Chapter 10

Collective Action and Cooperation

The literature on collective action in political science and economics is large, but the relationship between heterogeneity and inequality among the individual or group actors and collective action is still relatively under-researched. At the end of chapter 2 we gave some general indications why this is an important area in the discussion of institutional failures in developing countries. In this chapter we discuss this relationship in some detail, first with respect to examples from the recent macro-economic literature, particularly on fiscal crisis in section II, and then in sections III and IV we construct a model to focus more on microeconomic issues that may arise in the matter of collective action in management of local environmental resources or in firms benefiting from a common provision of infrastructural services. We also discuss the commonality and difference in the two types of models considered in the two parts of this chapter. In the concluding section V we review the main issues of this chapter.

Nearly 20 years back I published a book --Bardhan (1984)-- where I looked at the fiscal crisis of the Indian state as the result of an intricate collective action problem (in an implicit analytical framework of a non-cooperative Nash equilibrium) among its fragmented elite groups, who could not get their act together in, for example, generating the public savings
that could be invested in correcting the severe deficiencies in public infrastructure from which all the groups could ultimately gain, instead of dissipating the surplus of the economy on short-run subsidies and unproductive expenditures out of the public-fisc common pool. In the last decade there has been some theoretical literature on the ‘common pool problem’ in fiscal policy in the context of fiscal stabilization in Latin America. In this section and the next we’ll review this literature and connect it with our general theme. The general analytic setup for the type of common pool problem considered here is an infinitely repeated game between a number of powerful interest groups that compete over a common pool of public saving. The two critical assumptions are: 1) that each interest group is powerful to the extent that it is able to draw from the common pool of public saving without the consent of other groups and according to its own discretion (within some exogenously determined limits); and 2) that the costs of spending by any one group (e.g., the rate of interest on public debt, or inflation taxes generated by seignorage) are borne uniformly by all groups. As a result, we have a situation that lends itself to aggregate overspending: those with the power to control public spending compare its (private) marginal benefits against only a fraction of its (social) marginal costs.

In what contexts might special interests exert this degree of control over the allocation of public resources? Velasco (1998) mentions two contexts that may be of interest to us. First, when governments cannot credibly commit to implement efficient re-financing decisions with respect to state-supported firms, the resulting soft budget constraint problem may place the managers of those firms in a position to have access to public funds with relative ease. Second, in countries in which state or provincial governments are strong in relation to the central government—Velasco mentions the examples of Argentina and Brazil in
particular—there are a number of mechanisms through which local governments might "pass on" their deficits so that they become the *de facto* responsibility of federal authorities (and thus of taxpayers at large); Velasco describes three possible mechanisms:

(a) borrowing from state development banks that in turn could rediscount their loans at the Central Bank—in effect monetizing the subfederal deficits;

(b) obtaining discretionary lump sum transfers from the federal government, generally requested around election time and after large debts have been accumulated;

and (c) accumulating arrears with suppliers and creditors, which (for legal or political reasons) were eventually cleared up by the federal authorities.

Finally, Velasco cites a number of cross-country empirical studies which have yielded results consistent with the hypothesis that decentralized control of fiscal policy erodes fiscal discipline (in particular, Von Hagen (1992), Von Hagen and Harden (1994), and Alesina, Hausmann, Hommes, and Stein (1996)).

Let us identify some distinctions between the general analytic setup of the common pool problems considered here and that of the credible commitment problems we discussed in chapter 4. One important distinction is that, while incentives to "defect" are in general *one-sided* in models of credible commitment, they tend to be *multilateral* in models of common pool. To develop this point further, it is instructive to think of most problems of credible commitment as something resembling a *sequential* prisoners' dilemma game, in which one player must choose between "cooperation" or "defection" (or analogous concepts particular to a given context) before the other player makes his choice between the two options. A
solution to this sequential prisoners' dilemma can be found in a credible commitment by the second-moving player to choose cooperation contingent on the first-moving player choosing the same. Such a commitment simplifies the payoff schedule facing the first-moving player (the one to whom a commitment has been made) to two -- rather than four -- possible payoff outcomes: (cooperate, cooperate), and (defect, defect). Under the assumption that mutual cooperation generates higher payoffs for everyone than mutual defection, a credible commitment by the second-moving to cooperate is therefore sufficient to eliminate the first-moving player's incentives to defect. Cooperative norms in the context of credible commitment therefore hinge only on the second moving player's incentives to defect -- a unilateral incentive problem.

On the other hand, it may be useful to think of most problems of common pool as a simultaneous prisoners' dilemma game, in which all players must choose at the same time between cooperation and defection. As in the standard prisoners' dilemma, and in contrast to the sequential prisoners' dilemma that characterizes most models of credible commitment, a credible commitment by any given player to cooperate is insufficient in the context of the common pool problem to eliminate the defection incentives of the other player(s). The reason is that, due to the simultaneous nature of the game, a commitment by any given player to choose cooperation simplifies his opponents' payoff schedule to the two possible payoff outcomes (cooperate, cooperate) and (defect, cooperate), rather than the two outcomes (cooperate, cooperate) and (defect, defect), and so defection remains a dominant strategy for those to whom commitments have been made. In other words, each player must not only receive a commitment but also make one of his own if cooperation is to be achieved.
Cooperative norms in the context of the common pool problem must therefore take account of every player’s incentives to defect—a multilateral incentive problem.

However, note that the multilateral incentive problem that characterizes the common pool problem may be transformed into a credible commitment-type unilateral incentive problem through mechanisms that transfer the burden of commitment to an outside party, such as a moderator who promises to compensate individual decision makers for engaging in disciplined consumption (or, equivalently, threatens to punish them for appropriative behavior). For example, suppose that there are \( n \) claimants on a common pool, and that the decentralized outcome yields a social surplus of \( D \) while the fully cooperative outcome yields a social surplus of \( C, C > D \): cooperation might be sustained through a credible commitment by a third party to pay each individual decision maker \( C/n \) conditional on that individual choosing to cooperate. The role of behavior-contingent compensation by third parties in overcoming problems of common pool is examined by the literature on conditionality in foreign aid (surveyed by Drazen (2000)).

The models of common pool considered in this chapter share two common elements. The first is that discretionary appropriation of public funds by any given interest group yields short term payoffs and long term costs. In some cases, these costs arise purely out of the economy’s production function, while in others, explicit punishment costs for norm-violation are introduced through trigger strategies or other disciplinary mechanisms. The second shared element of the models is the assumption that the various interest groups competing over the common pool of public saving are essentially identical, to the extent that appropriation incentives are symmetric across groups in equilibrium (implying, in cases where
cooperation is modeled explicitly, that either all or none of the groups' incentive compatibility constraints are satisfied in any given state of nature). This assumption stands in contrast with that in our working model in the next section on the relationship between inequality and cooperation, where we try to understand the channels through which economic asymmetries among prospective collaborators (and the asymmetric payoffs, punishment costs, and incentive compatibility constraints they imply) affect the sustainability of cooperation.

Given the symmetric nature of agent behavior in the common pool models considered in this section, a primary objective of the models is (i) to identify the conditions under which sustainable consumption from the common pool can be achieved, and in some cases, (ii) to identify mechanisms through which the economy may switch from one state (being on either a sustainable path or an unsustainable one) to the other as part of a dynamic process. The models differ in important ways with regard to their predictions of when and where sustainable consumption is most likely to arise, and in particular, with regard to whether sustainable consumption is more likely to be achieved when assets to be appropriated are low (i.e., when the common pool is nearly depleted) or high (i.e., when the common pool is rich in resources). As Drazen (2000) observes, whether it is high or low levels of wealth that better facilitate cooperation is determined primarily by model specification, including functional forms. However, let us make a quick observation about the two types of cases before moving on to the models in the next section. In cases where cooperation is more easily achieved at high levels of wealth, initial conditions matter, giving rise to the possibility of multiple equilibria: economies that start off with sufficiently high levels of wealth tend to be predisposed towards virtuous cycles of cooperation and wealth accumulation, whereas
economies that start off poor tend to be doomed to uncooperative and growth-stifling equilibrium paths. In contrast, in cases where cooperation is more easily achieved at low levels of wealth, initial conditions tend to be irrelevant: regardless of starting position, economies tend to converge to a stable equilibrium path once common resources fall to a certain threshold level and cooperation is triggered.

II*

Let us begin with the model of Mondino et al (1998), which provides an example in which sustainable consumption is only triggered after available resources have fallen to a sufficiently low level. Mondino and his co-authors are interested in explaining recurrent inflation-stabilization cycles, wherein economies that have experienced prolonged spans of mounting inflation respond with dramatic policy changes to right the economic ship, only to once again start down unsustainable policy trajectories and begin the process anew. In their model, agents at large are organized into two pressure groups of equal size. In each period of an infinitely repeated game, each pressure group may demand from the government a subsidy of fixed size $S$ for its members. This assumption implies that the benefits of "grabbing" from the common pool are fixed, and allows Mondino and his co-authors to focus on how the costs of grabbing evolve over time. Three additional assumptions complete the basic setup of the model. First, the government always accommodates every subsidy demand; second, it does
so by printing money, which generates an inflation tax that must be borne by everyone holding the domestic currency; and third, in order to escape the inflation tax associated with seignorage, agents at large have the ability to transfer any portion of their wealth out of the domestic currency and into alternative assets which are shielded from domestic inflation, but are otherwise comparatively inconvenient stores of wealth—the overall extent to which agents choose to hold their wealth in these alternative assets is what Mondino and his co-authors call the economy's degree of "financial adaptation."

The basic intuition behind the emergence of recurrent cycles of inflation followed by stabilization may now be summarized. Suppose that the economy starts off in a position where monetization is high (implying that inflation must be low, since agents are willing to hold large amounts of the domestic currency). Note that since the stock of domestic currency in circulation composes the "tax base" for seignorage, a high degree of monetization implies that a given amount of seignorage may be financed by a relatively low inflation tax. Thus under current conditions (high monetization), the interest groups are likely to find that the marginal benefits of seignorage-financed government spending exceed its marginal costs. But over time, as pork-barrel spending continues to characterize the equilibrium of the stage game, inflation starts to mount, and agents respond by steadily increasing their financial adaptation. This raises the costs of continued inflation, since 1) financial adaptation is costly in and of itself (recall that agents prefer to store their wealth in the domestic currency, all else equal), and 2) continued seignorage in the face of increasing financial adaptation must be financed by an ever-decreasing tax base of domestic currency. Mondino and his co-authors assume that there is a non-convexity in the cost schedule of financial adaptation, in that any agent who
chooses to \textit{completely} substitute out of the domestic currency must incur a substantial fixed cost $K > S$. This assumption ensures that the economy will eventually reach a threshold level of inflation after which the cost of continued inflation (and thus of further financial adaptation) exceeds the one-period benefit from receiving the government subsidy $S$. Once the economy reaches this threshold point, both pressure groups will find it optimal to accept a stabilization package (which is modeled simply as zero subsidies for that period) that brings the economy back to its low-inflation starting point. But since it is assumed that agents prefer, all else equal, to store their wealth in the domestic currency, stabilization brings us back to the initial high monetization equilibrium, and the process begins anew.

The model of Mondino \textit{et al} is fully decentralized, so that a switch from unsustainable to sustainable consumption occurs if and only if it suits the private interests of individual decision makers. The model is specified in such a way that the benefits of grabbing from the common pool are fixed (the fixed subsidy $S$), whereas the costs of grabbing steadily increase over time so long as everyone continues to grab—as a result, grabbing ceases only after the common pool has been sufficiently depleted to bring the costs of additional grabbing above its fixed benefits. Velasco (1998 and 1999) reaches a similar conclusion in a model that introduces explicit punishments for non-cooperative behavior. As in the previous model, grabbing in the Velasco model exhibits increasing costs over time so long as everyone continues to grab; but in addition, the benefits that can be secured through grabbing also decrease over time with continued grabbing. The reason is that discretionary grabbing reduces the income flow accruing to public assets, and therefore reduces the pool of resources to be split among a given number of claimants.
The government in Velasco's model is composed of a number of powerful fiscal authorities, each representing a different interest group at large. These fiscal authorities are powerful in the sense that each may simply set the level of public spending to be devoted to his particular constituency, so long as aggregate spending in any one period does not violate a financial solvency condition (which prevents government debt from growing without bound). Public spending by any one group benefits that group exclusively, but must be financed with a common pool of public saving; in other words, the benefits of public spending are group-specific, while the costs are shared by all members of society. Given this setup, it is obvious that the economy will exhibit a tendency towards aggregate overspending, as each fiscal authority weighs the marginal benefit of spending against only a fraction of its social marginal cost; more specifically, each sees the cost of spending not as the rate of interest on public saving, but as the rate of interest minus what the others will either choose to take out (if public saving is positive) or have to pay (if public saving is negative). In a dynamic setting, which is Velasco's focus, this means that the economy may accumulate debt even in contexts in which there is no social incentive to engage in intertemporal smoothing and the practice of running deficits creates deadweight losses for the economy as a whole.

As is always the case in this type of setting, reputational mechanisms may help to enforce cooperative outcomes that leave everyone better off. Velasco formalizes the conditions under which cooperation between the fiscal authorities can be sustained as a trigger strategy equilibrium, in which defection by any one fiscal authority is punished by a multilateral reversion to the uncooperative equilibrium. He shows that cooperation is only sustainable after public resources have been depleted beyond a threshold point. At low levels
of public debt (or when resources to be appropriated are plentiful), the deadweight losses from additional debt accumulation are small in relation to the benefits of aggressive spending from a large asset stock, so it pays to deviate from the cooperative policy trajectory. As debt accumulates, the balance starts to shift: the deadweight losses from continued debt accumulation become more severe, and at the same time, the government's financial solvency condition becomes more restrictive, meaning that there are less funds available to be appropriated through deviation. Only when the private payoff from stabilization eventually exceeds that of deviation does the cooperative equilibrium becomes sustainable under trigger strategies.

In contrast to Mondino et al and Velasco, Aizenman (1998) provides a model in which sustainable consumption is more likely to be achieved when the common pool is rich rather than depleted, raising the possibility of multiple equilibria. Aizenman models fiscal policymaking in a union. The model is similar to the Velasco model—to the extent that public resources are up for grabs before a number of powerful fiscal authorities—but there are two important added features: first, there now exists a central budget director that tries to coordinate the behavior of the numerous fiscal authorities; and second, the fiscal authorities now have personal motives to secure public spending, so that political representation is now subject to an agency problem. This is a rather complex model, and let us cut straight to its core result. The government consists of a budget director and a large number of governors, each of whom represents a province or state belonging to the union. At the start of each period, the budget director determines the government's planned level of aggregate expenditure for that period, of which each governor is granted an equal share. Each governor's discretion
lies in whether he will comply with this spending limit, or instead, rely on public debt in order to finance additional spending. Governors who violate their assigned budgets face a risk of detection by the budget director \textit{ex post} (that is, after the offending governors have secured whatever personal gains are associated with spending). Any governor who is caught violating his allotted budget is expelled from office. As a result, governors are less likely to engage in excess spending when the long-term benefits of remaining in office are perceived to be greater.

The public has preferences against debt accumulation, and, like the budget director, can also punish fiscally undisciplined policymakers by removing them from office. The model's most critical assumption involves how this punishment is carried out. Specifically, Aizenman assumes that the public removes \textit{all} governors from office in a single stroke when the public debt is perceived to have reached an inordinately high level.\footnote{This assumption is incorporated into the model as follows: before the start of any given period, the \textit{probability} that the incumbent administration (meaning all governors) will be allowed to remain in office is a decreasing function of the current level of public debt.} Aizenman interprets this assumption to mean that the public uses the level of public debt as a basis for judging the competence of the existing administration as a whole. An alternative interpretation is that the public's tendency to levy wholesale rather than targeted punishments upon incumbent policymakers reflects constraints that are imposed either by informational disadvantages or by the political process itself. The second interpretation may be preferred, since, in the absence of \textit{ex ante} heterogeneity across prospective administrations, it is not clear why the public would choose to adopt this practice—which further externalizes the costs of "bad behavior" by individual policymakers—when it could secure better outcomes by being more
discriminating in choosing whom to punish. In any case, we may now see why the economy in the Aizenman model is likely to exhibit excessive public spending at precisely the worst possible times. Specifically, at times when the public debt is already substantial, each governor recognizes that his chances of remaining in office are slim regardless of whether he exercises fiscal discipline—due to the fact that all governors are likely to be ousted by the public in the very near future in any case. This induces all governors to shift the focus of their intertemporal cost-benefit analysis towards the short run, and thereby raises the comparative appeal of securing a short-run windfall through aggressive spending relative to sacrificing current consumption in an unlikely bid for job security. In other words, at high levels of public debt, each governor perceives that his days are limited, and opts to go out in a final fiscal blaze of glory. Note that the budget director's goal to avert this type of race to the bottom may require him to set an otherwise inefficiently high target level for the public debt. The reason is that a high debt target increases the payoffs that accrue to governors in the cooperative equilibrium, and thereby discourages deviation.

Finally, let us briefly consider the model of Benhabib and Rustichini (1996), which is similar in structure to the Velasco model and, like the Velasco model, also examines cooperative norms of sustainable consumption based on trigger strategies. What distinguishes the Benhabib and Rustichini model is that cooperation is in some cases more easily achieved at low levels of wealth and in other cases more easily achieved at high levels of wealth (whereas in the Velasco model, cooperation is only possible at low levels of wealth). Whether one or the other situation characterizes the economy is determined by the parameters of the model, and in particular, the curvature of the production function (where
production is increasing in the stock of a common pool of capital) relative to the curvature of agents' utility functions. (where utility is increasing in consumption of existing capital). When returns to consumption diminish at a high enough rate relative to the rate at which returns to capital diminish, cooperation is more likely achieved under "desirable" conditions, when wealth and consumption are both high. The reason is that, when the curvature of the utility function is pronounced in relation to the curvature of the production function, the marginal return to consumption is high relative to the marginal return to capital when both consumption and wealth are low (so that the benefit of appropriating additional consumption is high relative to the cost of future retaliation by other groups), while the marginal return to consumption is low relative to the marginal return to capital when both consumption and wealth are high (so that the benefit of appropriating additional consumption is low relative to its punishment costs). Conversely, and by symmetric logic, cooperation is more likely achieved under "duress" (when wealth and consumption are both low) when returns to capital diminish at a high rate relative to the rate at which returns to consumption diminish.

III

In this section we focus our attention more to the micro context and deal with inequality rather than group heterogeneity per se. The model presented here aims to delineate some possible mechanisms through which economic inequality might determine the necessary features as well as overall sustainability of social norms of cooperation. The model stays
neutral on the important but contested issue of the direct effects of inequality on aggregate economic outcomes, and instead limits its focus to indirect effects that might operate via social norms. In particular, the model is specified in such a way that social welfare depends only on whether or not cooperation takes place; as a result, inequality affects social welfare if and only if it affects the sustainability of social norms of cooperation.

It is important to note that the model considers distributional inequality with respect to a nontransferable resource, which we simply call private capital. Relatedly, it considers a setting in which cooperation promises to yield efficiency gains relative to the status quo not by promoting efficiency-enhancing redistributions of capital (i.e., from where its marginal return is already low to where it is still high), but instead, by promoting the adoption of efficient technologies for producing capital-complementing inputs. Examples of such inputs could be R&D, irrigation, infrastructural services, etc. The need for cooperative norms arises in the model out of a common pool problem in the production of the capital-complementing inputs. When agents combine their efforts to produce the inputs in an efficient manner, the fruits of their combined labor can be freely appropriated by any individual claimant; as a result, agents choose in decentralized equilibrium to adopt alternative production technologies that are economically inferior but free of the common pool problem.

Like the models of common pool that we have reviewed from the fiscal policy literature, the basic analytic setting is an infinitely-repeated game, and the prospective social norm of cooperation is one based on trigger strategies: at any given point in time, all agents are provided with the prospect of continued cooperation into the future so long as no one agent defects at the present. Unlike those models, which consider the effects of various state
variables (e.g., the economy's aggregate wealth) on the sustainability of cooperation among identical agents, our primary focus here is on the implications of asymmetry across agents—caused by distributional inequality—and our primary state variable of interest is the one that describes the extent of this asymmetry. As one result, whereas the perceived desirability of cooperation relative to defection is symmetric across players in standard models of common pool (so that cooperation is enforceable through trigger strategies for either everyone or no one), incentives for cooperation tend to differ across players in our model.

In our model, incentives for cooperation tend to be stronger for wealthier agents (all else equal) and there is always a threshold level of distributional inequality beyond which cooperation is enforceable for the richly endowed and unenforceable for the poorly endowed. This result is similar to that of common pool models in which cooperation is more easily sustained in times of plenty, due to the fact that wealth accumulation and incentives for maintaining cooperation are mutually reinforcing (Aizenman (1998), Benhabib and Rustichini (1996)). However, at intermediate levels of distributional inequality, it is not always the richly endowed who are willing to cooperate and the poorly endowed who are not. This is due to another point of departure between our model and standard models of common pool: we assume that agents have access to asymmetric rather than symmetric exit options. And when the richly endowed have access to sufficiently attractive exit options relative to those of the poorly endowed, it may be the rich who fail to exercise discipline even though they are the ones who stand to gain the most in absolute terms from cooperation. These differential exit options are particularly important in the context of management of environmental
resources in the rural areas of developing countries. For example, as we illustrate in our next Chapter, rich farmers with better urban connections or better access to alternative water supplies (say, from privately owned pumps) may be less interested in fostering cooperation on management of public irrigation systems. One may have a similar story on conservation of resources in a fishery with differential exit options for fishermen with different boat sizes.² For a general survey of the existing small literature on the role of inequality in collective action on the local commons, see Baland and Platteau (forthcoming).

Whether it is the rich or the poor who face stronger incentives to defect can have important implications for whether or not cooperation can be salvaged in some instances with the aid of sidepayments. In particular, we show that inequality acts to diminish the incentive effects of any given side-payment on the richly endowed, and to magnify its effects on the poorly endowed. As a result, cooperation aided by sidepayments tends to be more feasible when it is the poor and not the rich who must be compensated.

² The model of Dayton-Johnson and Bardhan (2002) examines the effect of inequality in different boat sizes on conservation in a fishery with differential exit options for big boat-owning fishermen and with linear technology.
The Model

Consider the following infinitely repeated game between two agents, one "rich" and one "poor." Time proceeds in discrete periods, and the agents have a common discount factor of $\beta$.

Stage Game

The economy enters each new period with a total private capital endowment of $K$, of which the rich agent controls share $\rho$ and the poor agent controls share $1 - \rho$, where $\rho > 0.5$. Each agent's utility is increasing in his consumption of own-produced output, $Y_i$, and decreasing in his effort level, $e_i$:

$$U_i = Y_i - ce_i$$  \hspace{1cm} (1)

Production of output $Y_i$ is assumed Constant-Returns-to-Scale Cobb-Douglas with respect to private capital $K_i$ and an unnamed complementary input, labeled $A_i$:

$$Y_i = A_i^\alpha K_i^{(1 - \alpha)}$$  \hspace{1cm} (2)

Property rights with respect to private capital are assumed to be fully secure. The complementary input $A_i$ is produced by agent effort via one of two possible technologies, and can be either private or public (in the non-excludable sense) depending on the technology selected. Each agent has access to a unique "status quo" technology, defined by one of the two equations.
\[ A_P = R_{LP}e_P \quad \text{(the poor agent's status quo technology)} \]
\[ A_R = R_{LR}e_R \quad \text{(the rich agent's status quo technology),} \quad (3) \]

where \( R_{LR} > R_{LP} \), so that the rich agent's status quo production technology is superior\(^3\) to that of the poor player. The agents may also choose to employ a "cooperative" technology which is defined by the equation

\[ A_i + A_j = R_H(e_i + e_j), \quad (4) \]

where \( R_H > R_{LR} > R_{LP} \). While the cooperative technology is therefore the most efficient option, its output (measured in units of the complementary input \( A \)) is subject to a pure common pool problem—namely, each agent has the ability to grab all the fruits of cooperation, \( R_H(e_i + e_j) \), for himself. In contrast, each agent is assured to secure the fruits of his own labor when using his status quo technology. As a result, use of the cooperative technology for production of the complementary input \( A \) cannot be sustained in the absence of the appropriate enforcement mechanisms. Formally, we can represent the stage game as a prisoners' dilemma whose only Nash Equilibrium involves mutual defection.

\[^3\] The assumption that the rich agent's status quo technology is superior to that of the poor agent is necessary for the existence of parameter values under which incentive compatibility is satisfied for the poor agent but not for the rich agent. If the two agents were to have access to the same status quo technology (or if the poor agent was the one with access to the superior status quo technology), incentive compatibility for the rich agent would be satisfied whenever incentive compatibility for the poor agent is satisfied.
Table 1
Payoff matrix (payoffs denoted in units of the production input $A$)

<table>
<thead>
<tr>
<th>Poor Agent</th>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperate</td>
<td>$R_H e_R$</td>
<td>$R_H (e_R + e_P)$</td>
</tr>
<tr>
<td>Defect</td>
<td>$R_L R e_R$</td>
<td>$R_{LP} e_P$</td>
</tr>
</tbody>
</table>

As is evident from Table 1, we are making two simplifying assumptions: the first is that mutual defection yields payoffs equivalent to those obtained when each player simply chooses his status quo technology (although all that matters is that mutual defection generates individuals payoffs strictly greater than zero, and strictly less than those from mutual cooperation); the second is that, in the fully cooperative outcome, the ratio of the players' payoffs is equal to the ratio of their contributions to the cooperative effort—namely, when both players cooperate, each player receives an amount of capital-complementing input equal to

$$A_i = R_{II} e_i. \quad (5)$$

This assumption allows us to keep the algebra simple (in particular by allowing us to restrict the issue of free-riding to instances where an agent chooses to defect from the cooperative
norm), but also has two important implications. First, as we can soon verify by checking equation (8), it yields the social surplus-maximizing distribution of the complementary input $A$, as the ratio $A_i/A_j$ is equal to $K_i/K_j$ in equilibrium. Second, it implies that social welfare is independent of distributional inequality, and depends only on whether or not cooperation takes place (i.e., social welfare can take on only one of two possible values, one corresponding to the status quo outcome and one corresponding to the decentralized outcome).

*Stage game payoffs*

In the decentralized equilibrium, each agent chooses $e_i$ to maximize equation (1), given equation (3). The solution to this optimization problem is

$$e_{i, status quo} = K_i \left( \frac{aR_i}{c} \right)^{\frac{1}{1-\alpha}}$$

(6)

As is expected, each agent's optimal choice of effort is increasing in his holdings of private capital and decreasing in the cost of effort. Substituting equation (6) into equation (1), and recalling that each agent has a discount factor of $\beta$, we can write the lifetime utility of each agent under the status quo as

$$S_i = \frac{1}{1-\beta} \left( K_i \left( \frac{aR_i}{c} \right)^{\frac{\alpha}{1-\alpha}} (1-\alpha) \right)$$

(7)

---

4 As can be easily verified, any equation of the form $S = A_1^{\alpha}K_1^{1-\alpha} + A_2^{\alpha}K_2^{1-\alpha}$, where the $A_i$'s and $K_i$'s must each sum to fixed quantities, is maximized when $A_1/A_2 = K_1/K_2$. 

21
Let us now turn to stage game payoffs under full cooperation. Since we have assumed that cooperation simply allows each agent to secure a return to effort equal to $R_H$ instead of $R_L$ (equation (5) vs. equation (3)), each agent's effort choice under full cooperation is equal to

$$e_{i, cooperate} = K e_i \frac{R_H}{c} \left( \frac{1}{c} \right)$$  \hspace{1cm} (8)

and we can express his lifetime utility under full cooperation as

$$C_i = \left( \frac{1}{\frac{1}{c} \left( K e_i \frac{R_H}{c} \left( \frac{1}{c} \right) \right)} \right)$$  \hspace{1cm} (9)

From equations (7) and (9), it is clear that the rich agent does better than the poor agent in both the status quo and cooperative outcomes, and in addition, that the incremental payoff to cooperation relative to the status quo is higher for the rich agent than for the poor agent.

Finally, let us calculate the stage game payoff from defecting when the other agent cooperates. Since a defecting agent $i$ appropriates the fruits of agent $j$'s effort, the latter factors into his own effort choice. Given that agent $j$ cooperates, agent $i$'s optimal choice of effort is equal to

$$e_{i, defect} = K e_i \frac{R_H}{c} \left( \frac{1}{c} \right) e_j$$  \hspace{1cm} (10)

revealing a free-rider effect for the defecting player $i$. Substituting equation (8) for $e_j$ into equation (10) yields
\[ e_{i, \text{defect}} = \text{Max} \left\{ (K_i - K_j) \left( \frac{\alpha R_i^{\alpha}}{c} \right)^{\frac{1}{1-\alpha}}, 0 \right\} \] (11)

We see from equation (11) that, when defecting, the poor agent chooses an effort level of zero, while the rich agent still chooses a strictly positive effort level (equal to his cooperative effort level minus the cooperative effort level of the poor agent). Substituting equation (11) into equation (1), we can calculate the stage game payoffs from defection for the poor agent and the rich agent, respectively, as

\[ D_{\text{poor}} = K \left( \frac{\alpha R_i}{c} \right)^{\frac{\alpha}{1-\alpha}} \rho^\alpha (1 - \rho)^{1-\alpha}, \] (12)

and

\[ D_{\text{rich}} = K \left( \frac{\alpha R_i}{c} \right)^{\frac{\alpha}{1-\alpha}} \left( \rho - \alpha (2 \rho - 1) \right) \] (13)

A quick glance at equations (12) and (13) reveals that both players' static payoffs from defection are a function of \( \rho \), or the extent of inequality in the distribution of private capital. Consider equation (13) first. Depending on the parameter \( \alpha \), \( D_{\text{rich}} \) is either monotonically increasing or monotonically decreasing in inequality \( \rho \); specifically, \( D_{\text{rich}} \) is monotonically increasing in \( \rho \) if and only if

\[ \alpha^{\frac{\alpha}{1-\alpha}} \geq 2\alpha^{\frac{1}{1-\alpha}}, \]
or simply $\alpha \leq 0.5$. In other words, increasing inequality increases (decreases) the rich agent's static payoffs from defection when production relies more (less) heavily on private capital.

Let us now consider equation (12). Intuitively, increasing inequality in the distribution of private capital exert two competing effects on the one-time payoff that the poor agent can secure by defecting ($D_{\text{poor}}$). On one hand, increases in inequality increase the effort level of the rich agent in the cooperative equilibrium, meaning that the poor agent can grab more of the complementary input $A$ by defecting. On the other hand, increases in inequality reduce the private capital holdings of the poor agent, meaning that the poor agent has less use for any given amount of complementary input that he is able to secure for himself. From equation (12), we see that the first effect dominates for increases in inequality $\rho$ up to $\rho = \alpha$, and that the second effect dominates when inequality rises beyond that level.

Therefore, whenever $\alpha \leq 0.5$ (so that production is more private capital intensive), increases in inequality increase the defection payoffs of the rich agent and decrease the defection payoffs of the poor agent (since the defection payoffs for the poor agent are decreasing in inequality whenever $\rho > \alpha$, and $\rho > 0.5$ by definition). On the other hand, whenever $\alpha > 0.5$, the rich agent's defection payoffs are decreasing in inequality, while the relationship between the poor agent's defection payoffs and inequality is inverse U-shaped.

**Social norms**

We are now interested in identifying the conditions under which trigger strategies can sustain cooperation in the infinitely repeated version of the stage game. A social norm of
cooperation will be considered sustainable via trigger strategies if both players choose to cooperate with the understanding that defection at any time \( t \) will trigger a permanent reversion to the Nash equilibrium of the stage game. That is, cooperation is sustainable so long as no agent at any time \( t \) is willing to forsake the incremental returns to cooperation (relative to the status quo) from time \( t + 1 \) onwards in exchange for the one-time payoff from defection at time \( t \). Formally, cooperation is sustainable if and only if

\[
\text{IC}_i: \quad C_i \geq D_i + \beta S_i \tag{14}
\]

for each agent \( i \). The incentive compatibility constraints particular to this model are:

\[
\text{IC}_P : \quad \left( \frac{1}{1-\beta} \right) \left( 1 - \beta \left( \frac{R_{LP}}{R_H} \right)^{\frac{\alpha}{1-\alpha}} \right) \geq \left( \frac{\rho}{1-\rho} \right)^{\frac{\alpha}{1-\alpha}} \left( \frac{1}{1-\alpha} \right) \tag{15}
\]

and

\[
\text{IC}_R : \quad \left( \frac{1}{1-\beta} \right) \left( 1 - \beta \left( \frac{R_{LR}}{R_H} \right)^{\frac{\alpha}{1-\alpha}} \right) \geq \left( \frac{1-\rho}{\rho} \right)^{\frac{\alpha}{1-\alpha}} + 1. \tag{16}
\]

As expected, the incentive compatibility constraints are identical when \( \rho = 0.5 \), and are both more likely to hold when the agents are more forward-looking (\( \beta \) is larger), and when the incremental return to cooperation relative to the status quo is larger (\( R_H \) is larger relative to the \( R_{Li} \)). We also see that increases in inequality \( \rho \) tend to make \( \text{IC}_P \) less likely to hold and \( \text{IC}_R \) more likely to hold. To understand intuitively why it is that increasing inequality tends to make defection more attractive for the poor agent and less attractive for the rich agent, we
can think of the returns to defection as the sum of two components: effort savings and production gains. Effort savings from defection fall with inequality for both the rich agent and the poor agent. Recall from equations (10) and (11) that a defecting rich agent always exerts an amount of effort equal to his cooperative effort choice minus the poor agent's cooperative effort choice, while a defecting poor agent always free rides completely. Because greater inequality reduces the cooperative effort of the poor agent, a defecting rich agent has "more ground to make up," while a defecting poor agent has "less to retract." Thus effort savings are lower for both players when inequality is greater (effect 1). The relationship between inequality and production gains from defection is more complicated. On one hand, increasing inequality means that the poor (rich) agent has less (more) private capital with which to complement whatever amount of the input \( A \) that he is able to appropriate from the opposing player by defecting (effect 2). On the other hand, because the rich agent exerts more effort than the poor player in the cooperative equilibrium, the poor agent has more of \( A \) to appropriate by defecting (effect 3). Finally, because the returns to \( A \) exhibit diminishing returns, the poor (rich) agent, who has less (more) of \( A \) in the cooperative equilibrium, has more (less) to gain from additional stocks of \( A \) (effect 4). For the poor agent, effects 3 and 4 dominate effect 1 and 2; for the rich agent, effects 1, 3, and 4 dominate effect 2.

A quick glance at equations (15) and (16) reveals that there is always a level of inequality \( \rho < 1 \) beyond which it is impossible to maintain incentive compatibility for the poor agent (since the RHS of equation (15) goes to infinity as \( \rho \to 1 \)), whereas there is always a level of inequality \( \rho < 1 \) beyond which incentive compatibility for the rich agent is
certainly satisfied (since the RHS of equation (16) goes to zero as $\rho \to 1$). Finally, note that for intermediate levels of inequality, it is entirely possible for the poor agent's incentive compatibility constraint to hold while that of the rich agent does not hold. Such situations are particularly likely to arise when the status quo technology of the rich agent is sufficiently superior to the status quo technology of the poor agent—that is, when the rich agent has an attractive "exit option."

**Sidepayments**

When one of the two agents' incentive compatibility constraint is violated by the terms of cooperation outlined above, cooperation might nevertheless be sustained with the aid of sidepayments: so long as one agent $i$'s incentive compatibility constraint is satisfied in the absence of sidepayments, the incorporation of a sidepayment from agent $i$ to agent $j$ into the terms of the social norm has the potential to sustain cooperation. Under this arrangement, the generic incentive compatibility constraint (equation (14)) becomes

$$\text{IC}_i: \quad C_i + \frac{\text{transfer amount}}{1 - \beta} \geq D_i + \beta S_i. \quad (17)$$

We must now also take account of the two agents' individual rationality constraints. Let $T$ denote the amount that is transferred from the rich agent to the poor agent (where $T$ can be negative), conditional on cooperation by both players. The individual rationality constraints for the two agents are
The incentive compatibility constraints of the poor agent and the rich agent in the presence of sidepayments are, respectively,

$$IR_P: \quad (1 - \rho)K \left( \frac{R_H}{c} \right)^{\frac{\alpha}{1 - \alpha}} - \left( \frac{\alpha R_{LP}}{c} \right)^{\frac{\alpha}{1 - \alpha}} (1 - \alpha) + T \geq 0$$ (18)

$$IR_R: \quad \rho K \left( \frac{R_H}{c} \right)^{\frac{\alpha}{1 - \alpha}} - \left( \frac{\alpha R_{LR}}{c} \right)^{\frac{\alpha}{1 - \alpha}} (1 - \alpha) - T \geq 0.$$ (19)

The incentive compatibility constraints of the poor agent and the rich agent in the presence of sidepayments are, respectively,

$$IC_P: \quad \left( \frac{1}{1 - \beta} \right) \left( 1 - \beta \left( \frac{R_{LP}}{R_H} \right)^{\frac{\alpha}{1 - \alpha}} \right) - \tau_P T \geq \left( \frac{\rho}{1 - \rho} \right)^{\frac{\alpha}{1 - \alpha}} \left( \frac{1}{1 - \alpha} \right)$$ (20)

$$IC_R: \quad \left( \frac{1}{1 - \beta} \right) \left( 1 - \beta \left( \frac{R_{LR}}{R_H} \right)^{\frac{\alpha}{1 - \alpha}} \right) - \tau_R T \geq \left( \frac{1 - \rho}{\rho} \right)^{\frac{\alpha}{1 - \alpha}} + 1$$ (21)

where

$$\tau_P = \left( \frac{c}{\alpha R_H} \right)^{\frac{\alpha}{1 - \alpha}} \left( \frac{1}{1 - \rho K (1 - \beta)} \right)^{\frac{1}{1 - \alpha}}$$ (22)

$$\tau_R = \left( \frac{c}{\alpha R_H} \right)^{\frac{\alpha}{1 - \alpha}} \left( \frac{1}{\rho K (1 - \beta)} \right)^{\frac{1}{1 - \alpha}}.$$ (23)

Note that while the amount of the sidepayment $T$ enters into the agents' individual rationality constraints directly, it enters into their incentive compatibility constraints with a multiplier $\tau$, and the size of this multiplier is a function of inequality. In particular, as $\rho$ increases, the
"sidepayment multiplier" in the poor agent's incentive compatibility constraint ($\tau_P$) becomes larger, while the sidepayment multiplier in the rich agent's incentive compatibility constraint ($\tau_R$) becomes smaller.

Another way to describe this result is to say that, as inequality increases, the effect of any given transfer amount is magnified with respect to incentive compatibility considerations for the poor agent, and diminished with respect to incentive compatibility constraints for the rich agent. This implies that, if cooperation requires a sidepayment from the rich agent to the poor agent, increases in inequality tend to make cooperation more attainable, all else equal. Intuitively, a very poor agent who chooses to defect in the absence of sidepayments might agree instead to cooperate for only a very small fee, since even a very small fee increases his payoffs from cooperation by a significant margin, in relative terms. At the same time, a very rich agent might be more than happy to offer this fee, since it is quite small in relation to the payoffs he can secure by convincing the poor agent to cooperate. On the other hand, and by similar logic, if cooperation requires a sidepayment from the poor agent to the rich agent, increases in inequality tend to make cooperation increasingly difficult to achieve, all else equal. Intuitively, a rich agent who demands a sidepayment in exchange for his cooperation is likely to ask too much of the poor agent, who is reluctant to sacrifice even a small share of his already meager prospective winnings from cooperation.

*Explicit punishments*
Let us now abandon the possibility of sidepayments, and suppose instead that each agent has the ability help enforce the cooperative norm by setting aside resources to punish his opponent in the event that his opponent defects. In particular, suppose that by setting aside $f_K$ units of capital at the beginning of the stage game, agent $i$ can impose a one-time cost of $xf_K$ on agent $j$ in the event that agent $j$ defects from the cooperative norm, where $x > 0$. Investments in punishment by either agent are unproductive in a strictly economic sense (they are "wasted" on the equilibrium path of cooperation), but may help to skew the incentive compatibility considerations of his opponent in a way that favors cooperation and is mutually beneficial. Note that an agent will only choose to invest in punishments if he is otherwise (under the standard norm of cooperation) willing to cooperate while his opponent is not. Therefore, either only the rich agent invests in punishment, or only the poor agent, or neither.

First consider the case where, in the absence of sidepayments, the rich agent is willing to cooperate and the poor agent is not. In this case, the individual rationality constraints of the two agents are

\[
\text{IR}_R : \quad R_H \geq R_{L,P} \quad \quad (24)
\]

\[
\text{IR}_P : \quad \rho \left( \left( \frac{R_H}{c} \right)^{\frac{\alpha}{1-\alpha}} - \left( \frac{R_{L,R}}{c} \right)^{\frac{\alpha}{1-\alpha}} \right) \geq f_s \left( \frac{R_H}{c} \right)^{\frac{\alpha}{1-\alpha}}. \quad (25)
\]

Note that the individual rationality constraint of the poor agent is automatically satisfied (and unchanged from the baseline model), since punishments are never delivered on the equilibrium
path of cooperation. The incentive compatibility constraints of the poor and rich agents are, respectively,

\[
\text{IC}_P : \left( \frac{1}{1 - \beta} \right) \left( 1 - \beta \left( \frac{R_{LP}}{R_H} \right)^{\frac{\alpha}{1 - \alpha}} \right) \geq \left( \frac{\rho}{1 - \rho} \right)^{\frac{\alpha}{1 - \alpha}} \left( \frac{1}{1 - \alpha} \right) - \phi_{fR}
\]

\[
\text{IC}_R : \left( \frac{1}{1 - \beta} \right) \left( 1 - \beta \left( \frac{R_{LR}}{R_H} \right)^{\frac{\alpha}{1 - \alpha}} - \phi_{fR} \right) \geq \left( \frac{1 - \rho}{\rho} \right) \left( \frac{\alpha}{1 - \alpha} \right) + 1,
\]

where

\[
\phi_P = \frac{1}{(1 - \rho)} \left( x \left( \frac{c}{R_H} \right)^{\frac{\alpha}{1 - \alpha}} \right) \left( \frac{\alpha^{\frac{\alpha}{1 - \alpha}}}{\alpha^{\frac{\alpha}{1 - \alpha}} - \alpha^{\frac{1}{1 - \alpha}}} \right)
\]

\[
\phi_P = \frac{1}{\rho}.
\]

We see from equation (25) that investments in punishment by the rich agent become more feasible with respect to individual rationality considerations when inequality increases (since increasing inequality increases the incremental payoff to cooperation relative to the status quo for the rich agent, the rich agent has more to potentially divert towards investments in punishing the poor agent). In addition, we see from equations (26) – (29) that when inequality is more pronounced, the incentive effects of any given investment in punishment \( f_R \) by the rich agent are diminished for the rich agent and magnified for the poor agent. This
implies that, all else equal, threats of punishment levied by the rich agent against the poor agent are more likely to enforce cooperation by the poor agent when inequality is more pronounced.

Let us now consider the opposite case, in which cooperation requires the poor agent to make investments in punishment. In this case, the individual rationality constraints are:

\[
\text{IR}_p: \quad a_p R_H - a_p R_{LP} \geq f_p R_H - f_p R_{LP}
\]

\[
\text{IR}_r: \quad R_H \geq R_{LR},
\]

while the incentive compatibility constraints are

\[
\text{IC}_p: \quad 1 - a_p R_H - a_p R_{LP} \geq f_p R_H - f_p R_{LP},
\]

\[
\text{IC}_r: \quad 1 - a_p R_H - a_p R_{LR} \geq f_p R_H - f_p R_{LR} + 1
\]

where

\[
\Box_p = \frac{1}{1 + f_p}
\]

\[
\Box_r = \frac{1}{1 + b_p}
\]
The situation here is in several important respects the opposite of the one just considered, in which the threat of punishment comes from the rich and is directed against the poor. From equation (30), investments in punishment by the poor agent become less feasible with respect to individual rationality considerations when inequality increases (since increasing inequality reduces the incremental payoff to cooperation relative to the status quo for the poor agent, the poor agent has less to potentially divert towards investments in punishing the rich agent).

In addition, we see from equations (32) – (35) that when inequality is more pronounced, the incentive effects of any given investment in punishment $f_{P}$ by the poor agent are magnified for the poor agent and diminished for the rich agent. This implies that, all else equal, threats of punishment levied by the poor agent against the rich agent are less likely to enforce cooperation by the rich agent when inequality is more pronounced.

We should note that for the sake of keeping the model in this section tractable, we have sacrificed a great deal of plausibility. For example, the two-person (or group) set up is quite restrictive. For more general distributions (even just adding a third group called the “middle class”) ‘increasing inequality’ can take many forms, even if one restricts attention to mean-preserving spreads. In Bardhan and Singh (2003) we have generalized to the case of many persons. Similarly restrictive is the presumption that social welfare is independent of inequality and depends only on whether cooperation takes place or not (of course, individual rationality conditions ensure that cooperation, if chosen, will be Pareto superior to non-cooperation, so that any concave aggregation of individual utilities will favor cooperation). And, of course, a problem with the trigger strategies is that once one embarks on the
punishment phase there is no looking back. So if there are small shocks in the model, in the long run every game will collapse to the punishment path. There is some existing literature which we can follow here, allowing for the possibility that some defection is tolerated because it could have been caused by an exogenous shock.

V

In this chapter we have discussed common pool problems in fiscal management when there are heterogeneous groups, and then related problems at the micro level for collective action when there is economic inequality. In the common pool model of fiscal policy the benefits of public spending on any one interest group are concentrated in that group, but the costs are uniformly distributed across all groups (for example, in the form of taxes, or interest on public debt)—the latter assumption can be interpreted to mean that all groups draw from a "common pool" of public resources. Each group therefore secures a large share of the total benefits from its particular brand of public spending while bearing a comparatively small share of the total costs. This sets the stage for a fiscal policy version of the Tragedy of the Commons: each interest group pursues its fiscal agenda until the private benefits of doing so no longer exceed the private costs, but by this time the private benefits have already fallen well below the social costs.

We began with a brief review of recent theoretical articles that all operate within this common pool framework. These articles are primarily interested in examining the
feasibility of cooperative agreements between interest groups to overcome the fiscal common pool problem. In the models, every interest group is better off when all groups adhere to a cooperative agreement than when none do; yet, when every other group cooperates, each group can secure a windfall gain by breaking its promise to adhere to the agreement—a classic prisoners' dilemma. The models differ sharply when it comes to predicting the conditions under which cooperation is sustainable. In some models, cooperation is only sustainable after fiscal conditions have become severely damaged, while in others, cooperation is only sustainable when fiscal conditions are still favorable.

The models we consider from the existing literature in section II all assume that the relevant interest groups are economically identical, so that incentives for cooperation are symmetric across groups: either all groups are willing to cooperate or none are. This assumption partly explains the existing literature's emphasis on aggregate fiscal conditions and their role in making cooperation more or less sustainable. After reviewing these models, we present a simple model in section IV that considers an alternative setting in which economic differences create differential incentives for cooperation. In our model, distributional inequality across interest groups creates unequal returns to cooperation. All else equal, the richly endowed stand to gain more from cooperation than the poorly endowed. At extreme levels of economic inequality, the rich are always eager to cooperate while the poor refuse to do so. However, at intermediate levels of inequality, we demonstrate that it may often be the poor who are willing to cooperate and the rich who are not. Our results in this vein arise from the very plausible assumption that the rich have access to more favorable exit options relative to those available to the poor.
We then enrich the basic model by considering two mechanisms that might be introduced to salvage cooperation when it is otherwise unsustainable: sidepayments from agents who are willing to cooperate to agents who are otherwise unwilling to cooperate, and relatedly, explicit punishments levied by the former against the latter (note that these mechanisms are only sensible in a setting of differential incentives for cooperation). We show that each of these two mechanisms may be quite useful for sustaining cooperation when it is the rich who must induce the poor to cooperate, but that they are likely to be much less effective when it is the poor who must attempt to enforce discipline on the part of the rich.
References


