinite nonnegative sequence $z = (x_\infty, g_\infty)$, where $x_\infty = (x_1^1, \dots, x_1^n, \dots, x_t^1, \dots, x_t^n, \dots)$ and $g_\infty = (g_1, \dots, g_t, \dots)$. We refer to $z_t = (x_t, g_t)$ as the allocation in period t . The utility U^j of agent j is a function of $z^j = (x_\infty^j, g_\infty)$, where $x_\infty^j = (x_1^j, \dots, x_t^j, \dots)$. We assume that U^j can be written as $U^j(z^j) = \sum_{t=1}^{\infty} \delta^{t-1} [x_t^j + u(g_t)]$, where $u(\cdot)$ is continuously twice differentiable, strictly increasing, and strictly concave on $[0, \infty)$, with $\lim_{g \rightarrow 0^+} u'(g) = \infty$ and $\lim_{g \rightarrow +\infty} u'(g) = 0$. The future is discounted at a rate δ . There is a linear technology by which the private good can be used to produce the public good, with a marginal rate of transformation $p = 1$. The private consumption good is nondurable, the public good is durable, and the stock of the public good depreciates at a rate $d \in [0, 1]$ between periods. Thus, if the level of public good at time $t - 1$ is g_{t-1} and the total investment in the public good is I_t , then the level of public good at time t will be $g_t = (1 - d)g_{t-1} + I_t$.

In an *Irreversible Investment Economy* (IIE) the public policy in period t is required to satisfy three feasibility conditions: (i) $x_t^j \geq 0 \forall j, \forall t$; (ii) $g_t \geq (1 - d)g_{t-1} \forall t$; (iii) $I_t + \sum_{j=1}^n x_t^j \leq W \forall t$, where W is the aggregate per period level of resources in the economy. The first condition guarantees that private allocations are nonnegative. The second condition is the irreversibility condition, and is equivalent to $I_t \geq 0 \forall t$. In contrast, a *Reversible Investment Economy* (RIE) is an economy where the second constraint is replaced by $g_t \geq 0$, so I_t can be negative. We compare IIE to RIE in Section IV.

It is convenient to distinguish the state variable at t , g_{t-1} , from the policy choice g_t and to reformulate the budget condition. If we denote $y_t = (1 - d)g_{t-1} + I_t$ as the new level of public good after investing I_t in the current period when the last period's level of the public good is g_{t-1} , then the public policy in period t can be represented by a vector $(y_t, x_t^1, \dots, x_t^n)$. Substituting y_t , the budget balance constraint $I_t + \sum_{j=1}^n x_t^j \leq W$ can be rewritten as $\sum_{j=1}^n x_t^j + [y_t - (1 - d)g_{t-1}] \leq W$. With this notation, we must have $x_t \geq 0$, $y_t \geq (1 - d)g_{t-1}$ in an IIE.

The initial stock of public good is $g_0 \geq 0$, exogenously given. Public policies are chosen as in the classic free rider problem, modeled by a voluntary contribution game. In period t , each agent j is endowed with $w_t^j = W/n$ units of private good. We

assume that each agent has full property rights over a share of the endowment (W/n) and in each period chooses on its own how to allocate its endowment between an individual contribution to the stock of public good (which is shared by all agents) and private consumption, taking as given the strategies of the other agents. The individual contribution by agent j at time t is denoted i_t^j , and in an irreversible investment economy we require $i_t^j \in [0, W/n] \forall j$. The total economy-wide increase in the stock of the public good in any period is then given by the sum of the agents' individual contributions.³

To study the properties of the dynamic free rider problem described above, we study symmetric Markov perfect equilibria, where all agents use the same strategy, and these strategies are time-independent functions of the state, g .⁴ A strategy is a pair $(x(\cdot), i(\cdot))$, where $x(g)$ is an agent's level of consumption and $i(g)$ is an agent's contribution to the stock of public good in state g . Given these strategies, by symmetry, the stock of public good in state g is $y(g) = (1 - d)g + ni(g)$. For the remainder of the paper we refer to $y(g)$ as the investment function. Associated with any Markov perfect equilibrium of the game is a value function, $v(g)$, which specifies the expected discounted future payoff to an agent when the state is g . An equilibrium is continuous if the investment function, $y(g)$, and the value function, $v(g)$, are both continuous in g . In the remainder of the paper we will focus on continuous equilibria. In the following we refer to equilibria with the properties described above simply as equilibria.

We are interested in studying the long-term properties of the allocation. Given an equilibrium $(y(g), v(g))$, an allocation y^o is a steady state if it is a fixed point of the investment function: $y^o = y(y^o)$. A steady state y^o is said to be stable if there is a neighborhood $N_\varepsilon(y^o)$ of y^o such that for any $N_{\varepsilon'}(y^o) \subseteq N_\varepsilon(y^o)$, $g \in N_{\varepsilon'}(y^o)$ implies $y(g) \in N_{\varepsilon'}(y^o)$. Intuitively, starting in a neighborhood of a stable steady state, g remains in a neighborhood of a stable steady state for all future periods. In what follows we will focus only on steady states that are stable and we will refer to stable steady states simply as steady states. We say that convergence to a steady state is gradual if it is not reached in finite time starting from a left neighborhood of it. We say that convergence is monotonic if the state converges monotonically to the steady state.

II. The Planner's Problem

As a benchmark with which to compare the equilibrium allocations, we first analyze the sequence of public policies that would be chosen by a utilitarian planner. The planner's solution is extremely simple in the environment described in the previous section: this feature will help to highlight the subtlety of the strategic interaction studied in the next two sections.

³The state variable in our model can have alternative interpretations. For example, one may think of g as physical public capital, which seems natural to assume is irreversible. Once a bridge is constructed, it cannot be demolished and costlessly transformed back to consumption. But we have in mind a more general interpretation. In some applications g could represent social capital or certain aspects of aggregate human capital (literacy rates, for example). The key aspect of g is its durability.

⁴We study the subgame perfect equilibria of the game in Section IV where we compare the predictions in the two equilibrium concepts.

The planner's problem has a recursive representation in which g is the state variable and $v_p(g)$, the planner's value function, can be represented as:

$$(1) \quad v_p(g) = \max_{y,x} \left\{ \begin{array}{l} \sum_{j=1}^n x^j + nu(y) + \delta v_p(y) \\ \text{s.t.} \quad \sum_{j=1}^n x^j + y - (1-d)g \leq W, \quad x^i \geq 0 \quad \forall i, \quad y \geq (1-d)g \end{array} \right\}.$$

By standard methods (see Stokey, Lucas, and Prescott 1989), we can show that a continuous, strictly concave, and differentiable $v_p(g)$ that satisfies (1) exists and is unique. The optimal policies have an intuitive characterization. When the accumulated level of public good is low, the marginal benefit of increasing y is high, and the planner finds it optimal to spend as much as possible on building the stock of public good: in this region of the state space $y_p(g) = W + (1-d)g$ and $\sum_{j=1}^n x^j = 0$. When g is high, the planner will be able to reach the level of public good $y_p^*(\delta, d, n)$ that solves the planner's unconstrained problem: i.e., $nu'(y_p^*(\delta, d, n)) + \delta v_p'(y_p^*(\delta, d, n)) = 1$. Applying the envelope theorem, we can show that at the interior solution $y_p^*(\delta, d, n)$ we have $v_p'(y_p^*(\delta, d, n)) = 1-d$. It follows that $y_p^*(\delta, d, n) = [u']^{-1}\left(\frac{1-\delta(1-d)}{n}\right)$. The investment function has the following simple structure. If $(1-d)g < (y_p^*(\delta, d, n) - W)$, then $y_p^*(\delta, d, n)$ is not feasible: the planner spends W on the public good so $y_p(g) = (1-d)g + W$. If $(1-d)g \in [y_p^*(\delta, d, n) - W, y_p^*(\delta, d, n)]$, instead, the planner can choose $y_p(g) = y_p^*(\delta, d, n)$ without violating the constraints. If $(1-d)g > y_p^*(\delta, d, n)$, then the irreversibility constraint is binding and $y_p(g) = (1-d)g$.

This investment function implies that the planner's economy converges to one of two possible steady states. If $W/d \leq y_p^*(\delta, d, n)$, then the rate of depreciation is so high that the planner cannot reach $y_p^*(\delta, d, n)$ (except temporarily if the initial state is sufficiently large). In this case the steady state is $y_p^o = W/d$, and the planner invests all resources in all states on the equilibrium path. If $W/d > y_p^*(\delta, d, n)$, $y_p^*(\delta, d, n)$ is sustainable as a steady state. In this case, in the steady state $y_p^o = y_p^*(\delta, d, n)$, and the (per agent) level of private consumption is positive: $x_p^*(\delta, d, n) = (W + (1-d)g - y_p^*(\delta, d, n))/n > 0$. For the rest of the paper, we assume that $W/d > y_p^*(\delta, d, n)$. The analysis in the other case is similar.

III. Equilibrium with Irreversibility

In contrast to the planner's solution, in equilibrium no agent can directly choose the stock of public good y : an agent (say j) chooses only his own level of private consumption x and the level of its own contribution to the stock of public good. The agent realizes that in any period, given g and the other agents' level of private consumption, her contribution ultimately determines y . It is therefore as if agent j chooses x and y , subject to three feasibility constraints. The first constraint is a resource constraint that specifies the level of the public good $y = W + (1-d)g - [x + (n-1)x(g)]$. This constraint requires that the stock of public good y equals total resources, $W + (1-d)g$, minus the sum of private consumptions, $x + (n-1)x(g)$.

The function $x(g)$ is the *equilibrium* per capita level; naturally, the agent takes the equilibrium level of the other players, $(n - 1)x(g)$, as given. The second constraint requires that private consumption x is nonnegative. The third requires total consumption nx to be no larger than total resources W . Agent j 's problem can therefore be written as:

$$(2) \quad \max_{y,x} \left\{ \begin{array}{l} x + u(y) + \delta v(y) \\ \text{s.t. } x + y - (1 - d)g = W - (n - 1)x(g) \\ W - (n - 1)x(g) + (1 - d)g - y \geq 0, \text{ and } x \leq W/n \end{array} \right\},$$

where $v(g)$ is his equilibrium value function.

In a symmetric equilibrium, all agents consume the same fraction of resources, so agent j can assume that in state g the other agents each consume $x(g) = [W + (1 - d)g - y(g)]/n$, where $y(g)$ is the equilibrium investment function. Substituting the first constraint of (2) in the objective function, recognizing that agent j takes the strategies of the other agents as given, and ignoring constant terms, the agent's problem can be written as:

$$(3) \quad \max_y \left\{ \begin{array}{l} u(y) - y + \delta v(y) \\ y \leq \frac{W + (1 - d)g}{n} + \frac{n - 1}{n}y(g), \quad y \geq \frac{(1 - d)g}{n} + \frac{n - 1}{n}y(g) \end{array} \right\}.$$

To interpret the second constraint, note that it can be written as $y \geq (1 - d)g + \frac{n-1}{n} [y(g) - (1 - d)g]$: the new level of public good cannot be lower than $(1 - d)g$ plus the investments from all the other agents (in a symmetric equilibrium, an individual investment is $[y(g) - (1 - d)g]/n$). Similarly, the first constraint can be written as $y \leq (1 - d)g + \frac{n-1}{n} [y(g) - (1 - d)g] + W/n$: the new level of public good can not be larger than $(1 - d)g$, plus the investments from all the other agents, plus the maximal individual contribution W/n .

A continuous symmetric Markov equilibrium is fully described in this environment by two functions: an aggregate investment function $y(g)$, and an associated value function $v(g)$. The aggregate investment function $y(g)$ must solve (3) given $v(g)$. The value function $v(g)$ must be consistent with the agents' strategies, so, for all g :

$$(4) \quad v(g) = \frac{W + (1 - d)g - y(g)}{n} + u(y(g)) + \delta v(y(g)).$$

An equilibrium in an Irreversible Investment Economy is a pair of functions, $y(\cdot)$ and $v(\cdot)$, such that for all $g \geq 0$, $y(g)$ solves (3) given the value function $v(\cdot)$, and for all $g \geq 0$, $v(g)$ solves (4) given $y(g)$. For a given value function, if an equilibrium exists, the problem faced by an agent looks apparently similar to the problem of the planner, but with two important differences. First, in the objective function the agent does not internalize the effect of the public good on the other agents. This is the classic free rider problem, present in static models as well: it induces a suboptimal investment in g . The second difference with respect to the planner's problem is that the agent takes

the current and future contributions of the other agents as given. The incentives to invest depend on the agent's expectations about the other agents' current and future contributions, which are captured implicitly by the investment function $y(g)$.

To understand the complications associated with constructing an equilibrium with irreversibility, consider what happens when the irreversibility constraint becomes binding in an economy with zero depreciation. Suppose first that, as in the planner's problem, the value function is strictly concave. In this case the investment function looks very much like the planner's investment function. The agents find it optimal to invest as much as possible until the unconstrained optimum is feasible, say y^o ; and then they find it optimal to stay constant at this level until the irreversibility constraint becomes binding: $y(g) = y^o$ for $g \leq y^o$, so y^o is the steady state. The binding irreversibility constraint, however, forces the agent to choose a higher level of public good on the right of y^o : $y(g) = g > y^o$ for $g > y^o$. Because of the free rider problem, y^o is strictly lower than the efficient level: the players recognize that any increase in their investment crowds out the other players' investments in the following periods and so they underinvest. The irreversibility constraint acts as a beneficial commitment device and limits the ability of other players to *eat* additional investments. A marginal increase in g is more valuable on the right than on the left of the steady state y^o . However, this generates a contradiction since y^o would be suboptimal: by marginally increasing g at y^o the agents obtain a higher utility. In equilibrium we must have that the rate of investment on the right and on the left of the point at which the irreversibility constraint is binding are the same. The investment function therefore must be tangent to the irreversibility constraint $y = g$, as in the equilibrium represented in Figure 1. Satisfying this smooth pasting condition, however, is impossible with a strictly concave objective function because it requires that the agents are willing to invest at a sufficiently high speed on the left of y^o to guarantee that $y'(g) = 1$ at the point of contact with the irreversibility constraint. The challenge is to construct equilibria with this property.

In what follows, we proceed in two steps. First, to prove existence of an equilibrium we construct equilibria in which the value function has a flat top, as for example in Figure 1. We then prove that there is no loss of generality in focusing on this particular class in order to study the set of equilibrium steady states when depreciation is sufficiently low.

In an equilibrium in which the objective function has a flat region the investment function may take a more general form than the planner's solution. Figure 1 represents a typical equilibrium when $d = 0$ to illustrate the logic of the construction. The equilibrium investment function takes the following form:

$$(5) \quad y(g) = \begin{cases} \min\{W + g, \tilde{y}(g^2)\} & g < g^2 \\ \tilde{y}(g) & g \in [g^2, y^o], \\ g & g \geq y^o \end{cases}$$

where equilibrium investment is characterized by two critical levels, g^2 , y^o and an investment function $\tilde{y}(g)$, which is a nondecreasing function with values in $[g, W + g]$. To see why $y(g)$ may take the form of (5), consider Figure 1. The right panel of the figure illustrates a canonical equilibrium investment function. The steady state is labeled y^o in the figure, the point at which the (bold) investment function intersects

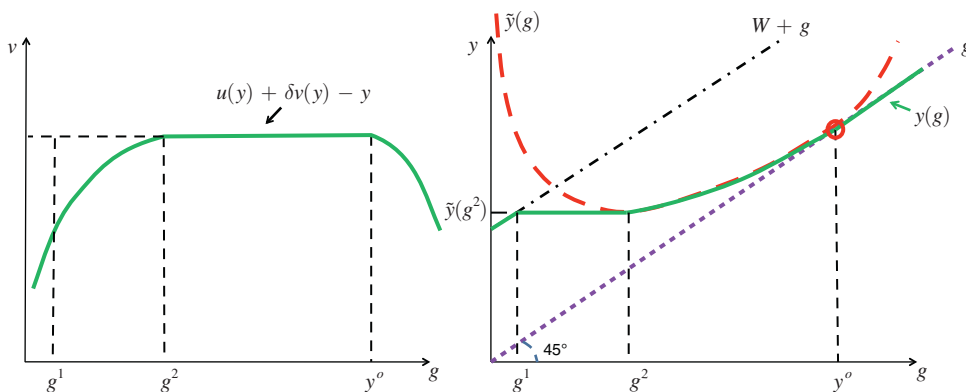


FIGURE 1. AN EXAMPLE OF FLAT TOP EQUILIBRIUM WITH NO DEPRECIATION

the (dotted) diagonal. The left panel graphs the corresponding objective function, $u(y) - y + \delta v(y)$. For $g < g^1$, the objective function of (3) is strictly increasing in y and resources are insufficient to reach the level that maximizes the unconstrained objective function.⁵ In this case it is optimal to invest all resources: $y(g) = W + g$ in $g \leq g^1$. For $g > y^o$, the objective function is decreasing. The agents would like to reduce g , but the irreversibility constraint is binding, so $y(g) = g$. For intermediate levels of $g \in [g^1, y^o]$, an interior level of investment $y \in (g^2, y^o)$ is chosen. This is possible because the objective function is flat in this region: an agent is indifferent between any $y \in [g^2, y^o]$. In the example of Figure 1, agents choose $y(g) = \tilde{y}(g^2)$ in $g \in [g^1, g^2]$, and the increasing function $\tilde{y}(g)$ in $y \in [g^2, y^o]$. The key observation here is that since the objective function has a flat region, the agents are willing to choose an *increasing* investment function in $[g^2, y^o]$ that is tangent to the irreversibility constraint.

Can we construct an investment function that makes the value function flat on top and that also satisfies the equilibrium conditions? For an investment curve as in Figure 1 to be an equilibrium, we need to make sure that the agents are indifferent between investing and consuming for all states in $[g^2, y^o]$. If this condition does not hold, the agents do not find it optimal to choose an interior level $y(g)$. The marginal utility of investment is zero if and only if

$$(6) \quad u'(g) + \delta v'(g) - 1 = 0 \quad \forall g \in [g^2, y^o].$$

Since the value function is (4), in the general case with $d \geq 0$ we have

$$(7) \quad v'(g) = \frac{1 - d - y'(g)}{n} + u'(y(g))y'(g) + \delta v'(y(g))y'(g).$$

Substituting this formula in (6), we see that the investment function $y(g)$ must solve the following differential equation in $[g^2, y^o]$:

$$(8) \quad \frac{1 - u'(g)}{\delta} = \frac{1 - d - y'(g)}{n} + u'(y(g))y'(g) + \delta v'(y(g))y'(g).$$

⁵In Figure 1 it is assumed that we have $W + g \geq g^2$ only for $g \geq g^1$, so the agent can afford to choose a level of y that maximizes the objective function (i.e., $y \in [g^2, g^3]$) if and only if $g \geq g^1$.

This condition is useful only if we eliminate the last (endogenous) term: $\delta v'(y(g))y'(g)$. To see why this is possible, note that $y(g)$ is in $[g^2, y^o]$ for any $g \in (g^2, y^o)$ in the example of Figure 1. In this case, (6) implies $\delta v'(y(g)) = 1 - u'(y(g))$. Substituting this condition in (8) we obtain the following necessary condition:

$$(9) \quad y'(g) = \frac{1 - d - \frac{n(1 - u'(g))}{\delta}}{1 - n}.$$

Condition (9) shows that there is a unique way to specify the shape of the investment function that is consistent with a *flat* objective function in equilibrium. This necessary condition, however, leaves considerable freedom to construct multiple equilibria; (9) defines a simple differential equation with a solution $\tilde{y}(g)$ unique up to a constant. An equilibrium $\tilde{y}(g)$ is pinned down when we impose the smooth pasting condition discussed above. Using (9), it can be verified that $y'(g) = 1$ at $g = [u']^{-1}(1 - \delta)$. The tangency condition $y(y^o) = y^o$ provides the initial condition for the differential equation (9), and so uniquely defines $\tilde{y}(g)$ in (5). The dashed line in Figure 1 represents this function.

The following proposition proves that an equilibrium exists for any $d \geq 0$ and indeed a continuum of equilibria exist for $d > 0$ (when $d = 0$ the upper and lower bounds of the equilibrium steady states in Proposition 1 coincide).⁶

PROPOSITION 1: *For any d, δ, n and $y^o \in [[u']^{-1}(1 - \delta(1 - d)), [u']^{-1}(1 - \delta(1 - \frac{d}{n}))]$, there is an equilibrium with steady state y^o in an irreversible economy. In all these equilibria convergence is monotonic and gradual.*

Perhaps unsurprisingly, for any $d > 0$ and $\delta < 1$, all equilibria in Proposition 1 are inefficient: first, the steady state is below optimum $y_p^*(\delta, d, n)$ (defined in Section III); second, convergence is inefficiently slow since convergence is in finite time in the planner's solution. The first type of inefficiency, however, disappears in the best Markov equilibrium as players become patient since the upper bound of the set characterized in Proposition 1 converges to the planner's first best:

COROLLARY 1: *For any d and n , the steady state in the most efficient equilibrium of an irreversible economy converges to the efficient level as $\delta \rightarrow 1$.*

Corollary 1 does not put bounds on the equilibrium set and so on potential inefficiencies in other equilibria. In the next result we show that multiplicity disappears as depreciation converges to zero. Let $\bar{y}_{IR}(\delta, d, n)$ and $\underline{y}_{IR}(\delta, d, n)$ be the supremum and the infimum of the set of equilibrium steady states. We have:

PROPOSITION 2: *For any δ and n , we have that $|\bar{y}_{IR}(\delta, d, n) - \underline{y}_{IR}(\delta, d, n)| \rightarrow 0$ as $d \rightarrow 0$. Moreover, there is a $\bar{d} > 0$ such that for $d < \bar{d}$, all equilibrium paths are gradual.*

⁶The details of the proofs of Proposition 1 and all the following results are in the online Appendix.

The first part of Proposition 2 shows that when depreciation is small, there is a small set of feasible equilibrium steady states. Indeed, the set of steady states converges to a singleton as depreciation converges to zero. This is a property entirely due to irreversibility since, as we discuss in the next section, with reversibility there is a continuum of equilibria for any level of depreciation (and for any δ and n as well). The second point shows that equilibria must look alike also in terms of the convergence path when d is sufficiently small: in all of them, convergence is gradual.⁷ This also is a property that is due to the irreversibility constraint, since in reversible economies we can have convergence in finite time, as we show below.

IV. Two Comparisons

In this section we discuss two natural comparisons to put the results of Propositions 1 and 2 in perspective. First, we compare the equilibrium sets with and without irreversibility; second, we compare Markov to subgame perfect equilibria.

A. Irreversible versus Reversible Economies

In economies with reversibility (RIE) the agents can choose a negative level of investment and so the public good can be reduced below $(1 - d)g_t$. The feasibility constraints defining the economy become: (i) $x_t^j \geq 0 \forall j, \forall t$; (ii) $y_t \geq 0$; and (iii) $\sum_{j=1}^n x_t^j + [y_t - (1 - d)g_{t-1}] \leq W \forall t$. In Battaglini, Nunnari, and Palfrey (2012) we characterize the symmetric equilibria of the dynamic free rider game in these economies assuming that the agents withdraw up to $1/n$ of the accumulated stock. Let $\bar{y}_R(\delta, d, n)$ and $\underline{y}_R(\delta, d, n)$ be the upper and lower bounds of the set of steady states in a RIE. In Battaglini, Nunnari, and Palfrey (2012) we show that $\bar{y}_R(\delta, d, n) = [u']^{-1}(1 - \delta(1 - d/n))$ and $\underline{y}_R(\delta, d, n) \leq [u']^{-1}(1 - \delta(1 - d)/n)$. This allows a neat comparison with the equilibria in an irreversible economy. Let E_R and E_{IR} be the sets of equilibrium steady states respectively with and without reversibility. The following result is an immediate implication of Propositions 1–2:

PROPOSITION 3: *For any δ and n , there is a $\bar{d} > 0$ such that $E_{IR} \subset E_R$ for all $d < \bar{d}$. Moreover, $|\underline{y}_{IR}(\delta, d, n) - \bar{y}_R(\delta, d, n)| \rightarrow 0$ as $d \rightarrow 0$.*

Proposition 3 makes clear that irreversibility has two effects. First, it reduces the set of equilibrium steady states, that with irreversibility is *strictly* included in the set with reversibility if depreciation is sufficiently small. Second, and more importantly, the irreversibility constraint eliminates the inefficient equilibria. When the economy is reversible, there are always very inefficient equilibria, with steady states not larger than $[u']^{-1}(1 - \delta(1 - d)/n)$ that are worse than the steady states reached by an agent in autarky (that is, alone by herself). On the other hand, as $d \rightarrow 0$, the lower bound of the steady states with irreversibility converges to the upper bound of the set

⁷The property that the convergence path is gradual has been highlighted as a general feature of subgame perfect equilibria with irreversibility by Lockwood and Thomas (2002). As we prove in Proposition 3, their result requires an assumption that d is exactly zero. When $d > 0$, graduality of the equilibrium path is not a general feature of subgame perfect equilibria.

with reversibility. In addition to these two effects, Battaglini, Nunnari, and Palfrey (2012) show that with reversibility there is always an (inefficient) equilibrium in which convergence to the steady state is not gradual. This shows that graduality is a feature associated to irreversibility.

The fact that reversibility affects so much of the equilibrium set may appear surprising. In a planner's solution the irreversibility constraint is irrelevant: it affects neither the steady state (that is unique) nor the convergence path.⁸ Even in economies with irreversibility, the irreversibility constraint is typically never binding on the equilibrium path starting from a low level of g (see Figure 1 for example). The reason why irreversibility is so important in a dynamic free rider game is precisely the fact that equilibrium investments are inefficiently low and the irreversibility constraint may limit the inefficiency by acting as a *commitment device*. The intuition is as follows. In the equilibria of a dynamic free rider problem, both with and without irreversibility, the agents hold back their individual contributions for fear that they will crowd out the contributions of other players, or even be appropriated by other agents in future periods. In economies with irreversibility, however, the irreversibility constraint limits the ability of the agents to appropriate the accumulated public good. In general, the irreversibility constraint is binding only for states that are so high that they are not reached on the equilibrium path; still, the fact that in these states free riding will be limited affects the entire equilibrium investment function. In states just below the point in which the constraint is binding, the agents know that the constraint will not allow the other agents to reduce the public good when it passes the threshold. These incentives induce higher investments and a higher value function, with a ripple effect on the entire investment function.

B. Subgame Perfect Equilibria versus Markov Equilibrium

Previous to our work, the effects of irreversibility have been studied in the literature on monotone games, in particular by Lockwood and Thomas (2002) and Matthews (2013).⁹ Both papers make two key assumptions: (i) a zero contribution is a dominant strategy for all players, and so the game can be reduced to a repeated version of a prisoners' dilemma game;¹⁰ (ii) the state variable g can only stay constant or increase because depreciation is exactly zero. Under these conditions Lockwood and Thomas (2002) have characterized the most efficient subgame perfect equilibrium (SPE) and Matthews (2013) has characterized all SPE.¹¹ This literature arrives at a conclusion that is opposite to ours: while we show irreversibility has a positive

⁸ Assuming the initial state g_0 is smaller than the steady state, on the convergent path the stock of public good is never reduced; it keeps increasing until the steady state is reached, and then it stops. The irreversibility constraint is, thus, never binding on the equilibrium path.

⁹ See footnote 2 for other important contributions in the monotone games literature that are not as closely related to our paper.

¹⁰ In our model we assume standard preferences $u(g)$ satisfying the Inada condition ($u'(g) \rightarrow \infty$ as $g \rightarrow 0$ and $u'(g) \rightarrow 0$ as $g \rightarrow \infty$). In this case the dominant strategy assumption is never satisfied since the players find it optimal to make strictly positive contributions, no matter what the other players do.

¹¹ Matthews (2013) presents the most general analysis of monotone games to date. Although Matthews does not allow for depreciation, his model assumes very general specifications for the players' preferences and makes weak assumptions on the timing of contributions. For this general version of the model, Matthews provides a necessary condition for a SPE; however, for a characterization of the SPE, this requires the same assumptions as in Lockwood and Thomas (2002).

welfare effect, they show it has a negative effect. They prove that SPE in economies with irreversibility are more inefficient than in economies with reversibility, at least when agents are sufficiently patient. This finding is shown by proving an *Anti-Folk Theorem*: with reversibility, the efficient allocation is achievable in a SPE for a sufficiently high discount factor (but less than one); with irreversibility, SPE are inefficient for all $\delta < 1$.¹² This literature has also stressed the conclusion that the most efficient equilibrium path is characterized by gradualism (and hence inefficiently slow) if and only if the investment is irreversible. The central new finding in this literature is the fact that *any comparison at all* can be made focusing on the large set of subgame perfect equilibria. This is interesting because the set of SPE is not generally very informative in games with perfect information for reasons similar to standard folk theorems for repeated games.

With this background, we next compare the most efficient SPE paths in RIE and IIE under the assumptions of our model, allowing a positive (but possibly arbitrarily small) rate of depreciation. We say that an investment path is a *SPE path* if it coincides with the equilibrium path of a SPE.

PROPOSITION 4: *For any $d > 0$ and $n > 1$, there is a $\bar{\delta} < 1$ such that the most efficient SPE path in a RIE and the most efficient SPE path in an IIE both coincide with the Pareto efficient investment path for all $\delta > \bar{\delta}$. Hence, neither the most efficient SPE path in a RIE nor the most efficient SPE path in an IIE are characterized by gradualism for all $\delta > \bar{\delta}$.*

Proposition 4 makes clear that the results on the effects of the irreversibility constraint obtained in the literature on monotone games critically depend on the assumption that depreciation is exactly zero; they can not be interpreted as results describing economies with a small—perhaps arbitrarily small—degree of depreciation, because when depreciation is not zero, results are *qualitatively* different.¹³ Perhaps more seriously, Proposition 4 suggests that when $d > 0$ a comparison between the most efficient SPE in RIE and IIE with patient players is no longer insightful, since the efficient allocation is an equilibrium in both environments.

To see why the rate of depreciation is so important, consider for simplicity the environment in Lockwood and Thomas (2002), where a zero contribution is a dominant strategy (an assumption that is not used in the proof of Proposition 4).¹⁴ When $d = 0$, the worst punishment for an agent is that all other agents stop making contributions, which is always the worst equilibrium by assumption (since zero contribution is a dominant strategy). But with $d = 0$ this punishment becomes increasingly irrelevant as g approaches the efficient steady state; in the worst case, a deviation is punished by an allocation that remains forever close to the efficient allocation. This

¹²The intuition behind this result is that irreversibility limits the ability of agents to punish each other; in the worst continuation equilibrium players stop making contributions, but they can not eat or destroy the accumulated state g .

¹³Lockwood and Thomas's (2002) result can be extended to the case of arbitrarily small amounts of depreciation, but only if the order of limits is reversed, i.e., by fixing δ and taking d to zero. Thus, it is always possible to find a sequence d^n, δ^n converging to $\{0, 1\}$ such that the efficient allocation path is a SPE path for any n both in a RIE and in an IIE: it follows that, for low d and high δ , a sharp comparison between the two is not possible.

¹⁴For consistency, Proposition 4 is stated under the assumptions described in Section II according to which players do not have a dominant strategy. The strategy of our proof can be immediately adapted (in a simplified form) to the environment considered by Lockwood and Thomas (2002).

is the reason why irreversibility makes it impossible to obtain an efficient allocation in Lockwood and Thomas (2002). The environment is very different if d is positive, even if arbitrarily small. In this case, after a deviation, the state would gradually decline and it would eventually approach zero. This convergence may certainly be slow when d is small; but as $\delta \rightarrow 1$ only the long run matters, so the punishment for a deviation is sufficiently high to induce all players to make the efficient contribution.

V. Conclusions

In this paper we have studied a simple model of free riding in which n infinitely lived agents choose between private consumption and contributions to a durable public good. We study the properties of the Markov equilibria in which contributions to the public good are irreversible. We present three main results. First, we show that, no matter what the depreciation rate is, the steady state of the best Markov equilibrium converges to the efficient level as the agents' discount factor converges to one. Second, as depreciation converges to zero, the set of steady states converges to a unique point. It follows that, when the discount factor is high and depreciation is low, all equilibrium steady states are approximately efficient. Third, however, convergence to the steady state is always inefficiently slow. From these results we conclude that, with irreversible investments, multiplicity of steady-state equilibria is not problematic when depreciation is small; and that the problem with dynamic free riding is not so much an inefficient steady state, but an inefficiently slow accumulation path. Finally, we compare these results with the more commonly studied case with reversible investment. We show that irreversibility has two effects. First, it reduces the set of equilibrium steady states that with irreversibility is strictly included in the set with reversibility if depreciation is sufficiently small. Second, and more importantly, the irreversibility constraint eliminates the inefficient equilibria which, with reversible investments, are possible for any level of depreciation, discount factor and group size.

REFERENCES

- Admati, Anat R., and Motty Perry. 1991. "Joint Projects without Commitment." *Review of Economic Studies* 58 (2): 259–76.
- Battaglini, Marco, and Stephen Coate. 2007. "Inefficiency in Legislative Policy-Making: A Dynamic Analysis." *American Economic Review* 97 (1): 118–49.
- Battaglini, Marco, Salvatore Nunnari, and Thomas Palfrey. 2012. "The Free Rider Problem: A Dynamic Analysis." National Bureau of Economic Research Working Paper 17926.
- Besley, Tim, and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton: Princeton University Press.
- Compte, Olivier, and Philippe Jehiel. 2004. "Gradualism in Bargaining and Contribution Games." *Review of Economic Studies* 71 (4): 975–1000.
- Dockner Engelbert J., and Ngo Van Long. 1993. "International Pollution Control: Cooperative versus Noncooperative Strategies." *Journal of Environmental Economics and Management* 25: (1) 13–29.
- Dutta, Prajit K., and Roy Radner. 2004. "Self-Enforcing Climate-Change Treaties." *Proceedings of the National Academy of Sciences of the United States of America* 101 (14): 5174–79.
- Fershtman, Chaim, and Shmuel Nitzan. 1991. "Dynamic Voluntary Provision of Public Goods." *European Economic Review* 35 (5): 1057–67.
- Gale, Douglas. 2001. "Monotone Games with Positive Spillovers." *Games and Economic Behavior* 37 (2): 295–320.
- Levhari, David, and Leonard J. Mirman. 1980. "The Great Fish War: An Example Using Cournot-Nash Solution." *Bell Journal of Economics* 11 (1): 322–34.

- Lockwood, Ben, and Jonathan P. Thomas.** 2002. "Gradualism and Irreversibility." *Review of Economic Studies* 69 (2): 339–56.
- Marx, Leslie M., and Steven A. Matthews.** 2000. "Dynamic Voluntary Contribution to a Public Project." *Review of Economic Studies* 67 (2): 327–58.
- Matthews, Steven A.** 2013. "Achievable Outcomes of Dynamic Contribution Games." *Theoretical Economics* 8 (2): 365–403.
- Rubio, Santiago J., and Begoña Casino.** 2002. "A Note on Cooperative versus Non-Cooperative Strategies in International Pollution Control." *Resource and Energy Economics* 24 (3): 251–61.
- Stokey, Nancy L, Robert E. Lucas, Jr., and Edward C. Prescott.** 1989. *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.