DISCUSSION PAPER SERIES

No. 4210

VIABLE TAXATION

Carlo Perroni and Kimberley Ann Scharf

PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4210.asp**

ISSN 0265-8003

VIABLE TAXATION

Carlo Perroni, University of Warwick
Kimberley Ann Scharf, University of Warwick and CEPR

Discussion Paper No. 4210
January 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4210

January 2004

ABSTRACT

Viable Taxation*

Taxation is only sustainable if the general public complies with it. This observation is uncontroversial with tax practitioners but has been ignored by the public finance tradition, which has interpreted tax constitutions as binding contracts by which the power to tax is irretrievably conferred by individuals to government, which can then levy any tax it chooses. In the absence of an outside party enforcing contracts between members of a group, however, no arrangement within groups can be considered to be a binding contract, and therefore the power to tax must be sanctioned by individuals on an ongoing basis. In this Paper we offer, for the first time, a theoretical analysis of this fundamental compliance problem associated with taxation, obtaining predictions that in some cases point to a re-interpretation of the theoretical constructions of the public finance tradition while in others call them into question.

JEL Classification: H10, H20, H30 and H40
Keywords: government, public goods and taxation

Carlo Perroni
Department of Economics
University of Warwick
COVENTRY
CV4 7AL
Tel: (44 24) 7652 8416
Fax: (44 24) 7652 3032
Email: c.perroni@warwick.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=154747

Kimberley Ann Scharf
Department of Economics
University of Warwick
COVENTRY
CV4 7AL
Tel: (44 24) 7652 3742
Fax: (44 24) 7652 3032
Email: k.scharf@warwick.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=127385

*We are grateful to Tim Besley, Ian Crawford, Ben Lockwood, Hamish Low, David Ulph, and other participants to the 2003 Warwick/IFS Public Economics Weekend for comments and suggestions. All errors are our own.

Submitted 29 September 2003

1 Introduction

The economics literature has traditionally characterized taxation as a coercive levy. The power to tax, in turn—in line with a contractarian view of government institutions which dates back to Hobbes and Locke—can be thought of as arising from a tax constitution to which individuals voluntarily and irreversibly submit in order solve the free-rider problem in public goods provision. A tax constitution is then the result of a once-and-for all agreement to confer coercive powers to a monopoly government, a ruling body which pursues goals that are private but nevertheless consistent with the pursuit of collective goals—e.g. surplus extraction, which implies a stake in total surplus—and to which enforcement power is irretrievably surrendered in some way—e.g. by providing the ruler with a strong and unflinchingly loyal army.[1]

Power, however, can never be irretrievably surrendered by citizens to government: formal constitutions can be (and are) changed; and even dictatorships depend on the ongoing support of a (possibly restricted) group of individuals for their survival. If the power of enforcement does not ultimately reside with government but remains within the group that supports government, the notion that the free-rider problem can be solved by a tax constitution seems simply to shift the problem to a different level: how can coercion be sustained in groups?

Since governments do not possess any independent power of enforcement beyond that of individuals, and since there are no independent powers of enforcement *outside* the tax constitution, we cannot view the latter as a binding contract in the usual way that we think of contracts between individuals. Rather, enforcement must come from *within* the tax constitution itself. Then, abiding to the tax constitution must be a continuously-renewed, individually and collectively rational choice for individuals and for the group, balancing individuals' temptations for opportunistic behaviour against the group's ability to punish defectors. In other words, a tax constitution must be self-enforcing. This idea is consistent with the point, often made informally, that formal norms are irrelevant if they cannot be enforced, and that policy making must take enforcement problems into account in order to produce relevant policies.[2]

The idea that cooperation must be self-supporting has gained some currency in the economics literature (e.g. collusion between firms and international agreements) but has never been systematically applied to the interpretation of tax institutions.[3]

[1] The contractarian interpretation of the state is generally attributed to Hobbes (1651) and Locke (1690), and was later popularized by Rousseau (1762); it was first applied to the free-rider problem by Wicksell (1834), and was later popularized by several influential writers (Olson, 1965; Buchanan, 1975; Brennan and Buchanan, 1980).

[2] See, for example, Bird (2003).

[3] A sparse literature, begun by McMillan (1979), has focused on the sustainability of voluntary contributions towards collective consumption under repeated interaction, but coercive taxation as such has not

In this paper we develop a theory of self-enforcing tax constitutions under repeated interaction, where constrained-efficient tax rules must be supported by renegotiation-proof punishment strategies that are both individually and collectively rational. We then ask to what extent the idea of self-enforcing taxation is able to rationalize the observed structure of tax institutions, and discuss how it can be reconciled with the presence of political processes presiding over collective choices.

Our analysis shows that accounting for the ongoing nature of incentives for opportunistic behaviour does much more than simply refine old contractarian ideas by giving them a theoretically more satisfying treatment. Sustainability involves stricter requirements on tax structures than contractarianism does—the basic contractarian requirements being that individuals must be better off under a tax constitution than under a "state of anarchy." As a result, the analysis of self-enforcing tax constitutions offers sharper predictions about the structure of taxes than contractarianism does. For example, a theory of self-enforcing tax constitutions predicts that individuals with the same preferences and income will pay the same tax (horizontal equity), and that higher income individuals will pay higher taxes (ability-to-pay principle), even if individuals are neutral towards redistribution.

A model of self-enforcing taxation can thus explain the most commonly observed features of tax systems, as traditional models of coercive taxation can, but it can derive these features from basic principles of individual and collective rationality—on which sustainability under repeated interaction is based—rather than from a set of distributional or ethical principles as traditional theories do; and in some cases it goes further in offering explanations for certain features that traditional theories are silent on or have difficulty explaining. For example, a model of self-enforcing taxation predicts that there exists a natural upper bound to the level of taxation that can be sustained—something that policymakers seem acutely aware of—an upper bound which lies strictly below income. It also predicts that there exists a close relationship between the establishment of private property rights and the public provision of collective goods—a relationship which is well documented throughout history but about which theories of coercive taxation have nothing to say; indeed, it is only when insecure property rights are accounted for that a convincing theory of self-enforcing tax equilibria takes shape.

The axiomatic parsimoniousness of such a model, combined with its ability to account

---

been examined in this light. There is a small but growing literature on the problem of enforcement in groups (Muthoo, 2003; Dal Bo, 2001; Acemoglu, 2002). While this literature does not focus on taxation, some of the ideas in that literature are related to the questions addressed here. A recent literature dubbed "Historical and Comparative Institutional Analysis" (Greif, 1998) uses ideas from theory of repeated games to analyze and rationalize differences in institutional structures across countries and time periods. This literature mostly focuses on cooperation in markets, with only a select few contributions addressing repeated interaction between state and citizens (Weingast, 1997); even those, however, start from the traditional view that the state does have some autonomous, albeit possibly limited, power over citizens.

for the sustainability problem and produce a richer set of predictions, makes it a compelling alternative to traditional models of coercive taxation. At the same time, a theory of viable taxation does not invalidate insights obtained from traditional analyses; rather it complements them: in mechanical terms, accounting for the problem of viability amounts to incorporating additional constraints—derived from a model of repeated strategic interaction—into the standard optimization problems that are used to characterize optimal taxes.

Such an analytical framework may help us better understand observed tax systems, but it also has a cost when compared to traditional approaches. Theoretical results for repeated-interaction environments (folk theorems) are fundamentally indeterminacy results; this means that, in order to base predictions on these constructs, one must apply refinements and heuristic argument, forgoing the predictive directness of the more standard optimization models which public finance researchers have become accustomed to. Moreover, the idea that a tax constitution is a self-enforcing contract calls into question the interpretation of politics as a system of procedures that substitute for bargaining in large groups—an interpretation that is common to much of the recent political economy literature. Instead, the idea of tax constitutions as self-enforcing contracts points back to welfare theory as a predictive rather than normative tool of tax analysis, and suggests viewing politics as an information-pooling device rather than as a mechanism for reconciling conflicting interests.

The paper is structured as follows: Section 2 examines constrained-efficient, equilibrium tax constitutions under repeated noncooperative interaction amongst a group of infinitely-lived individuals, supported by renegotiation-proof punishment strategies in which individuals can only be punished through a reduction in collective consumption; Section 3 examines the relationship between collective consumption and property rights, in situations where individuals can be also punished by income expropriation but where property rights must themselves be self-enforcing; Section 4 describes predictions that are specific to a model of self-enforcing taxes. Section 5 examines other aspects of taxation in the context of a self-enforcing equilibrium framework, namely the incentive effects of taxes, the role of distributional goals, tax evasion and monitoring; finally, Section 6 deals with how self-enforcing cooperation can be interpreted as being the result of a social contract, and how it can be reconciled with politics. The main body of the paper offers a mostly verbal discussion of the theoretical arguments, with an aim to convey a broad picture of how these various arguments combine to produce a theory of viable taxation; formalities are relegated, for the most part, to the appendix.

2 Renegotiation-proof Contribution Equilibria with No Income Expropriation

Taxation has traditionally been interpreted as a means of overcoming, through coercion, the free riding problem in public good provision. This problem can be formally described

as follows. Consider an economy with n individuals, engaging in private and collective consumption. Individual i receives an income y^i.[4] Individuals can contribute to collective consumption an amount c^i, resulting in a level of collective consumption equal to $g = \sum_i c^i$, and in a level of private consumption $x^i = y^i - c^i$. The payoff of an individual in each period is $u^i(x^i, g) = x^i + \theta^i v(g)$, where $v'(g) > 0$ and $v''(g) < 0$ (quasilinear preferences) and $\theta^i > 0$. In a single round of interaction (as examined by, among others, Bergstrom, Blume, and Varian, 1980), a noncooperative equilibrium in contributions will result in a level of collective consumption, $g^N = \sum_i c^{iN}$, that falls short of the efficient level of collective consumption, g^*. Furthermore, if n is sufficiently large, noncooperative voluntary contributions will be zero.[5]

Coercive taxes can thus be interpreted as an institutional device for supporting collective consumption in large economies where individuals are selfishly-motivated and behave in a rational manner. But, as we have argued in the introduction, unless the power to tax can be irrevocably transferred to an absolute ruler, tax compliance must itself be interpreted as being the result of a deliberate, continuously renewed choice on the part of individuals. Such voluntary compliance can be characterized as an equilibrium phenomenon in a game where strategic interaction is repeated, resulting from a balance between individuals' temptation not to comply and the punishment that the rest of the group could administer against noncompliers.

Once we embrace the idea that tax compliance is voluntary, however, there seems to be little left to distinguish between taxes and voluntary contributions. And indeed the idea that the free riding problem in voluntary contributions to collective consumption can be solved by repeated interaction has been suggested before in the literature (McMillan, 1979). These contributions, like much of the applied literature on repeated interaction, focus on "Nash-reversion" punishment strategies, whereby a reduction in contributions below a given trigger level results in indefinite reversion to the one-shot noncooperative outcome. They conclude that if individuals are infinitely-lived and the game is repeated indefinitely, levels of contributions above the one-shot noncooperative levels can be supported by the threat of credible punishment.

There are two fundamental problems with this construction, one empirical and the other theoretical in nature. Whenever we observe institutions that support cooperation, we never observe indefinite Nash reversion being applied (we do not observe the public sector shutting down if someone does not pay taxes), and this seems also to be the case for voluntary contributions to collective consumption. The second objection might seem, at first sight, to be unrelated to the first one. Nash-reversion punishment strategies are analytically attractive, but there seems to be no compelling reason to view them as natural candidates

[4]For now, we shall assume that income cannot be forcibly expropriated; this assumption will be relaxed later on in our discussion.

[5]A formal characterization of a noncooperative equilibrium is given in Appendix A.

for an equilibrium under repeated interaction—and indeed there are good reasons for not doing so.

To begin with, if contributions are not perfectly observable, punishment by Nash reversion could be triggered "by mistake" in an equilibrium where no actual deviations occur, rather than remaining an idle, out-of-equilibrium threat (albeit a credible one). As was first shown by Green and Porter (1984), when players cannot perfectly monitor other players' actions, efficient punishment strategies must concentrate punishment immediately after defections, reverting to cooperative play afterwards—so as to minimize the expected cost of mistakes while maximizing deterrence; and even when monitoring is perfect, Abreu (1988) has shown that concentrating punishment in the periods that immediately follow a defection can outperform Nash reversion.

Additionally—and independently of other considerations—limited-duration punishment can make equilibria robust to the possibility of renegotiation. Nash reversion implies threats that are individually but not collectively credible: if a high-contribution outcome can be sustained by the threat of Nash reversion, then, upon entering the punishment phase, all parties would have an incentive to renegotiate away from the prescribed punishment—forgoing punishment and re-coordinating to a high-contribution equilibrium in the continuation game; this makes the threat of Nash reversion implausible. To deal with this type of objection, a number of writers (Farrell and Maskin (1989); Abreu, Pearce, and Stacchetti (1993); Van Damme (1989)) have proposed *renegotiation proofness* as a further refinement. This imposes a collective rationality requirement for *plausible* (rather than just credible) punishment strategies: upon entering the punishment phase, it must not be possible to find alternative equilibrium strategies in the continuation game that are favoured by all players to the stated course of action; i.e. at least one of the parties involved must have an active interest in carrying out the punishment. As Van Damme (1989) has shown with respect to a repeated prisoners' dilemma, renegotiation proofness may require concentrating punishment immediately after defections—as in Green and Porter (1984) and Abreu (1988), albeit for different reasons.

It turns out that, with imperfectly observable contributions to collective consumption, efficient, renegotiation-proof punishment strategies should not rely on reductions in the level of collective consumption in order to induce compliance—not even temporarily. Thus, invoking renegotiation proofness in the presence of imperfect monitoring can also take care of the first, empirical objection, producing predictions that parallel more closely the way in which we observe punishment being actually administered. At the same time, the structure of the resulting equilibria—as detailed below—makes it also more natural to interpret them as *tax* equilibria.

A formal definition of renegotiation proofness as we apply it here (*restricted strong renegotiation proofness*) is given in the appendix; loosely speaking, a renegotiation-proof equilibrium is an equilibrium that is not Pareto dominated by another equilibrium and is supported by a continuation equilibrium which is also Pareto-undominated. Renegotiation

proofness thus requires that the punishment that is triggered by defections must be rational for nondefectors, so that players could not renegotiate to an alternative, Pareto-dominant equilibrium in the continuation game.

Renegotiation proofness embeds the idea of efficient selection not just of continuation equilibria but also of the candidate equilibrium. This, however, does not necessarily mean that the level of contributions in a renegotiation-proof equilibrium must be the maximum level that can be sustained with heterogeneous individuals, there may exist non-Pareto ranked renegotiation proof equilibria involving different levels of collective consumption. In order to identify equilibria that support the maximum sustainable level of collective consumption, we need to invoke a further equilibrium selection criterion, namely selection according to ex-ante expected utility maximization—from the point of view of a representative individual who has yet to observe her own characteristics (*veil-of-ignorance* selection). Then, quasilinear preferences imply that ex-ante utility uniquely depends on the level of collective consumption, independently of the ex-post distribution of contributions. Thus, equilibrium selection on the basis of the maximum sustainable level of collective consumption—as long as this falls short of g^*—amounts to selection of an ex-ante undominated equilibrium. Quasilinear preferences also imply that expected utility maximization (or equivalently selection according to a utilitarian principle) involves no ex-ante risk aversion (i.e. no ex-post inequality aversion).

Consider then a scenario where an individuals' incomes are perfectly observable but their contributions are not. Specifically, if individual i chooses to contribute c^i, with some positive probability her actual observed contribution will be less than c^i (and it will be c^i otherwise). In a large economy, the following punishment strategies can be shown to be ex-ante undominated renegotiation-proof equilibrium strategies (see the proof of Proposition 1 in Appendix A): each player, i, contributes an amount c^i; if the observed contribution of any player, i, in period t is less than c^i, then in period $t+1$ all the other players (the punishers) collectively reduce their contributions to $g - y^i$ and expect player i (the defector) to contribute her full income, y^i, in period $t+1$. If the defector is observed to comply ("repents"), the other players increase their contributions back to the initial levels ("forgive") from $t+2$ onwards; otherwise forgiveness is postponed further. Renegotiation proofness prevents the level of collective consumption to be lowered to any arbitrary low level during punishment, since moving to a such a continuation equilibrium would not be collectively rational. Moreover, with imperfect monitoring, any reduction in g during punishment would only impact negatively on ex-ante utility, given that punishment is triggered with positive probability; on the other hand, punishing perceived defections by requiring the defector to temporarily increase her contribution does not lower ex-ante utility. Therefore, in an ex-ante efficient equilibrium, the level of collective consumption will remain unchanged during punishment.

Proposition 1 *In a large economy where income cannot be forcibly expropriated and in-*

dividual actions are not perfectly observable, constrained-efficient renegotiation-proof equilibrium strategies punish a reduction in contributions below their equilibrium-play level by requiring that such a reduction be followed by a temporary increase above the equilibrium-play level.

The structure of the above renegotiation-proof punishment strategies lends itself to a natural interpretation: the increase in the contribution required from defectors in the period following defection is consistent with the idea of a *fine*. Fines cannot be applied coercively here (the case where stronger forms of collective coercion can be applied will be discussed later), but are only credible as a way of punishing deviators because of the reduction in collective consumption that results from postponing payment of the fine, and must be temporary, because the promise of a reversion to the normal state of cooperation is the only way of inducing a deviator to comply with the punishment. Unlike in the case of Nash-reversion punishment strategies, punishing defectors in this way is collectively rational, and hence plausible, as it does not entail a break-up of cooperation.[6]

Suppose that—for the time being—we embrace an interpretation of the equilibrium we have just described as a *tax* equilibrium. Does this interpretation produce predictions that are in line with the observed structure of taxes?

The first implication concerns the maximum sustainable level of contributions (taxes): a contribution $c^i = y^i$ cannot be supported, i.e. the ratio of contributions to income (the average tax rate) is bound to be strictly below unity, both at the individual level and in aggregate, and any increase in disposable income raises the sustainable level of contributions. This is simply because a contribution $c^i = y^i$ would leave no room for a fine to be payable in the punishment phase. Note that this is the case even though the structure of preferences (quasilinear) assumed here does not imply that the ratio of collective to private consumption should remain below unity.

Proposition 2 *The maximum contribution level by an individual that can be sustained in a constrained-efficient, renegotiation-proof equilibrium in a large economy is always less than that individual's income and is increasing with it.*

The application of Nash-reversion punishment strategies produces very different (and implausible) predictions in this respect: with quasilinear preferences, the maximum contribution level, c^i, that can be sustained by the threat of indefinite reversion to g^N can exhaust income, y^i.

[6] In the discussion that follows, for the sake of brevity we shall omit explicit references to imperfect monitoring. Invoking imperfect monitoring is analogous to applying an equilibrium refinement, whose only effect is to rule out continuation equilibria involving a drop in the level of collective consumption (as shown in the proof of Proposition 1.)

If we next look at the relationship between the maximum sustainable level of contributions and group size—focusing for simplicity on the case of a homogenous economy—we find that, for a given per capita income level, the overall level of collective consumption that can be sustained in a constrained efficient equilibrium is increasing in the number of individuals involved. However, unlike Nash-reversion punishment equilibria, where the supportable level of collective consumption is independent of numbers, renegotiation-proof punishment is limited by the requirement of collective rationality, implying that the strength of the punishment that can be plausibly administered decreases with the number of individuals contributing to collective consumption. Thus free riding cannot be fully eliminated. Nevertheless, as group size approaches infinity—and the free-riding problem becomes increasingly acute—the equilibrium level of collective consumption approaches a level which—provided the individual marginal valuation for collective consumption is sufficiently large for g approaching zero (see the proof of Proposition 3)—remains strictly above the one-shot noncooperative level of collective consumption. Thus, renegotiation proofness notwithstanding, repeated interaction makes it possible to improve on the single-shot outcome, even in a large economy.

Proposition 3 *In a large, homogenous economy, for a given level of per-capita income, the maximum contribution level that can be sustained in a constrained-efficient, renegotiation-proof equilibrium is increasing in n. For n large, this level (weakly) exceeds the corresponding equilibrium level of contributions under a single round of interaction.*

What does the model predict about the distribution of the burden of collective provision?

As noted earlier, with quasilinear preferences, the efficient level of collective consumption is independent of the distribution of contributions, and inequality aversion plays no role. Thus, enforcement constraints are the only determinants of the distribution of contributions in equilibrium. Despite this, when we look at the contributions required from individuals with the same income in a constrained-efficient equilibrium with $g < g^*$, the model predicts that individuals with the same income and same preferences will pay the same contribution: since the incentive to postpone repentance in the punishment phase—which determine the sustainable level of contribution—is increasing in c^i, and since these conditions are otherwise the same for individuals who share the same income and preferences, the structure of contributions that can best ensure that these inequalities be satisfied for all players involves equal contributions for identical players. This feature is consistent with the principle of *horizontal equity* in the taxation of identical individuals. Here, however, it is just a consequence of the need for supporting a constrained-efficient level of collective consumption, and is fully unrelated to distributional considerations: there is no reason for requiring contributions to be the same across individuals other than the need for supporting the highest possible level of collective consumption.

If we next look at what the model predicts with respect to how contributions should vary with income. Focusing on a large-economy scenario where individuals have different incomes but identical preferences ($\theta^i = \theta, \forall i$), we find that constrained-efficient, individual contribution levels are increasing in the level of income. This results from the fact that the maximum punishment that can be administered against a defector (the fine) depends positively on her income. In the interpretation of contribution levels as taxes, this feature is consistent with the *ability-to-pay principle* in the determination of taxes on individuals of different income levels. Once more, this is unrelated to distributional concerns or income effects (given that preferences are quasilinear in the model). Neither does the positive relationship between contributions and income levels result from the fact that low income individuals may not have enough income to channel into contributions: constrained-efficient contributions will be higher for high-income individuals even when the contribution levels of all income types fall short of the income level of the poorest individual.

Proposition 4 *In a large economy where there is no preference heterogeneity in the population, a constrained-efficient, renegotiation-proof equilibrium features equal contributions by individuals of the same income level, and higher income levels are associated with higher equilibrium contributions, even when the value of collective consumption to an individual is independent of income and contributions do not exhaust income.*

Contrasting constrained-efficient, renegotiation-proof equilibria with constrained-efficient, Nash-reversion equilibria, we can note that no positive relationship between contribution levels and income levels is predicted by the latter, absent income effects.

It can also be shown that constrained-efficient contribution levels increase in income at an increasing rate. The degree of rate progressivity is again unrelated to distributional concerns. Instead, it is directly related to the structure of preferences towards collective consumption—which in an economy with no preference heterogeneity is the same for all individuals—i.e. to the shape of the valuation function, v, and to the degree of impatience, which is inversely related to the discount factor δ. Then, other things being equal, contributions will be comparatively less progressive in scenarios where the valuation of the public good provided with the revenue is inelastic (for example, because of the availability of private substitutes to collective consumption) and where individuals are comparatively more farsighted.

Proposition 5 *In a large economy where there is no preference heterogeneity in the population, an individual's contribution level in a constrained-efficient, renegotiation-proof equilibrium is increasing more than proportionally to her income level. Equilibrium contributions will be comparatively more progressive the more elastic is the marginal valuation of collective consumption with respect to changes in the level of its provision and the higher the degree of impatience.*

To investigate the implications of preference heterogeneity for the distribution of the burden of collective provision, we can focus on an alternative scenario where individuals have the same income level but different valuations, θ^is. Assume first that preferences are publicly observable. Focusing on a constrained-efficient equilibrium with $g < g^*$, we find that contributions in a constrained-efficient equilibrium are positively related to individual valuations for collective consumption; this is because, for a given fine, a defector's temptation not to pay the fine is lower the higher is her valuation for collective consumption. This prediction is consistent with the application of the *benefit principle* in the allocation of the cost of provision amongst individuals (Lindahl-style taxation). However, it can be also shown that the equilibrium contribution level is less elastic with respect to changes in an individual's valuation for collective consumption than it is with respect to changes in her income level. Therefore, even when preferences are fully observable, income differentials have comparatively more effect than valuation differentials in producing a spread in contribution levels, i.e. ability to pay dominates benefit considerations in determining an individual's contribution level in a constrained-efficient equilibrium.

Anonymity with respect to preferences does not rule out the use of the punishment strategies we have described. Preferences-dependent equilibrium contributions can also be supported in a scenario where preference characteristics are not publicly observable[7]—so that individual trigger contribution levels cannot be conditioned directly on preferences—as long as the overall distribution of preference types is fully known. The reason for this is that punishment involves a reduction in contributions by nondefectors, and therefore need not be specifically targeted towards a certain individual. Then, rather than describing the punishment as being triggered by a reduction in an individual's contribution, we can describe it as being triggered by a reduction in the number of contributions of each type below the corresponding equilibrium value. Repentance will then follow on the part of any defector in an equilibrium in which, off the path of equilibrium play, nondefectors stick to the punishment strategies that have been described (i.e. they refrain from increasing contributions themselves in order to induce repentance, knowing that the defector will repent), a behaviour which is in turn rationally anticipated by the defector. Thus an equilibrium featuring differentiated contributions levels by preference types is fully consistent with anonymity with respect to preference characteristics.

Proposition 6 *In a large economy where there is preference heterogeneity in the population, an individual's contribution level in a constrained-efficient, renegotiation-proof equilibrium is increasing with her valuation of collective consumption; however, individual equi-*

[7]This would be the case, for example, if individual preference characteristics, as summarized by the preference parameters, θ^is, are nonverifiable; then, even if a player's past actions could conceivably be used by others to infer her characteristics, punishment strategies must remain anonymous with respect to the parameter θ^i. An alternative but equivalent scenario would be one where the parameters θ^is are time dependent and serially uncorrelated.

librium contribution levels are comparatively more sensitive to changes in income than they are to changes in individuals' valuation for collective consumption.

To sum up our discussion so far, if we interpret individual contributions as taxes, in a self-enforcing renegotiation-proof equilibrium where punishment of defectors only consists of a temporary withdrawal of contributions by others, we obtain predictions that are in line with those from models of coercive taxation—such as the prediction that taxes increase with income—as well as additional predictions that can help explain certain features of real-world tax systems which traditional analysis are silent upon—such as the prediction that there exist a natural upper bound for the sustainable ratio of collective consumption to income.

Even when predictions parallel those of the traditional analyses, they are derived from different principles. For example, in a constrained-efficient self-enforcing equilibrium, tax progression is not related to inequality aversion but depends on preferences towards collective consumption. Similarly, the positive relationship between the level of tax and the benefit received is obtained as a direct result of enforcement constraints rather than having to be derived from the application of ethical principles. Thus, in comparison with traditional theories of coercive taxation, a model of viable taxation makes it possible to say more with less.

Yet, the model we have analyzed so far still has two serious shortcomings as a model of taxation. First, if the only recourse for the group against defectors is a reduction in contributions, only limited punishment can be administered in a large economy. Although we have shown that increasing the number of individuals, for a given level of per capita income, does not result in less contributions, it does necessarily reduce the volume of contributions *as a proportion of income*, as the following theoretical experiment shows. Consider a homogeneous economy with $y^i = y, \theta^i = \theta, \forall i$, and suppose that, starting at a given constrained-efficient equilibrium, we increase the number of individuals and simultaneously reduce the per-capita income level and the individual valuation parameter θ so as to hold both aggregate income and the aggregate valuation for collective consumption constant. In the modified economy, aggregate income and the efficient level of collective consumption will both be the same as in the original economy, and so will the individual contribution rate (as a fraction of income) required to support a given level of collective consumption. Nevertheless, it can be shown that punishment will be comparatively less effective in the larger economy, implying that the maximum sustainable level of collective consumption will be lower.[8] Thus, although repeated interaction can make it possible to improve on the one-shot noncooperative outcome, the free-riding problem remains.

[8]Let n increase to $n' > n$, and simultaneously change the per-capita income level from y to $y' = (n/n')y$ and the public good valuation scale parameter from θ to $\theta' = (n/n')\theta$. In the resulting modified economy, aggregate income is the same as in the original economy, and the aggregate valuation for collective consumption also remains the same. The efficient level of collective consumption is constant. The per-

The second shortcoming of this model as a model of taxation is that it runs against the observation that, in addition to fines, stronger forms of coercion *are* available and *are* actually employed to secure compliance. Yet, in order to rationalize the use of stronger forms of punishment in the context of a self-enforcing equilibrium, it must be the case that their application is ex-post collectively rational and hence ex-ante plausible. For example, supporting cooperation by punishing a reduction in contributions with death is not plausible if administering the penalty is costly, simply because applying the penalty cannot be part of an ex-post collectively rational response for the group once the defector adopts her one-shot optimal level of contribution (given that, by construction, renegotiation to a Pareto-dominant, cooperative equilibrium would then be possible.[9])

In the next section we show that stronger punishment can be rationally administered in an economy where income expropriation is feasible.

3 Renegotiation-proof Tax Equilibria with Income Expropriation

One channel through which punishment can be plausibly administered is income expropriation, if feasible. The reason why expropriation can be both individually *and* collectively rational is that the punishers stand to gain from expropriating a defector's income, and therefore continuation equilibria where defectors are punished can remain undominated by a renegotiated reversion to cooperation. However, expropriation adds a new source of temptation to behave opportunistically; so, if we call expropriation into play as a form of punishment against defectors, we must also account for an economy's ability to prevent expropriation attempts from occurring in equilibrium. For this to be possible, there must be force in numbers: a group of individuals acting against a defector must be able to confiscate from the defector an amount that is greater than the total amount an individual can gain by individually attempting expropriation against members of the same group. This idea is formalized below.

We will begin our discussion by focusing on the expropriation problem in isolation from the collective consumption problem. We will then examine how the possibility of income expropriation affects equilibria with collective consumption; as we will show, the availability

---

capita level of contribution required to support a given level g, however, is now smaller, and so is per capita income; then, the expressions $y - c$ and $\theta v(g)$ on the left-hand side of (4) (Proof of Proposition 1 in the Appendix) and $y - c$ on the right-hand side are all proportionally scaled down; the expression $\theta v(g - y)$, however, decreases less than proportionally, implying that (4) will be violated in the modified economy.

[9]More generally, renegotiation proofness rules out reputation-type equilibria—equilibria where actions that are not ex-post rational are supported by the credible threat of other players reverting to the belief of an ex-post rational action being adopted indefinitely following a single deviation.

of stronger means of punishment produces outcomes that can be even more convincingly interpreted as viable tax equilibria. For the sake of simplicity, our discussion will abstract from imperfect monitoring, although we shall keep focusing on punishment strategies in which the level of collective consumption remains unchanged during punishment.[10]

3.1 Equilibrium Property Rights

Suppose first that income is only used for private consumption ($\theta = 0$), and suppose that each individual is endowed with a certain amount of time—normalized to unity without loss of generality—which can either be used either for income generation, obtaining a maximum income y^i, or to engage in expropriation attempts against others up to a total fraction, ρ, of her total available time. Only a fraction, ζy^i, of income, representing income generated from market activities, can be expropriated, with the remaining income representing nonmarket income.[11] An expropriation attempt by i against an individual j uses a fraction $\gamma < \rho$ of i's time, and therefore involves a cost γy^i.[12] The individual can thus at most participate in $\bar{n} = \rho/\gamma$ expropriation attempts. For simplicity, we shall assume $\bar{n}$ to be an integer number. We also assume that γ, ρ, and ζ are the same for all individuals,[13], and that $\rho < 1 - \zeta$—implying that the expropriable income produced by i is ζy^i, independently of whether i engages in the expropriation of others.

If a group of individuals other than j, consisting of $s \leq n$ individuals, attempt expropriation against j, they can collectively expropriate a fraction, $\beta(s)$, of her income, y^j—obtaining collectively an amount equal to $\beta(s)y^j$—where $\beta(s) = \xi(s)\zeta$, $0 \leq \xi(s) \leq 1$, $\xi'(s) > 0$,.[14] We shall also assume that costless, voluntary transfers between individuals are feasible, implying that the expropriating group can divide $\beta(s)y^j$ amongst themselves as they choose. The relationship between the size of the expropriating group and the expropriation outcome—represented by $\beta(s)$—captures the strength of the group vis-à-vis the

[10]This feature could be rationalized—in a scenario with income expropriation and imperfect monitoring—by invoking the same arguments of Section 1, without otherwise affecting the structure of punishment strategies as we describe it here (see the proof of Proposition 10.)

[11]This implies that the individual's market productivity is assumed to be the same as her nonmarket productivity.

[12]Assuming γ to be the same for all individuals that all individuals have the same strength when expropriating others—an assumption that will be relaxed later in section 3.4.

[13]Allowing for heterogeneity with respect to such parameters is not necessary to present our arguments, but would considerably enrich the model's predictions.

[14]As producer of the income y^j, individual j can be thought of as enjoying a natural advantage in her claim (as explicitly modelled in Grossman, 2001), such as the advantage conferred by the initial physical possession of a tradeable good, which limits the expropriation that can be carried out by others. This feature is consistent with the observation that in all societies property rights have been enforced by a combination of both legal and physical means of protection (enclosures, locks, etc.) conferring a possession advantage to the owner.

individual, which depends upon both the expropriation technology and the extent to which income is generated in an expropriable form.

In this framework, property rights over transferable goods are not exogenously enforced and can only emerge as an equilibrium phenomenon (when the equilibrium level of expropriation effort is zero). On the other hand, initial possession of income entails a preferential claim because forcible expropriation is costly, so that efficient outcomes will feature as little expropriation as possible: aggregate disposable income will be maximized when no expropriation attempts take place. But if $\beta(1)$ (the fraction of income that can be expropriated by a single individual) is large enough and γ is not too large (i.e. if expropriating others' income is sufficiently easy), an efficient, no-expropriation outcome cannot be supported within a single round of interaction.[15]

Repeated interaction, on the other hand, may make it possible to prevent expropriation attempts from occurring. Specifically, a no-expropriation outcome can be supported by renegotiation-proof punishment strategies in which punishment only lasts for one period, and which are described as follows: each player i, $i \in I$, does not attempt expropriation against any other player as long as the other players do the same; if any player i deviates from this course of action at t, then in period $t+1$ player i becomes an expropriation target for a group of $s_i \leq n$ of the other players and fully accommodates expropriation by others; if player i accommodates punishment in this way at $t+1$, all players revert to no expropriation from $t+2$ onwards, otherwise forgiveness of player i is postponed until player i accommodates punishment.[16] As long as the amount collectively expropriated from the defector during punishment is greater than the expropriation cost collectively incurred by the expropriators, at least one of the expropriators will be better off carrying out the punishment in comparison with any alternative continuation equilibrium.

Proposition 7 *Absent collective consumption, property rights can be supported by the threat of collective expropriation of violators' income.*

In an economy where the number of individuals of each income type is large, individuals will be most tempted to attempt expropriation against those individuals with the highest productivity, and punishment through collective expropriation in a constrained-efficient equilibrium will be carried out most effectively by individuals having the lowest market productivity—thus minimizing expropriation costs and maximizing the collectively

[15]For example, in the symmetric case where all individuals are identical ($y^i = y, \forall i$), if $\beta(1) > \gamma$, a symmetric, pure-strategy, no-expropriation outcome does not constitute a one-shot noncooperative equilibrium.

[16]As before, our discussion focuses on equilibria that are both ex-ante constrained-optimal and anonymous (identical individuals play identical strategies). Asymmetric equilibria that are ex-post constrained-efficient are in principle also possible—for example, equilibria where a subset of individuals permanently "gang up" against one individual and expropriate her income—but these equilibria violate anonymity and are ex-ante dominated.

sustainable level of punishment. The gross gain from unilateral expropriation attempts will thus be the same for all productivity types, whereas the cost of individual expropriations and the ensuing punishment will be increasing with the expropriator's market productivity. Then, an individual's temptation to attempt unilateral expropriation will be larger, relative to the punishment, the lower her market productivity; it follows that property rights will be more difficult to establish the more unequal is the distribution of income. Also, property rights will be easier to establish the weaker individuals are relative to groups as expropriators.

Proposition 8 *Absent collective consumption and direct transfers in equilibrium, the establishment of property rights requires expropriation costs to be comparatively small and group strength to be comparatively high, and requires a higher degree of patience the more unequal is the distribution of income.*

If direct transfers between players are used along the equilibrium path, they may make it possible to prevent expropriation from occurring even when incomes are very unequal. Suppose that it is possible for an individual to make voluntary transfers to other individuals. Such transfers will never be used as part of a one-shot noncooperative equilibrium—individuals cannot directly gain from transferring income to others—but they may be used under repeated interaction as part of an equilibrium where the withdrawal of transfers is used as a credible punishment against defections alongside collective expropriation, and where the transfers themselves are supported by the threat of collective expropriation of individuals who decide unilaterally to withhold them. Voluntary transfers can thus ease distributional tensions by increasing the opportunity cost of defections for low-income individuals—the individuals who are most tempted to expropriate. Even if the potential victims of expropriation must surrender part of their income to the potential expropriators, these equilibria can Pareto dominate equilibria with expropriation because they avoid expropriation costs.[17] Such transfers (which can be interpreted as representing a system of redistributive taxes) are not the result of preferences for redistribution but serve to complement collective expropriation in support of property rights.

Proposition 9 *Transfers (negative taxes) may be needed to support a no-expropriation equilibrium.*

On the other hand, the sustainable level of transfers may not be large enough to prevent expropriation attempts in equilibrium.[18]

[17]Note that transfers would never be used in any way absent income expropriation, either in equilibrium or during punishment.

[18]This can be shown by the following example. Suppose that half of all individuals have a certain income positive y' and the other half have an income of zero. With n large, the net gain a zero-income

3.2 Income Expropriation and Taxation

Having dealt with the question of how property rights can be sustained, we are now in a position to ask how the feasibility of income expropriation affects the sustainability of private contributions to collective consumption.

In an economy where collective consumption has a positive value ($\theta > 0$), if expropriation is possible, the following punishment strategy can be used in place of the punishment strategy described in Section 2: if an individual i deviates from equilibrium play in period t, the other individuals expropriate i in period $t+1$, extracting from i an amount equal to $\beta_i y^i$, which is directed to funding collective consumption, and in the same period they collectively reduce their contributions by an amount equal to $(\beta_i + \eta_i(1-\beta_i))y^i$, $0 \leq \eta_i \leq 1$, (the amount of income expropriated from i plus a fraction of the portion of i's income that is not expropriated); player i accommodates expropriation at $t+1$ and at the same time changes her contribution to $\eta_i(1-\beta_i)y^i$; if player i does this, then all players revert to equilibrium play from $t+2$ onwards, otherwise forgiveness is postponed until player i accommodates punishment. This punishment strategy (weakly) dominates the no-expropriation strategy described in Section 2 in terms of the punishment that can be inflicted on defectors, since the temptation not to accommodate punishment by making contributions is comparatively reduced: the contribution required by the defector in the punishment phase is now $\eta_i(1-\beta_i)y^i$ rather than y^i.[20]

Proposition 10 *In a large economy where income can be forcibly expropriated, constrained-efficient renegotiation-proof equilibrium strategies punish a reduction in contributions below their equilibrium-play level by a temporary increase above the equilibrium-play level in the defector's payment, part of which is obtained by direct expropriation and part of which is voluntarily contributed by the defector.*

Contributions that are supported by a combination of fines and forcible expropriation more closely resemble real-world taxes; from here on we shall refer to them simply as taxes.

Relying on expropriation during punishment strictly dominates relying only on increased contributions as a means of punishment, and is effective in large economies as well as

individual gets from expropriating others is then $(\beta(1)\rho/\gamma)y'$. In order to be induced not to expropriate, the zero-income individuals must each receive a net transfer greater than $(\beta(1)\rho/\gamma)y'/\delta$; then, since the number of high- and low-income individuals is the same, the cost of funding such transfers to each one of the high-income individuals would also be equal to $(\beta(1)\rho/\gamma)y'/\delta$. Therefore, if $1 > \beta(1)\rho/\gamma > \delta^2\bar{\beta}'$, where $\bar{\beta}'$ represents the maximum collectively rational rate of expropriation towards an y' type individual (see Proof of Proposition 7), the defection gain for a high-income individual is larger than the punishment incurred;[19] in this case, a no-expropriation equilibrium cannot be supported by transfers.

[20]Direct earmarking of expropriated income to collective consumption (weakly) dominates a punishment strategy whereby expropriated income is distributed to the punishers, because it reduces a punishers' ability to gainfully deviate from the stated punishment strategy during punishment.

in small ones. This can be seen by repeating the scaling experiment we described at the end of the previous section—whereby n is increased holding both aggregate income and the efficient level of collective consumption constant in a homogenous economy: with expropriation, and if the maximum collectively rational rate of expropriation of a defector's income, $\bar{\beta}_i$, is greater than $1-\delta$, as n approaches infinity the ratio of sustainable collective consumption to aggregate income approaches a positive asymptote. Thus, expropriation can make it possible to overcome the effect of large numbers on free riding, in the sense that it makes it possible to support a level of collective consumption that does not vanish in size—when measured relative to its efficient level—as the economy becomes progressively larger.

Proposition 11 *In a large, homogenous economy where income expropriation is feasible, and if the collectively rational rate of expropriation against noncontributors is sufficiently large, an increase in the number of individuals, holding total income and the optimal level of collective provision constant, leaves the maximum sustainable level of collective consumption in a constrained-efficient equilibrium unchanged at a level that is strictly above zero.*

However, expropriation also increases the temptation to defect by expropriating others. With expropriation it may be impossible to support any positive level of tax for low income individuals, for whom attempting unilateral expropriation of others is profitable, in a constrained efficient equilibrium featuring no expropriation. In such situations, a constrained efficient equilibrium may require transfers (negative taxes) to low-income individuals—as discussed in the previous subsection—which in turn reduce the fraction of total revenues that can be channelled towards collective consumption.[21] Nevertheless, if we focus on a scenario with identical individuals, we can show that in economies where property rights can be independently established—i.e. in economies where the conditions of Proposition 8 are met—expropriation possibilities always help to solve the free-riding problem in the provision of collective consumption.

Proposition 12 *The maximum level of collective consumption that can be supported in a large economy where expropriation is feasible and where property rights are otherwise established is higher than the corresponding level in an economy where expropriation is not feasible.*

Thus, if expropriation goes hand in hand with sustainable property rights, it can make it possible to support higher levels of collective consumption.

Conversely, it can be shown that the need to engage in collective consumption can facilitate the establishment of property rights. Specifically, in situations where transfers are

[21] For these individuals, the negative tax, unlike a positive tax, will remain unchanged during deviations and will vanish in the punishment phase.

required to prevent unilateral expropriation attempts from occurring, indirectly channeling the transfers through collective consumption can make it easier to keep expropriation incentives in check. This is because, although the threat of reduced provision of collective consumption can deter expropriation attempts just as the threat of reduced transfers can, indirect transfers can be more easily sustained by those making them if they also benefit from the increased collective consumption—both because this lowers the temptation to withhold the transfer, and because it increases the punishment that other contributors can administer to those who fail to contribute.

Proposition 13 *The need to engage in collective consumption can make it possible to sustain property rights even in situations where it is not otherwise possible to do so.*

Furthermore, the value of an indirect transfer can be larger to the receivers than that of a direct transfer. This will be the case if collective consumption is at a suboptimal level for which the sum of the recipients' marginal valuations is greater than the cost; i.e. the reduction in the level of collective consumption during punishment could be worth more to the low-income individuals than the implicit money transfer associated with it, for if they received a direct transfer instead, they would themselves be unable to sustain the same level of collective consumption. In this case, the high-income individuals vicariously solve the free-riding problem for the zero-income individuals through indirect transfers. As an extreme illustration of this mechanism, suppose that only low-income people value public goods, but are unable to sustain collective provision because their low income limits the punishment they face. Higher-income individuals will then be indifferent between making direct transfers and contributing to collective consumption; contributions towards the public good, however, are worth more to the low-income recipients, and the plausible threat of terminating them is therefore comparatively more effective as a deterrent against expropriation attempts by low-income individuals.

This prediction is consistent with governments' use of public good provision as a redistributive tool—a practice to which a sizable theoretical literature has been devoted. Our discussion would suggest that public provision of collective goods is more likely to be used to moderate distributional tensions in this way the greater are the specific collective consumption needs of lower-income individuals. In some cases, however, direct transfers may be additionally required to support property rights—for example, if only a subset of lower-income types value collective consumption. Thus, redistributive provision of collective goods can co-exist with direct redistributive transfers.

The equilibria we have described above can be viewed as representing self-enforcing social contracts that combine two distinct dimensions of social cooperation—the provision of collective consumption and the collective enforcement of individual property rights. Our analysis has shown how these two aspects become tightly knit together, with one dimension of social interaction affecting the ability to sustain cooperation in the other. This illustrates a general point concerning sustainable social contracts: the question of how taxes

can be sustained cannot be meaningfully addressed abstracting from other forms of social cooperation.

3.3 Voluntary Contributions

If taxes are simply interpreted as self-enforcing contributions—albeit supported by the threat of income expropriation—then how do we explain the existence of "voluntary" contributions, contributions which individuals view as being separate from taxes and non-compulsory? This observation can be reconciled with our previous discussion once we consider the implications of unobservable heterogeneity. Combined with observable income differentials, heterogeneity in preferences may result in equilibria supported by punishment strategies that only partially rely on expropriation, and where taxes and voluntary contributions can be separately identified, as the following discussion illustrates.

Suppose that there is preference heterogeneity in the economy and that preferences are not observable, and consider the case where the income level is the same for everyone. There can then be three possibilities for a constrained-optimal equilibrium: (a) everyone is required to pay a common contribution level (tax) supported by the punishment strategies described in the previous subsection, and there is full compliance; (b) everyone is required to pay a common contribution level (tax) supported by the same punishment strategies, but there is only partial compliance, with lower-preference individuals not paying the tax; (c) everyone pays a certain minimum contribution level (tax) supported as above, but, in addition, a subset of players make a contribution above the tax, with these additional contributions being supported by the punishment strategies described in Section 2—anonymous punishment by reduction in collective contributions only, without invoking the threat of expropriation. An equilibrium such as that described under (c) will always (weakly) dominate an equilibrium as under (a), since using the anonymous punishment strategies in addition to the expropriation-based strategies can never result in lower aggregate contributions. An outcome such as (b) may be favoured if the number of high-preference individuals relative to that of low-preference individuals is sufficiently large that inducing the highest possible level of contributions by them—by relying on the threat of expropriation—dominates the cost of defection by the low-preference individuals.

Proposition 14 *In the presence of unobservable preference heterogeneity, a constrained-efficient equilibrium may feature differentiated punishments with differentiated triggers, i.e. taxes paid by everyone under penalty of expropriation and additional contributions paid by a subset of individuals under penalty of a reduction in contributions by others.*

When collective consumption involves multiple collective goods, a constrained-efficient equilibrium may feature earmarking of taxes and voluntary contributions to different goods. Suppose, for example, that preferences towards a certain collective good are concentrated within a subset of individuals. Then, it will be comparatively easier to support contributions

towards that good without reliance on the threat of expropriation. A constrained-efficient equilibrium may then involve earmarking of any voluntary contributions towards this good, with tax revenues being directed to the provision of the remaining public goods.

Voluntary contributions can also be instrumental in preventing expropriation attempts from occurring in equilibrium (i.e. they can be motivated by the need to preserve property rights even when they are not themselves supported by the threat of expropriation). Going back to the earlier example, where half of the individuals are haves and the other half are have-nots, suppose that only a subset of the haves value public consumption highly, while the remaining haves do not. Then, the contribution level by the haves that can be supported by the preference-anonymous punishment of collective expropriation may be too small to prevent expropriation attempts by the have-nots. In this case, then, it is possible for a constrained-efficient equilibrium to feature voluntary contributions by the high-income, high-preference types that are above the tax required from high-income individuals of all preference types *under threat of* expropriation, and whose effect is *to prevent expropriation* by low income individuals. Such additional contributions can only be significant in size if the group making them is sufficiently restricted (because as we noted in Section 2, contributions supported by renegotiation-proof strategies that do not rely on expropriation are limited in large groups); i.e. they must be made by a comparatively small "elite" of higher-income individuals. These additional contributions provide a collective service to the members of the elite in that they secure property rights for that elite. Compliance by individual elite members, however, is not elicited by the threat of expropriation by low-income individuals, but by the threat of a reduction in contributions by other elite members.

3.4 Punishment by Jail

What about more severe forms of punishment beyond expropriation? As we noted earlier, renegotiation proofness requires punishment to be both individually and collectively rational. However, punishment strategies that are not collectively rational under full information may be collectively rational on the basis of beliefs that are sequentially updated under asymmetric information. To illustrate this idea, consider the case of punishment by jail.

So far, we have focused on equilibria where expropriation is the only collectively sustainable form of punishment besides the withdrawal of collective provision (or transfers) but where property rights can nevertheless be enforced. In such equilibria, reductions in contributions face the same penalty as expropriation attempts: both are punished by expropriation. In reality more severe forms of punishment—such as jail or execution—are used against offenders; moreover, the infringement of property rights is typically punished more severely than failing to contribute to collective consumption is.

Suppose, however, that it is possible to jail (or execute) an individual, preventing the individual from attempting expropriation indefinitely. Also suppose that jailing an indi-

vidual involves a cost which is greater than the expropriation loss incurred by any single individual but less than the collective expropriation loss incurred by the group from a unilateral expropriation attempt. Since the cost of jailing an individual is larger than the expropriation cost incurred by any single individual, jailing cannot be an individual best response to expropriation attempts—and therefore cannot be part of one-shot equilibrium strategies. Neither can it be part of a renegotiation-proof equilibrium under repeated interaction if it is possible to prevent expropriation attempts through transfers that are less than the cost of jailing expropriators—since unlike expropriation, punishment by jail does not generate a gain for the punishers.

If jailing is a lower-cost alternative for the victims to using transfers, or if transfers cannot prevent expropriation attempts from occurring and, at the same time, the cost of jailing offenders is less than the expropriation costs offenders impose on the victims, pre-emptive jailing (possibly alongside transfers) may be part of a renegotiation-proof equilibrium in which some individuals are jailed before they have a chance to offend, at a cost which is shared amongst all remaining individuals (contributions towards the cost of jailing can in turn be supported by the threat of expropriation). In a renegotiation-proof equilibrium, however, pre-emptive jailing can only be directed against those individuals who cannot be dissuaded from offending by any other viable means, e.g. by transfers that are less costly than jailing. Then, if these individuals—whom we shall call "irredeemable offenders"—can be directly identified by others before they can attempt expropriation, ex-post punishment of offences by jail cannot be used as a plausible threat against all other individuals in a renegotiation-proof equilibrium. In other words, under full information, jail can be used for the purpose of prevention but it cannot be used as a deterrent.

On the other hand, if individuals cannot perfectly identify potential expropriators, jail can act as a credible deterrent. Consider a scenario where the irredeemable offenders are individuals who have a low expropriation cost parameter, γ^i, which is here assumed to be individual-specific and private information. Then an irredeemable offender cannot be identified ex ante, but carrying out expropriation attempts may reveal her to others. In this case, jail can become a credible and plausible threat for all individuals in a separating, renegotiation-proof, perfect Bayesian equilibrium in which ex-post beliefs are derived from ex-ante behaviour consistently with Bayesian updating: suppose individuals believe that only irredeemable offenders attempt expropriation; on the basis of this belief, it is rational for the other players to jail offenders ex post; if this is how offences are punished, it is rational ex ante for individuals to expect that all expropriation attempts will be punished by jail; if, facing this threat, no individual outside the group of irredeemable offenders finds it optimal to offend, then the belief that only irredeemable offenders attempt expropriation will be consistent with observed behaviour. Such an equilibrium will be renegotiation-proof in an expected sense, because at $t+1$, if the punishers believe a defector to be an irredeemable offender, there is no scope for recoordinating to a continuation equilibrium where offenders are not jailed and all individuals experience a higher expected continuation

payoff.[22] Thus, anonymity can play a crucial role in sustaining cooperation.[23]

In contrast with the case of expropriation attempts, it can never be rational to use jail against noncontributors, neither as an individual nor as a collective response: jailing does not generate a gain from the jailers nor can it prevent any individual from noncontributing, and cannot therefore be used as a deterrent against reductions in contributions no matter what the information structure is. Then a constrained-efficient equilibrium may involve expropriators and noncontributors being punished differentially: expropriators face jail, noncontributors face expropriation.

Proposition 15 *If individual expropriation costs are private information, a constrained-efficient perfect Bayesian equilibrium may involve unilateral expropriation attempts being punished by jail, with noncontributors only facing expropriation and/or fines.*

With jail punishment, property rights will be comparatively more robust, and it will be comparatively more likely—in comparison with an economy where expropriation is not

[22]In reality, there may be gray areas. For example, there may be several types of crime, some more serious than others, and some of them may be attributable to irredeemable offenders while others may not. In this case, jail may be sustainable in a perfect Bayesian equilibrium only for some types of crime and not for others, implying that there will still be scope for using transfers and/or fines/expropriation in order to deal with some types of crimes. Beliefs may also be historically determined—as discussed by Greif, Milgrom, and Weingast (1994)—and self-propagating. For example, suppose that there are really no irredeemable offenders, and that active expropriations are in fact only the result of accidents. Nevertheless, starting from a belief that the probability of irredeemable offenders occurring in the population is greater than the probability of a mistake occurring, it may be a best response for those who hold such belief to punish accidental deviators by jail (without chance of redemption), therefore never testing the belief that irredeemable offenders actually exist. Such equilibria satisfy the conditions for a perfect Bayesian equilibrium, but not those for a sequential equilibrium in the sense of Kreps and Wilson (1982), since they are supported by sequentially rational beliefs at nodes that are reached with probability zero in equilibrium. Diverging and self-sustaining beliefs concerning the application of punishment could also help rationalize situations where only a subset of individuals comply with a formal tax that is not generally enforced, as well as differences in compliance standards across groups, producing an effect analogous to that of different collective behavioural standards, as has been described in the social norms literature. For a recent treatment of the role of social norms in tax compliance, see Ratto, Thomas, and Ulph (2004).

[23]Anonymity, however, can also limit the possibility of partial-compliance equilibria. Although our analysis has focused on constrained-efficient, high-compliance equilibria, even under full information lower-compliance equilibria are possible alongside high-compliance ones. Then, lower-compliance equilibria could be thought of as the result of some sort of equilibrium selection failure on the part of the group. Consider, for example, compliance equilibria supported by expropriation-based punishment strategies. Expropriation is by its nature nonanonymous, and there will exist equilibria where punishment strategies are such that only certain individuals—identified in some fashion—and not others are punished by expropriation if they fail to comply. However, such equilibria would be ruled out under an anonymity restriction requiring that punishment strategies be nondiscriminatory. Thus, anonymity provides another mechanism on the basis of which one can rationalize selection of full-compliance equilibria (nondiscrimination is indeed held as a fundamental principle in most institutional systems).

feasible—that the possibility of expropriation will help to support collective consumption, rather than collective consumption helping to support property rights. Thus, jail punishment can help sustain high levels of collective consumption. But this is not because jail is used directly to enforce taxation; it is because it can help establish independently secure property rights, which in turn reduces the need to use transfers for the same purpose, freeing tax revenues for funding collective consumption.

4 Renegotiation-proof Tax Equilibria: Predictions

Our discussion so far has mostly focused on retracing predictions that can also be generated by models of coercive taxation. However, a model of self-enforcing taxation and insecure property rights also yields additional predictions that are either distinctive of such a model or find no parallel in analyses of coercive taxation.

4.1 Property Rights and Collective Consumption

The results in the previous section points to the existence of a relationship between property rights and collective consumption. The enforcement of property rights has been recognized as one of the basic functions of government, and has even been described by public finance writers as a public good.[24] Technically, the enforcement of laws and contracts cannot be classified as a form of nonrivalrous collective consumption, as single acts of enforcement benefit specific individuals; rather the collective dimension of enforcement consists of its reliance on the strength of the group vis-à-vis the individual. Nevertheless public finance writers have long recognized that law enforcement and the provision of collective goods have always co-existed.[25] The theoretical framework we have described enables us to explore this relationship formally.

As noted earlier, one can think of expropriable output as being associated with market activities. For given characteristics of the expropriation technology, as summarized by the function $\xi(s)$, what is the effect on the supportable level of collective consumption of an increase in the relative importance of market activities if this simply translates into an increase in the portion of income that can be expropriated? In our model, this amounts to examining the effects of an increase in ζ, resulting in an equiproportional upward shift in the expropriation schedule $\beta(s) = \xi(s)\zeta$.

It can be shown that in an economy where, absent collective consumption, property rights can be established without the help of transfers, an increase in ζ directly translates

[24]Public finance textbooks, such a Buchanan and Flowers (1975), have often characterized Law and Order as a publicly provided public good.

[25]Adam Smith's *Wealth of Nations* (Book V, Chapter II, Part II) lists law enforcement as "the second duty of the sovereign," after defence but before public good provision.

into an increase in the severity of the punishment for noncontributors and hence in tax revenues.

Proposition 16 *In a large, homogeneous economy where income expropriation is feasible but where property rights are independently secure, an increase in the share of income that can be expropriated can raise the maximum sustainable level of collective consumption.*

We can interpret this result as predicting that the emergence of market activities can facilitate the development of collective consumption activities.[26] This prediction is consistent with the observation that collective consumption is a characteristic of advanced market economies.[27] a pattern that has been typically associated with the idea that publicly provided infrastructure and education is a key ingredient of economic growth.[28] While this latter form of causation is undoubtedly present, our analysis shows that the reverse form of causation may also be important.[29]

4.2 Tax Progression

If we examine the question of how taxes vary with income, we find that, as in the no-expropriation scenario discussed in Section 2, individuals with the same income will pay equal taxes (horizontal equity) and contributions will be increasing with income (ability to pay). As noted earlier, however, in the presence of expropriation, taxes can be negative. Furthermore, the rate of tax progression in this case will also depend on the expropriation technology, and will be different depending on whether taxes are positive or negative.

Focusing on the large-economy case, it can be shown that average rates will be increasing with income and marginal rates will be positive. Other things equal, marginal tax rates will be higher for low-income individuals who are potential expropriators—and more so for individuals who are net recipients of transfers—than for higher-income individuals who are not potential expropriators: this is because a marginal increase in income for a potential expropriator directly reduces the temptation to engage in expropriation; this relaxes the

[26] Another channel through which the relationship between market development and collective consumption may be established is the possibility of expropriating the capital assets that are a pre-condition for the rise of economic specialization and market exchange—a channel that is absent from our analysis.

[27] Tax revenues to GDP ratios are comparatively higher in OECD countries.

[28] See, for example, Mankiw, Romer, and Weil (1992).

[29] Warfare may also be an important channel through which the linkage between property rights and collective consumption is established. If a group is subjected to an external threat of expropriation, and if individual group members are protected from outsiders by fellow groups members—at a cost to them—then individuals who fail to contribute to collective consumption may be punished by other members of the group withholding their protection; which they can do credibly and consistently with collective rationality. This suggests a possible positive relationship between warfare and collective consumption. Modelling this relationship requires a somewhat richer framework than the one we present here, and is left for future work.

relevant enforcement constraint(s), which in turn makes it possible to further raise the tax required from the individual. This effect is absent in the case of high-income individuals for whom unilateral expropriation of others is unprofitable, or when jail is an effective deterrent against unilateral expropriation attempts.

Proposition 17 *In a constrained-efficient, self-enforcing tax equilibrium for a large economy, average tax rates will be (weakly) increasing in income and marginal tax rates will be positive. Negative taxes will be associated with comparatively higher marginal tax rates, and marginal tax progression may taper out with income, with a constant marginal tax rate applying at the higher end of the income distribution.*

Real-world tax systems do rely on coercive expropriation of noncontributors, and, at the bottom end of the income distribution, feature transfers that are characterized by high replacement rates (progression); also, as predicted by our analysis, progressivity typically flattens out at the top of the income distribution—a feature that traditional analyses of optimal taxation have been at pains to explain.[30]

4.3 Income Inequality and Collective Consumption

Given that collective consumption, can be more effective at preventing expropriation attempts than direct transfers are, we could expect that more income inequality, which increases the temptation of unilateral expropriation, may lead to increased collective consumption. However, the maximum tax that is sustainable for high-income individuals is independent of the expropriation gain by low-income individuals; thus, while collective consumption can help prevent expropriation attempts from occurring, an increased threat of unilateral expropriation by the poor need not translate into higher taxes—since the highest sustainable tax for an individual is only a function of the plausible threat of collective (rather than unilateral) expropriation by the group.

On the contrary, an increase in income inequality may trigger a switch from a scenario in which low-income individuals are not tempted to expropriate higher-income individuals to one in which they are; then, if direct rather than indirect transfers can be more effective at preventing expropriation attempts (say, because the potential expropriators do not value collective consumption) and such transfers are part of an equilibrium, increased inequality would require more direct transfers, reducing the level of collective consumption that can be supported. In fact, it can be shown that if marginal rates of sustainable taxation are decreasing in income, albeit still positive (Proposition 17), increased inequality will lower net tax revenues.

[30]Optimal income tax models can predict *falling* marginal rates at the top of the income distribution, but have difficulty explaining marginal rates remaining constant beyond a certain income level. See Bevan (2004) for a recent discussion and alternative theoretical explanation.

Proposition 18 *In a large economy, a mean-preserving income spread that raises the income level of a given above-mean income group and lowers the income level of a different below-mean group, leaving income unchanged for all other groups, weakly lowers the level of collective consumption.*

This predictionis unrelated to the presence of redistributional goals. Models of optimal redistributive can predict decreasing marginal rates, but do not predict that increased inequality should lower public good provision.

4.4 Public and Private Provision

How does public provision crowd out voluntary provision in a model of self-enforcing taxation?

As we have described in Section 3.3, when there is unobservable preference heterogeneity, contributions in excess of the tax can arise in a renegotiation-proof equilibrium. Then, for those individuals, contributions are identified by the same conditions that apply to a no-expropriation scenario.

Consider now an increase in tax revenues that can be collected from individuals other than those who make voluntary contributions (this may be the result of a change in the distribution of income within that group of individuals, to an exogenous change in the expropriation technology or to any such other change that leaves voluntary givers otherwise unaffected), focusing for simplicity on a homogeneous-economy case. In this scenario, a model of self-enforcing contributions predicts less than perfect crowding out of private provision by public provision.

Proposition 19 *An increase in the public provision of collective consumption in a homogeneous economy causes less-than-perfect crowding out of private provision.*

This prediction is consistent with evidence from OECD countries but cannot be readily generated by static models of voluntary contributions.[31]

5 Other Dimensions of Taxation

The preceding discussion of tax equilibria has abstracted from a number of issues that have occupied a pre-eminent role in traditional analyses of enforceable taxation. This section describes how the framework described in the previous section can be extended to accommodate some of these additional aspects of taxation.

[31] See Bergstrom, Blume, and Varian (1980).

5.1 Taxation and Incentives

Traditional analyses of coercive taxation have stressed the incentive effects associated with taxation. Self-enforcing taxation can call into play incentive effects analogous to those that stem from coercive taxes, alongside incentive effects that stem specifically from the presence of enforceability constraints, as illustrated by the following discussion (a more detailed treatment is provided in Appendix B).

Suppose that individuals also use some of their time for leisure consumption, l^i, which affects utility in a quasilinear fashion, i.e. $u^i(x^i, l^i, g) = x^i + \theta^i v(g) + h(l^i)$, where $h'(l^i) > 0$, $h''(l^i) < 0$. Assuming the individual does not engage in expropriation, her gross of tax income is $(1 - l^i)y^i$, where y^i can be taken to represent the individual's market productivity (wage), which may or may not be private information.

Consider first a scenario where individual productivity is publicly observable. Also focus on a large-economy, no-expropriation scenario. Along a constrained efficient equilibrium path of play, leisure will be selected so that its marginal value to the individual equals its market productivity, y^i, i.e. a level l^{i*} such that $h'(l^{i*}) = y^i$. By the same arguments presented in Section 2, defectors can be punished most harshly by a renegotiation-proof punishment strategy that requires them to take no leisure in the punishment period—earning an amount equal to y^i—and to contribute their full earnings to collective consumption.[32] This punishment strategy is renegotiation-proof, because once punishment has been triggered we cannot find an alternative continuation equilibrium that the punishers favour over the prescribed course of action.

Thus, with endogenous leisure, the punishment cost to defectors also includes the value of the leisure forgone $h(l^{i*})$ in the punishment phase; thus, the tax that can be viably levied on an individual will depend on that individual's "full income" rather than on her earnings alone. It can then be shown that the progression of the marginal tax rate with respect to y^i will be less than with fixed leisure (since the punishment depends on full income and full income is concave in y^i). On the other hand, as the ratio of earnings to full income increases with y^i (since l^{i*} is decreasing in y^i), the ratio of equilibrium tax to earnings will tend to fall, implying that taxes, when measured as a proportion of earnings, will tend to increase comparatively less. However, since earnings increase at a decreasing rate as a proportion of full income (because of convexity of h), this also implies that the positive effect of this mechanism on the marginal tax rate out of earnings will tend to flatten out with earnings.

If only earned income, $(1 - l^i)y^i$, is publicly observable (or verifiable), then punishment strategies must be conditioned on income and will produce labour-leisure distortions as in

[32] When expropriation is feasible, individual i's tax in a constrained-efficient equilibrium will be identified again by one of three possible regimes as discussed in the preceding section. In this case, however, the level of leisure taken in the punishment phase will play the same role that η_i plays in the fixed-leisure case.

conventional models. To illustrate, consider an economy with two productivity types, $\underline{y}$ and $\overline{y} > \underline{y}$, each present in numbers respectively equal to $\underline{n}$ and $\overline{n}$, and suppose that the minimum contribution levels that trigger punishment are a function of earned income. Then, the respective taxes for each type, $(\underline{c}, \overline{c})$, and the levels of leisure, $(\underline{l}, \overline{l})$, taken respectively by low- and high-productivity type individuals must satisfy the enforcement constraints that apply to each respective type, as well as a no-mimicking (self-selection) constraint requiring that high-productivity individuals have no incentive to mimic low productivity individuals by adjusting their labour supply so as to match the low-productivity type's earnings level and pay a correspondingly lower tax.

In this economy, if we were to abstract from the presence of enforcement constraints, or if these constraints are not binding, information problems could only matter if a redistributional goal is present: with coercive taxation, if the only reason for taxation is the funding of collective consumption, any desired level of collective consumption can be achieved with taxes that are uniform across productivity types, thus ensuring self-selection. If taxes must be self-enforcing, however, and if the maximum common level of self-enforcing taxation that can be supported is not enough to finance the efficient level of collective consumption, enforcement constraints will come into play. Starting from a common level of taxation that is sustainable for both high- and low-productivity types, additional revenue can be raised by raising the level of taxation for high-productivity types (given that their constraint is slack at that level); a higher tax on the high-productivity type, however will cause the self-selection constraint to bind.

Thus, as in a standard two-class economy redistributive tax problem, an optimal self-enforcing tax structure will entail distortions in labour supply choices at the bottom end and no distortion at the top end, even in the absence of a distributional goal. However, unlike in the case of coercive redistributive taxation, here a constrained-efficient tax structure is determined by a comparison between the marginal valuation of collective consumption and the marginal cost of the distortion associated with the no-mimicking constraint.

As shown in Appendix B, if the social value of a marginal revenue increase exceeds the cost of the associated marginal distortion on labour-leisure choices, it is possible for an optimum to be defined uniquely by marginal revenue effects. At such an optimum any further increase in the high-productivity tax would require a decrease in the low-productivity labour supply, which in turn, through the enforcement constraint, would require a decrease in low-productivity taxes and thus further tighten the no-mimicking constraint; the result could then be a decrease in revenues. In this case, the impossibility to observe (or verify) ability makes self-enforcing contributions to collective consumption distortionary just as in conventional models of distortionary coercive taxation, even if the choice of contribution levels is only dictated by revenue-maximization considerations.

5.2 Preferences for Redistribution

So far, we have only looked at situations where individuals are ex-ante risk neutral, which translates into neutrality towards ex-post inequality. This let us isolate the specific implications of enforcement constraints for the structure of taxes. Inequality aversion, if present, will have a role alongside enforcement considerations in determining tax progression.

In the case of coercive taxes, optimal tax structures result from a trade-off between efficiency and equity considerations. In contrast, with self-enforcing taxes enforcement constraints may end up dominating equity considerations, even when the resulting pattern of taxes is redistributive.

To see this, suppose all individuals have the same valuation for collective consumption, but there are $n/2$ high-income individuals and $n/2$ low-income individuals. Consider a no-inequality-aversion, constrained-efficient equilibrium as that which we have described, where high-income individuals pay positive taxes, low-income individuals receive a subsidy (a negative tax) and the net revenues are directed to collective consumption. Suppose now there is inequality aversion; then one possibility would be to divert some of the net revenues towards funding private consumption. By doing so, collective consumption is reduced and private consumption for low-income individuals is increased. However, if g is sufficiently smaller than g^*, diverting net revenues in this way will actually lower welfare for low-income individuals, each of whom values an extra dollar's worth of collective consumption more than she values an increase in private consumption of $1/(n/2)$ dollars—the additional per capita subsidy that could be funded with one dollar's worth of revenue. In this case, inequality aversion will have no additional effects on the structure of taxes: taxation will be redistributive, but its structure will be independent of inequality aversion—i.e. the same in a risk-neutral (utilitarian) society as in an infinitely risk-averse society (one which only cares for the welfare of the individual who is least well off ex post). Analogously, if g is close to g^* but high-income individuals do not value collective consumption, redistribution and enforcement will both be best served by a structure that directs all sustainable payments by the rich towards financing of collective consumption. Here, the structure of self-enforcing taxes in a constrained-efficient equilibrium will be invariant to changes in the degree of inequality aversion.

On the other hand, there will be scenarios where redistribution concerns bite. If, for example, only the rich value collective consumption and the poor have no ability to expropriate others individually ($\beta(1) = 0$), a constrained-efficient tax system in the absence of inequality aversion dictates that all revenues be directed to fund collective consumption (which is only enjoyed by the rich and which the poor may nevertheless be required to fund in part), whereas in the presence of inequality aversion, a constrained-efficient equilibrium would require that some of the revenues be directed to fund private consumption for the poor. This, however, will come at a cost in terms of enforceability, because diverting funds away from collective consumption will make it more difficult to sustain payment by the

rich. There exists, in this case, an enforcement-equity tradeoff, whereby taxes are more difficult to enforce if the revenues they generate are directed towards redistribution.

5.3 Monitoring, Evasion, Avoidance

Self-enforcing taxes as characterized so far are not inconsistent with the existence of monitoring activities against tax evasion, an aspect extensively studied in the literature on coercive taxation. Let us stress that the viability problem is quite separate from the problem of tax evasion and enforcement as it is usually intended. We never think of a tax evader as someone who refuses to pay a tax that she acknowledges she is liable to pay; we would describe this comparatively uncommon behaviour as civil disobedience. By tax evasion we typically refer to a situation where individuals make misrepresentations to the tax authority (e.g. underdeclaring their income) in order to reduce their tax liability. The framework we have described can be easily augmented to account for tax evasion and monitoring choices.

Consider the following variant of our model: income is exogenous but is only observable by others through costly monitoring activities. Efficiency in this case would require individuals to truthfully declare their income, so as to minimize the social costs of public monitoring. However, if monitoring of income never takes place, individuals are always induced to misrepresent it and so pay a lower tax—a tax whose level is nevertheless consistent with the application of renegotiation-proof punishment as described earlier. On the other hand, constantly monitoring all individuals may be too costly. Then, constrained efficiency may require the use of randomized monitoring, just as predicted by standard analyses. However, unlike in standard analyses of tax enforcement derived from Beckerian models of crime and punishment (Becker, 1968)—which have trouble justifying the use of limited fines in conjunction with frequent monitoring rather than reliance on extreme fines—here punishment would be automatically limited by the requirements of subgame perfection and renegotiation proofness.[33]

Our framework also makes it possible to provide a formal characterization of tax avoidance as distinct from tax evasion. Consider a situation where income is not directly observable but is believed ex ante to be perfectly correlated with a certain observable characteristic. Then punishment strategies could be conditioned on that observable characteristic. Suppose, however, that ex post there is a realization that the correlation was not perfect as initially thought, and that some individuals exhibiting a certain publicly observable characteristic actually have a higher income than that which the punishment strategies associate with that characteristic (i.e. a loophole is discovered in the tax system). Behaving rationally, those individuals would then only pay taxes that are consistent with the punishment

[33]In addition, if monitoring can lead to mistakes whereby low income individuals are assessed as being high income, the application of limited punishment would also be dictated by the need of containing the cost of mistakes.

strategies as written (i.e. they take advantage of the newly discovered loophole); those punishment strategies, however, were written incorrectly, although this only becomes apparent ex post. In these situations, although there is no misrepresentation involved, punishment strategies could be revised ex post to strike at the avoiders. Such punishment may range from the requirement to make up retroactively for unpaid taxes to the application of the maximum sustainable combination of expropriation and fines. However, if individuals are risk averse, the possibility of punishment strategies being revised ex post in this way would induce an ex-ante risk cost on individuals, a consideration which may, in a constrained efficient equilibrium, limit the application of punishment against avoiders. This would be consistent with the observation that tax administrations often impose only limited penalties on avoidance. In contrast, in any constrained-efficient equilibrium, tax evasion would always be punished as harshly as it is viable to do so.

6 Self-enforcing Tax Constitutions as Social Contracts

We have characterized a tax constitution as a self-enforcing contract that must be consistent with an equilibrium sequence of actions which are ex-post rational both from an individual and a collective perspective, but which is selected ex ante according to ex-ante expected utility maximization. Thus, the idea of viable taxation takes us naturally back to the Rawlsian/Wicksellian idea of unanimous collective choices under an ex-ante veil of ignorance (Rawls, 1971), which has traditionally been the intellectual foundation of utilitarianism and welfarism in general, and, more specifically, of theories of optimal taxation. In this context, however, utilitarianism is a predictive rather than a normative criterion: individual and collective rationality naturally suggests equilibrium selection according to ex-ante optimality, and so the outcome that is predicted is that which maximizes ex-ante welfare subject to ex-post sustainability constraints.

This interpretation is fully consistent with the traditional contractarian view of individuals consensually agreeing to a *binding* tax constitution, where ex-ante consensus guarantees that everyone gains from the existence of taxes.[34] Selection of an equilibrium that maximizes ex-ante expected utility means that the selected course of action gives everyone a higher expected payoff than any alternative (including the alternative of an indefinite repetition of one-shot noncooperative equilibria), and implies ex-ante Wicksellian unanimity. Furthermore, the continuation punishment equilibria that are used to support constrained-efficient outcomes are themselves self-supporting, renegotiation-proof equilibria; this means that, even once the veil of ignorance is lifted, each individual, even if not necessarily better off relative to *symmetric* Nash reversion, is better off along the equilibrium path than she

[34]Arguments for the use of benefit-based taxes are also often based on this principle (see, Brennan and Buchanan, 1980).

would be under some alternative, worst-case, feasible equilibrium (namely, the continuation equilibrium that would be adopted to punish her). In this sense, we can say that everyone can gain from a self-enforcing tax constitution both from an ex-ante and an ex-post perspective. Thus, the characterization of a self-enforcing tax constitution that we have presented in this paper follows very much in the contractarian tradition (and in doing so is counter to some more recent literature trends)—although, as noted earlier, enforceability places much more stringent requirements on the resulting structure of taxes than simple contractarianism does.

But, unlike in the traditional contractarian view, we cannot really think of ex-ante equilibrium selection as a contractual agreement in the literal sense of the word, given that no agreement can be binding, and no contracting is therefore involved. Rather, equilibrium selection can be interpreted here as individuals each independently acknowledging that a rational group of individuals will do best, from an ex-ante perspective, coordinating to a particular self-enforcing equilibrium. And, unlike in the standard contractarian story, ex-ante equilibrium selection must account for interim and ex-post incentives, i.e. all behaviours prescribed by the selected tax constitution must be compatible with ex-post individual rationality once the veil of ignorance is broken.

Our characterization of self-enforcing tax constitutions admittedly abstracts from organizational details of real-world government institutions, but it can be fully reconciled with them, and can be thought of as being complementary with analyses of government delegation structures—such as, for example Laffont (2000)—that in turn abstract from enforcement considerations. A fully developed theory of self-supporting government institutions would also need to integrate those aspects, an undertaking which is beyond the scope of the present paper.

Nevertheless, in the characterization of collective choices as resulting from a self-enforcing contract there arise significant problems of interpretation that we wish to address before concluding our discussion.

6.1 Uncertainty and Incomplete Contracting

Equilibrium selection according to ex-ante optimality does not mean that all contingencies must be explicitly be dealt with ex ante. In principle, under each contingency, a suitable self-enforcing continuation equilibrium can be reconstructed ex post and identified by all individuals on the basis of the principles that guide ex-ante selection; once the correct equilibrium has been identified, all individuals, anticipating that other individuals will do the same, will select the corresponding strategies as a best response.

But what about the "dimensionality curse," which is typically regarded as the main obstacle to complete state-contingent contracting? In the case of binding contracts, the problem would be that it is impossible to write down a contract detailing what all parties must do in all possible contingencies, including contingencies that are thought of as being

very unlikely. With a self-enforcing contract, however, it is not strictly necessary to do that: once the ex-ante criterion that drives equilibrium selection (e.g. ex-ante expected utility maximization) is known to all parties, when a contingency arises, individuals should in principle be able to reconstruct, from the ex-ante selection criterion, which particular self-enforcing equilibrium strategies individuals are supposed to adopt in that state of the world.

This idea can also be consistent with an environment where beliefs about the state of the world evolve as a result of learning, i.e. where individuals update their beliefs on the basis of new experience according to Bayes' rule: for example, individuals may think a certain state to be unlikely and so select a certain self-enforcing course of action, but after observing that state occurring, they may update their beliefs and move to another self-enforcing continuation equilibrium, again identified on the basis of the commonly known ex-ante selection criterion.[35] Thus, learning can be consistent with the idea that people play according to a pre-ordered scheme.

What is more difficult to accommodate in our construction is the idea that people may not have well defined beliefs about future events, e.g. a situation where something occurs which was wholly unforeseen.[36] However, this is a limitation of the standard way we think about choices under uncertainty and is not specific to collective choices—analogous problems arise in models of individual rational choice. In principle, the non-expected utility models that have been put forward in recent theoretical literature to model decisions in the presence of ill-defined priors could be adapted to our framework.

6.2 Finite Lives

A potential objection to the idea of repeated interaction supporting cooperation is the observation that individuals have finite lives. However, the idea of self-enforcing cooperation can also be applied to environments with overlapping generations (OLG) of finitely lived individuals (Kandori, 1992). Equilibrium strategies in OLG models involve a mechanism

[35] If different individuals have access to different pieces of new information, there will be an information pooling problem: if individuals do not concur about which state of the world they are in, they will also not concur about which continuation equilibrium the ex-ante selection criterion prescribes. We return to this issue below when discussing politics.

[36] We are not referring here to unforeseen contingencies which people cannot name but whose consequences may be understood. For example, individuals may be aware of the fact that certain unforeseen needs may arise with some probability (e.g. because this type of things tend to happen with some regularity) and that on average these events will affect them in a certain way (for example, requiring them to lower their private consumption). This type of uncertainty can be incorporated into ex-ante expectations. We are referring to situations where individuals are unable to form well-defined assessments of the likelihood of such unforeseen events, a situation that has been associated with the notion of "Knightian" uncertainty, and is referred to by more recent literature as a situation in which individuals have ill-defined priors (as, e.g., in Chen and Epstein, 2002).

akin to a retirement bonus: individuals cooperate when young and defect when old, but their defection when old will be accommodated by others (young individuals) if the old defectors have played cooperatively when young. Applied to taxation, this is consistent with unfunded social security systems as observed in many real-world systems.

In any subgame perfect equilibrium of an economy with finitely-lived individuals, individuals will always adopt their one-shot noncooperative best response in the final period of their life. Along a "cooperative" path of play, the other (young) players will accommodate this behaviour if the old player has in turn done the same when young, but will punish an old player who has failed to accommodate others when young. Punishment will thus consist of best responding to, rather than accommodating, the action chosen by the old player. The structure of such equilibria implies that renegotiation proofness will manifest itself differently in this case. Think of an economy where individuals live for two periods and where income cannot be expropriated. If a young player defects in the first period, in the last period of the defector's life, when the punishment is applied, there is little scope for the young and the defector, now old, to recoordinate to an alternative equilibrium: given that the old player's behaviour is simply a best response to the young players's actions, renegotiation proofness simply requires that the action chosen by the young be optimal for them taking the old player's reaction into account (a Stackelberg equilibrium). In the case of contributions to collective consumption, it will always be unilaterally optimal for an old player in a large economy to select zero contributions, so a fine cannot be part of punishment. Punishment by the young player will simply consist of collectively reducing the level of collective consumption from its equilibrium level to a level that is collective rational for the young. Punishment will then be particularly effective if there exist multiple forms of collective consumption some of which are old-cohort specific (i.e. used by the old and paid by the young), so that it will be collectively rational for the young to reduce provision of these goods to zero in the punishment phase. Thus, equilibrium contributions will be age-dependent.[37]

With income expropriation, punishment will consist of expropriation and a reduction in collective consumption for adult defectors and expropriation and a fine for young defectors, the latter supported by a punishment consisting of expropriation and a reduction in collective consumption against young defectors who do not repent when adults.[38]

[37]Not only will contributions not be constant through an individual's life, it may also be nonmonotonic with respect to age (see Appendix B).

[38]Again, this will entail contributions that are first increasing and then decreasing with age. An issue that arises naturally in connection with expropriation in a context where lives are finite is that accumulated assets may be expropriable. Then, since the profile of asset holdings is also first increasing and then decreasing with age, punishment by expropriation would also entail higher taxes for middle-aged individuals than for young and old individuals, and may in fact produce a contributions profile that is consistent with wealth rather than income taxation. If direct conditioning of taxes on age or wealth is not feasible, this may have to be achieved indirectly by differentiated taxation of labour and capital income.

6.3 Politics

How can we reconcile the idea that collective choices are the result of a self-enforcing contract, selected ex ante and once-and-for-all, with the observation that real world institutions feature formalized mechanisms of collective choice—such as voting? Once more we can note that groups cannot rely on any outside power of enforcement to make a political constitution permanently binding; therefore, whatever procedures of collective choice are adopted in equilibrium should be interpreted as being the result of a continuously-renewed, individually and collectively rational choice, and the ensuing collective choices must in turn be independently sustainable. The idea of ex-ante selection of a self-enforcing equilibrium then seems to leave no room for politics. Why should individuals vote, if coordination to a certain constrained-efficient outcome can be achieved spontaneously, and if the political mandate that is conferred by the vote does not per se convey any additional power of enforcement?

It may be tempting to argue that political competition, in this framework, may work as an ex-post equilibrium selection mechanism: although one self-enforcing contract may be ex-ante optimal, other equilibria are also sustainable; then, ex post, once a distribution of individual types is realized, the different types may use the political process as an imperfect substitute for bargaining in order to agree to the ex-post selection of a certain outcome. This interpretation, however, clashes directly against the self-enforcing nature of cooperation. If disagreement occurs, what should the fallback outcome be? Given that cooperation is self-enforcing, it can be any continuation equilibrium, including those featuring cooperation.[39] Hence, the application of bargaining ideas to a repeated interaction framework appears problematic.

Another possible interpretation of politics, related to the above, is that it is necessary to deal with contractual incompleteness. The written constitutions of modern states were devised by individuals who could not have fully anticipated all contingencies, and accordingly chose to write a mainly procedural constitution, leaving it to the political process to identify the appropriate course of action in each contingency. However, as noted earlier, a self-enforcing contract does not need to provide for all contingencies explicitly, and is indeed consistent with the notion of a formal constitution as establishing basic principles. Furthermore, as we have just noted, in a self-enforcing arrangement, there is no meaningful way of interpreting politics as an ongoing bargaining process that makes up for contractual incompleteness.

Thus, the idea that tax policies reflect a self-enforcing contract calls into question the

[39]Furthermore, from an ex-ante perspective, there would be little reason to coordinate to any arrangement that allows for political competition to drive the ex-post selection of self-enforcing equilibrium, since an arrangement where coordination takes place to the ex-ante optimal equilibrium is also self-enforcing (and ex-ante optimal).

interpretation of politics as a system of procedures that substitute for bargaining in large groups—an interpretation that is common to much of the recent political economy literature.

There is, however, a possible interpretation of political processes—distantly related to the idea of incompleteness—that is fully consistent with the idea of a self-enforcing tax constitution, namely that political competition provides an information-pooling device which enables individuals to coordinate to a "good" equilibrium. With incomplete information, as time goes by individuals may each have the opportunity to make new observations about the state of the world (and particularly about their own private characteristics) obtaining new information that is private to them, so there will be a problem of pooling this new private information for the purpose of Bayesian updating; this will generate scope for people trying to misrepresent what they know in order to push the outcome in directions they favour. In other words, individuals will try to persuade others that they have new compelling information suggesting that the right thing to do, according to the very principles all individuals concur upon, is indeed what they argue should be done. All of this could look a lot like politics. A formal illustration of this idea is given in Appendix B.

We can then think of appointed policymakers as flag-bearers, whose very identity and appointment via a given political process reveals information that would otherwise be private to individual voters, and who are entrusted to produce a single, publicly observable coordinating signal—their announced policy—an announcement that has no autonomous decision force beyond its function of inducing coordination on specific self-enforcing continuation equilibrium. In this interpretation there is a natural separation between the announcement of taxes—which is delegated to policymakers—and their enforcement—which can be delegated to independent government institutions representing the group and applying punishment consistently with collective rationality as prescribed by the particular continuation equilibrium selected.

The general picture that emerges from interpreting politics in this way is still consistent with the observed features of real-world politics, and broadly consistent with the predictions of traditional models of political competition: political candidates will publicly disagree with each other about which policies will be appropriate, individuals will vote for candidates according to their preferences and to the candidates' preferences as they understand them, policymakers will be selected by majority voting and, once elected, will make choices that are as close as possible to their preferred choice. What is different here is the interpretation of politics: political competition is about reconciling different opinions on the true state of the world rather than about reconciling conflicting interests in the definition of policies—which, in principle, can be taken care of ex ante.

6.4 Self-enforcing Contract or Evolved Behaviour?

There is a possible alternative explanation for group cooperation that does completely away with the problem of enforcement: the idea that cooperative behaviour could be innate to individuals within groups and be the result of evolutionary selection (see Bergstrom, 2002, for a survey).[40] The arguments that are presented in the literature combine the idea of evolutionary group selection (whereby more cooperative groups reproduce more successfully as groups) with the idea of individual selection within groups (whereby defectors reproduce more within groups) and show that the tension between these two can result in equilibria where cooperative behaviour prevails (defectors do invade groups but those groups end-up self-destructing, so that eventually only cooperators remain). Similarly, evolutionary arguments have been offered to rationalize individuals' willingness to punish defectors even when punishing is not individually rational.

One key argument against the idea that cooperation in collective consumption can be sustained because of hard-wired, evolutionarily selected behaviour is that there are independent reasons for selfish behaviour to be evolutionarily selected in an environment where the problem of sustaining collective consumption within groups is not the only problem individuals face. According to Robson and Kaplan (2003), individual optimizing behaviour coupled with learning enables individuals to cope better with a changing environment and/or to adopt a hunter/gatherer survival strategy (whereby learnt knowledge about one's environment becomes essential). Moreover, we cannot expect individual rationality with respect to output production and other individual choices to be simply decoupled from individual rationality in collective consumption choices, as it may be impossible for a rational nonopportunist to survive in a world of rational opportunists. Consider, for example, the following variant of the evolutionary models mentioned above. Suppose that individuals interact with each other within groups both according to situations that consist of either positive-sum games (giving rise to a collective choice problem) or zero-sum games (where competition is private in nature), and suppose that the positive-sum or zero-sum nature of the situation is not immediately obvious to individuals. Then, individuals who selectively behave cooperatively in positive-sum situations may be vulnerable to misrepresentation by other individuals, and may therefore fare comparatively worse than purely opportunistic individuals.

Another problem of interpretation arises if we try to apply ideas from the literature on the evolution of cooperation to explain tax constitutions. Much of the literature on evolution and cooperation—although rather abstract and necessarily vague in its discussion of how its constructs map into real-world institutions—refers to some sort of meta-

[40] Other characterizations of cooperative behaviour that have been proposed in the literature in the context of a static collective consumption problem—such as the idea of *warm glow* in voluntary giving (Andreoni, 1990)—can also be broadly related to evolutionary ideas.

game whereby people choose to behave cooperatively or noncooperatively in positive-sum situations. But there is no obvious reason why Nature should restrict itself to evolving cooperation with respect to collective consumption problems; it could also evolve suitable hard-wired behaviour in all sort of economic contexts where selfish behaviour may generate collectively undesirable consequences. So, for example, Nature should also take care of selecting people's behaviour in such a way that they do not respond at all to the presence of taxes in their budget constraints and no not take advantage of any tax avoidance opportunities, or so that, as producers, they price at marginal cost independently of profit maximization considerations (e.g. even if they happen to be monopolists). Thus, if we are prepared to abandon the idea that individuals behave in a selfish, individually rational way in the case of collective consumption, we should be prepared to do so with respect to all possible forms of group interaction—including market interaction.

An even more compelling argument against the idea of evolutionarily selected cooperative behaviour is that, as Robson (2001) notes, opportunistic individuals are able in any event to solve coordination failures through repeated interaction, so long as lifespans are sufficiently long. Then, the group selection advantage of hard-wired cooperators vis-à-vis opportunists would vanish, and the only thing that would remain is the comparative vulnerability of hard-wired cooperators to misrepresentations by opportunist invaders: individually rational, selfish behaviour will thus be selected over hard-wired cooperation.

On the other hand, the idea of self-enforcing tax constitutions that we put forward here is not inconsistent with the idea of evolutionary selection, as long as this takes place in parallel with repeated interaction amongst long-lived individually rational agents. There are, however, ways in which evolution and repeated interaction may interface. For example, in an overlapping generations setup, initial beliefs of newborns about their own characteristics may be systematically biased relative to the actual characteristics of the population in a way that facilitates the support of cooperation without contradicting rationality and learning. And, as noted earlier, under incomplete information efficient outcomes may need to be supported by a specific set of rational, self-sustaining beliefs that may themselves be the result of evolutionary selection—which is not inconsistent with Hayek's (1967) idea that social norms are a product of cultural evolution. Finally, evolutionary ideas could be applied to group rather than individual selection as an alternative to ex-ante utility maximization in order to rationalize equilibrium selection of efficient outcomes under repeated interaction (Boyd and Richerson, 1994). Accordingly, veil-of-ignorance selection of a self-enforcing equilibrium could be given an evolutionary interpretation: to the extent that the expectation of utility coincides with mean fitness, equilibrium selection on the basis of ex-ante utility maximization would be consistent with evolutionary group selection.

7 Conclusion

While the traditional public finance approach to the analysis of taxation has produced very valuable insights, the implicit assumption underlying that approach is that governments have the power to impose any tax they choose to. In that approach enforcement is only a problem in so far as individual attempts to evade taxes must be detected and punished. In contrast with this tradition, the informal debate on taxation has long acknowledged that general compliance depends, in a fundamental way, upon the general consensus of the public—i.e. it would be impossible to enforce compliance if the public did not find the tax system acceptable and were collectively non-compliant.[41] The question of the viability of tax rules—a different and more basic issue than the question of how viable taxes can be enforced—has effectively been assumed away by previous theoretical literature.

Early contractarian philosophers appeared to recognize that laws and institutions rest on the ongoing support of the public. Rousseau (1762), for example, notes that "sovereignty, being nothing less than the exercise of the general will, can never be alienated;" and that "the clauses ... [of the social contract] ... are so determined by the nature of the act that the slightest modification would make them vain and ineffective; so that, although they have perhaps never been formally set forth, they are everywhere the same and everywhere tacitly admitted and recognized."[42] It is only later formalizations of contractarian ideas (such as Wicksell's) that framed them within a theory of enforceable contracts. While contract theory can be useful for exemplifying the key ideas of social contractarianism, it posits an individually rational but irrevocable surrender of power from individuals to government, which in turn entails limited constraints for the structure of tax institutions. The public finance writers in that tradition thus had to look elsewhere for principles that could sharpen their models' predictions.

Our analysis here is a first attempt to provide a theoretical foundation for the idea that tax rules must be viable. In doing so, it shows that viability constraints can alone account for a number of observations that traditional theories are only able to explain by invoking distributional or ethical principles—such as, for example, taxation on the basis of ability to pay. At the same time, it casts some light on a number of questions that traditional theories have been silent on; and it suggests that a full integration of tax theory with a theory of property rights could be a fruitful way forward for advancing our understanding of tax institutions.

[41] A relatively recent example of an unsustainable tax was the poll tax introduced in the UK in the 1980s. This turned out to be unenforceable, which is consistent with our predictions that viable taxation must satisfy certain progressivity requirement.

[42] Such statements from Enlightenment Era writers can be taken as setting out a political manifesto; but they also be read as spelling out the individual and collective incentives that institutional arrangements must satisfy in order to be viable—i.e. institutional arrangements that are consistent with the "natural order" as they saw it.

Appendix A

Nash Equilibrium Contributions with No Repeated Interaction

Equilibrium contributions $c^{Ni} \in [0, y^i]$, are identified by the conditions:

$$\theta^i v'\left(\sum_i c^{iN}\right) - 1 \equiv \omega^i, \qquad c^{iN}\omega^i \geq 0, \qquad (y^i - c^{iN})\omega^i \leq 0, \quad \forall i. \tag{1}$$

Due to the quasilinearity of preferences, the above conditions only identify a total volume of collective consumption $g^N = \sum_i c^{iN}$, but not the distribution of contributions among individuals. Quasilinearity also implies that the efficient level of collective consumption, identified by the conditions

$$\sum_i \theta^i v'(g^*) - 1 \equiv \omega, \qquad g^*\omega \geq 0, \qquad \left(\sum_i y^i - g^*\right)\omega \leq 0, \tag{2}$$

is independent of how provision is funded. Conditions (1) and (2) imply $g^N \leq g^*$ (with $g^N = g^*$ if and only if $g^* = 0$). In the following discussion, we shall assume $g^* > y^i, \forall i$.

Renegotiation Proofness

Let $\mathcal{S}(h^t)$ be the set of subgame perfect equilibrium strategy profiles for the subgame defined by history h^t; in a repeated game $\mathcal{S}(h^t)$ is independent of history, and can be denoted as $\mathcal{S}$. Let R be a subset of $\mathcal{S}$. Then, continuation strategy profiles in R are weakly renegotiation proof (WRP) if the continuation equilibria they prescribe in all subgames are also elements of R and if no element of R induces a payoff profile that is Pareto dominated the payoff profile induced by another element of R (in short, no equilibrium in R is Pareto-dominated by another equilibrium in R). [43] Apart from this condition, the selection of R is arbitrary, which makes weak renegotiation proofness less than compelling as a predictive criterion.

An alternative refinement that does not suffer from the same limitation is strong renegotiation proofness, which is so defined: a WRP equilibrium set R is also a strongly renegotiation-proof (SRP) if there is no other WRP equilibrium set $R' \subseteq \mathcal{S}$ containing equilibria that Pareto dominate equilibria in R'. Identifying SRP equilibria requires a systematic comparison of *all* possible equilibria in $\mathcal{S}$, which makes this criterion impractical in concrete applications. A more manageable version of strong renegotiation proofness can be obtained by exogenously restricting attention to WRP equilibria within a strict subset $\mathcal{S}'$ of $\mathcal{S}$.

What we do in our analysis is identify *restricted* SRP equilibria that are defined within the set of equilibria that involve a single-period punishment phase, and then proceed to investigate whether these are robust with respect to small changes in the structure and length of the punishment—i.e. whether these equilibria remain SRP when we allow for a local relaxation of $\mathcal{S}'$.

[43] It may be tempting to extend this refinement by requiring that continuation equilibria be robust to objections by *any* coalition of players, as in the Coalition-Proof Nash equilibrium refinement of Bernheim, Peleg, and Whinston (1987). That refinement, however, simply guides equilibrium selection, and is not concerned with sequential rationality in dynamic games as is renegotiation proofness. Requiring continuation equilibria to be unobjectionable for any possible coalition of players (including single players) would effectively eliminate the possibility of any form of punishment. Milder extensions, however, have been suggested—see, for example, the notion of a Consistent Bargaining Equilibrium proposed by Abreu, Pearce, and Stacchetti (1993).

Proof of Proposition 1

Consider first a scenario with perfect monitoring. In this case, in addition to the punishment described in the text, the level of collective consumption in the punishment phase may also fall to a level $\underline{g}_i < g$ that is compatible with collective rationality. This may require arbitrarily differentiating contributions amongst nondefectors in such a way that the resulting outcome is not Pareto undominated by the equilibrium outcome, but doing so does not affect incentives for potential defectors beyond its impact on $\underline{g}_i < g$. Sufficient conditions for such punishment strategies to be renegotiation-proof equilibrium strategies are then:

$$(c^i - c^{iD}) - \theta^i \left(v(g) - v(g - c^i + c^{iD})\right) \leq \delta \left(y^i - c^i + \theta^i \left(v(g) - v(\underline{g}_i)\right)\right), \quad \forall i, \tag{3}$$

$$(y^i - c^{iR}) - \theta^i \left(v(\underline{g}_i) - v(\underline{g}_i - y^i + c^{iR})\right) \leq \delta \left(y^i - c^i + \theta^i \left(v(g) - v(\underline{g}_i)\right)\right), \quad \forall i, \tag{4}$$

$$\sum_{j \neq i} \theta^i \left(v(\underline{g}_i) - v(g)\right) + (g - \underline{g}_i + y^i - c^i) \geq 0, \quad \forall i, \tag{5}$$

where δ is the one-period time discount factor—the same for all individuals—and where c^{iD} and c^{iR} denote the one-shot unilaterally optimum levels of contribution for player i respectively when the other players contribute levels c^j, $j \neq i$, and when the other players contribute the lower levels $c^j - (g - \underline{g}_i + y^i - c^i)/(n-1)$, $j \neq i$; these are respectively identified by

$$\theta^i v' \left(g - c^i + c^{iD}\right) - 1 \equiv \omega^{iD}, \qquad c^{iD}\omega^{iD} \geq 0, \qquad (y^i - c^{iD})\omega^{iD} \leq 0, \quad \forall i; \tag{6}$$

$$\theta^i v' \left(\underline{g}_i - y^i + c^{iR}\right) - 1 \equiv \omega^{iR}, \qquad c^{iR}\omega^{iR} \geq 0, \qquad (y^i - c^{iR})\omega^{iR} \leq 0, \quad \forall i. \tag{7}$$

In a large economy, both the defection and the no-repentance contribution levels, c^{iD} and c^{iR}, will both be zero, and (3)-(5) are simplified accordingly .

Condition (3) states that the maximum gain for player i in period t from adopting an alternative strategy must be less than the present value of the loss from having to increase her contribution in period $t+1$ (no defection). Condition (4) states that in period $t+1$, following defection in period t, player i wishes to secure reversion to cooperation from $t+2$ onwards by paying a higher contribution in period $t+1$ rather than postponing repentance and contributing her unilateral optimum level (no postponement of repentance). The left-hand side of condition (5) represents the gain to players other than player i from carrying out the punishment, and is trivially satisfied. This latter condition is what makes such an equilibrium renegotiation-proof (collective rationality of punishment). If the punishment is carried out and the defector repents, the other players benefit from a reduction in the contributions required from them in period $t+1$ and at the same time the level of collective consumption is collectively optimal for them. This implies renegotiation proofness, given that the punishers cannot be made better off by increasing provision above $\underline{g}_i$—which is collectively (constrained) optimal for them—and given that there is no scope for further compensation from the defector—who is already directing all of her income to collective consumption.

The level of provision $\underline{g}_i \leq g$ must be supported, at $t+1$, by analogous punishment strategies against any individual, other than i, who, at $t+1$, reduces her contribution below the contribution $c_i^j \leq c^j$ required from her during the punishment phase. This can always be achieved, since the contributions of individuals other than the defector in the punishment phase are less than their equilibrium contributions, and so is their temptation to defect from the stated course of punishment.

However, if monitoring is imperfect, it can be shown that ex-ante efficiency in a large economy requires $\underline{g}_i = g$, as long as the probability of mistakes is positive. Let c^i denote i's intended equilibrium contribution level in an equilibrium with $\underline{g}_i = g$, and let μ be the probability of a reduction in an individual's contributions being observed in an equilibrium where no player, i, ever actually deviates from an intended c^i. If a collective consumption level $\underline{g}_i < g$ were to be adopted as part of the punishment against perceived defections, the sustainable intended contribution level could be increased to $\bar{c}^i > c^i$. Provided punishment were never triggered, this would produce an increase in g and a corresponding ex-ante gain. However, in a large economy, and as long as $\mu > 0$, punishment will be triggered in each period with a probability that approaches unity as n become large, implying that lowering $\underline{g}_i$ below g will reduce ex-ante utility.

For $\underline{g}_i = g$, the left-hand side of (4) is greater than or equal to the left-hand side of (3), and therefore the latter inequality is always satisfied if the former is. What this means is that in a large economy the sustainable level of contributions will be identified by (4).[44]

The remainder of the proof is devoted to showing that requiring a payment of y^i during the punishment phase is consistent with ex-ante constrained optimality.

Ex-ante constrained efficiency requires making a defector's contribution in the punishment phase (the fine) as high as possible, i.e. equal to her income, y^i (as implied by (4)), because this maximizes the sustainable level of contribution. This can be most easily seen by focusing on the symmetric case with $y^i = y$, $\theta^i = \theta$, $\forall i$. Denote the required payment in the punishment phase as f, with $y \geq f > c$. With n large we can write (4) as $f - \theta(v(g) - v(g-f)) - \delta(f-c) \equiv \Psi \leq 0$. The expression Ψ is increasing in c and is greater than $c - \theta(v(g) - v(g-c) - \delta(f-c)$ (since $f > c$ and v is concave), implying that the maximum sustainable contribution will be identified by the condition $\Psi = 0$. Note that $\Psi = 0$ implies $\theta(v(g) - v(g-f))/f - (1-\delta) = \delta(c/f) > 0$, and therefore $\theta v'(g-f) - (1-\delta) > 0$ (since $v'(g-f) > (v(g) - v(g-f))/f$). Totally differentiating the condition $\Psi = 0$, we obtain

$$\frac{\partial c}{\partial f} = \frac{\theta v'(g-f) - (1-\delta)}{n\theta\left(v'(g-f) - v'(g)\right) + \delta} > 0. \tag{8}$$

Thus c is maximized by raising f to its upper bound, y.[45] Notice that adding more punishment after period $t+1$ does not make punishment more effective, as the following experiment shows. Consider an alternative punishment strategy in which the increase in the contribution required

[44]One could conjecture that, since (3) is slack, it may be possible for equilibrium contributions to exceed the level of contributions to which defectors revert following repentance. (Note that raising the reversion contribution level after period $t+2$ above c^i is not possible unless the reversion contribution level at $t+1$ is reduced below c^i; in other words, c^i, as identified by (4) represents an upper bound on the present value of discounted reversion contributions following a defection.) If, however, contribution levels $\bar{c}^i > c^i$ were part of a renegotiation-proof equilibrium, then this equilibrium would Pareto dominate the supporting continuation equilibria which are triggered by defections and which need to be invoked in support of such an equilibrium—a contradiction.

[45]Renegotiation-proof equilibria that are not ex-ante constrained-efficient and involve a payment $f^i < y^i$ by the defector in the punishment phase may be possible. Suppose, for example, that all individuals have the same preferences but one-half of the population has an income of zero and the other half an income y'. Denote with $g^*(n/2)$ the level of collective consumption that is ex-ante efficient for a group of $n/2$ individuals. Then an equilibrium with $g = g^*(n/2)$ and $f' < y'$ for all individuals with income y' could be a renegotiation-proof equilibrium even if there exist another renegotiation-proof equilibrium with $f' = y'$ and a level of collective consumption above $g^*(n/2)$ (but still short of g^*). Thus, renegotiation-proof punishment strategies have the same structure even if we do not invoke ex-ante utility maximization as a criterion for equilibrium selection.

by the defector at $t+1$ is accompanied by an increase in period $t+2$ by an amount $z \leq y^i - c^i$, with reversion to normal play starting in period $t+3$ rather than $t+2$. This would increase the punishment for deviations (the right-hand side of (3)) by $\delta^2 z$; the right-hand side of (4), however, would decrease by $(\delta^2 - \delta)z$. Then, since (4) always becomes binding before (3) does when $z = 0$, an increase in z would either have no effect or cause (4) to bind. Furthermore, as long as δ is sufficiently close to one, concentrating punishment in a single period makes it possible to support the highest contribution. To see this, suppose that the fine in period $t+1$ is reduced from y to $y - z$ but a fine γz is also applied in period $t+2$, with $\gamma > 0$. Then, as we increase z, punishment becomes more spread out and the total punishment will increase or decrease depending on the size of γ. The repentance constraint is then $y - z - \theta v(g) + \theta v(g - y + z) \leq \delta(y - c) + \delta^2 \gamma z - \delta(1 + \gamma)z$. Consider next an increase in z starting from $z = 0$. The repentance constraint that refers to period $t+2$ will be slack (the one-shot gain γz for noncompliance is less than $y - z$ and punishment is greater than that for no-repentance at $t+1$ since we can make punishment restart). Totally differentiating the repentance constraint for $t+1$, we obtain $\partial c / \partial z = (1 - \theta v'(nc) + \delta^2 \gamma - \delta(1 + \gamma))/(n\theta(v'(g - y) - v'(g)) + \delta)$: the denominator is positive, and, for δ approaching unity, the numerator approaches $-\theta v'(nc) < 0$ for any γ, implying that two-period punishment will be outperformed by single-period punishment.

Proof of Proposition 2

A contribution level $c^i = y^i$ cannot be supported, since (4) could not be satisfied in this case. If we differentiate (4)—for $\underline{g}_i = g$ and $c^{iR} = 0$—with respect to c and y, we obtain

$$\frac{\partial c}{\partial y} = \frac{\delta - 1 + \theta v'(g - y)}{\delta - n\theta \left(v'(g) - v'(g - y) \right)}. \tag{9}$$

Since $(\delta - 1)y + v(g) - v(g - y) = \delta c \geq 0$ in equilibrium, we must have $(v(g) - v(g - y))/y > (1 - \delta)$ and therefore $v'(g - y) > 1 - \delta$ (since $v'' < 0$), which implies that (9) is positive.

Proof of Proposition 3

Totally differentiating (4) with respect to n, with $g = nc$, $\underline{g}_i = g$, and $c^{iR} = 0$, we obtain

$$\left(\frac{\partial g}{\partial n} \right)_{y \text{ constant}} = \frac{\delta c}{n\theta \left(v'(g - y) - v'(g) \right) + \delta} > 0. \tag{10}$$

On the other hand, holding aggregate income constant, the sustainable level of contributions that can be sustained is decreasing in n, but asymptotically approaches a level that exceeds the one-shot noncooperative level g^N. Letting $y = Y/n$, with Y constant, and totally differentiating (4) with respect to n, we obtain

$$\left(\frac{\partial g}{\partial n} \right)_{ny \text{ constant}} = \frac{\delta c - \left(\theta v'(g - y) + (1 - \delta) \right) y}{n\theta \left(v'(g - y) - v'(g) \right) + \delta}. \tag{11}$$

Since $(1 - \delta)y - \theta(v(g) - v(g - y)) + \delta c = 0$ at an optimum and $\theta v'(g - y)y > \theta(v(g) - v(g - y))$, expression (11) is negative. However, as n approaches infinity, collective consumption approaches the level $\hat{g} > 0$ for which $\theta v'(\hat{g}) = 1 - \delta(1 - \hat{g}/Y) < 1$, or zero if $\theta v'(0) < 1 - \delta$; whereas the one-shot noncooperative level of collective consumption is either equal to a positive level identified by the condition $\theta v'(g^N) = 1$, or zero if $\theta v'(0) < 1$. All of this implies $\hat{g} \geq g^N$, with the inequality being strict if $\theta v'(0) > 1 - \delta$.

Proof of Proposition 4

Suppose we totally differentiate (4)—for $\underline{g}_i = g$ and $c^{iR} = 0$—with respect to c^i and y^i, holding the total level contributions constant. Holding g constant implies holding the distribution of income constant as well as its aggregate level; this experiment thus amounts to comparing the equilibrium contributions of two individuals with different but arbitrarily close income levels. We obtain

$$\left(\frac{\partial c^i}{\partial y^i}\right)_{g \text{ constant}} = \frac{\delta - 1 + \theta v'(g - y^i)}{\delta} > 0, \qquad \forall i. \tag{12}$$

Proof of Proposition 5

Further differentiating (12) with respect to y^i, we obtain

$$\left(\frac{\partial^2 c^i}{\partial (y^i)^2}\right)_{g \text{ constant}} = -\frac{\theta v''(g - y^i)}{\delta} > 0, \qquad \forall i. \tag{13}$$

Proof of Proposition 6

Totally differentiating (4), we obtain

$$\left(\frac{\partial c^i}{\partial \theta^i}\right)_{g \text{ constant}} = \frac{v(g) - v(g - y)}{\delta} > 0, \qquad \forall i. \tag{14}$$

The elasticity of c^i with respect to changes in θ^i is $\theta^i \left(v(g) - v(g-y)\right)/(\delta c^i)$, which can be rewritten as $(y^i/(\delta c^i))\theta^i \left(v(g) - v(g-y)\right)/y^i$, which is less than $(y^i/(\delta c^i))\theta^i v'(g - y^i)$, representing the elasticity of c^i with respect to changes in y^i.

Proof of Proposition 7

In the following, it will be notationally convenient to order individuals by the size of their potential income, i.e. $y^{i+1} \geq y^i, \forall i$.

Suppose that voluntary transfers are zero in equilibrium (the case with non-zero transfers is dealt with in Proposition 9). If player i attempts expropriation, the highest gross deviation gain she can obtain is by directing her expropriation attempts against the $n'_{-i} \leq \bar{n}$ highest income individuals—other than herself—for which $\beta(1)y^j \geq \gamma y^i$, $j \geq n - n'_{-i}$. Denote the set of such individuals as M_{-i}; the net gain, $\Lambda^i \geq 0$, that can be obtained from these unilateral expropriation attempts is

$$\Lambda^i \equiv \sum_{j \in M_{-i}} \max\{\beta(1)y^j - \gamma y^i, 0\}. \tag{15}$$

Λ^i will be decreasing in y^i (higher income individuals are less tempted to expropriate others), and may reach zero at a certain income level (remaining at zero thereafter).

The set of players, $P_i(s_i)$, participating in expropriation of the defector in the punishment phase consists of the largest possible number, $\bar{s}^i$, of lowest-income individuals, other than player i, who can collectively profit from carrying out the punishment: $\bar{s}^i$ the largest s_i for which

$$\beta(s_i)y^i - \gamma \sum_{j \in P^i(s_i)} y^j \geq 0. \tag{16}$$

Let $\bar{\beta}_i \equiv \beta(\bar{s}^i)$. Then, assuming that $\rho/\gamma < n$ (which will always be the case with n large), the punishment strategy described will prevent unilateral expropriation attempts if

$$\Lambda^i \leq \delta \bar{\beta}_i y^i, \quad \forall i. \tag{17}$$

Condition (17) ensures that the gain to player i from deviating and/or from postponing repentance after deviation is less than the present value of the punishment.[46]

In order to show that (17) and (16) define a renegotiation-proof equilibrium, we can proceed as follows. Condition (16) says that expropriating the defector must produce a net benefit for the punishers in comparison with the no-expropriation path. As long as this is the case, at least one of the expropriators will be better off carrying out the punishment in comparison with any alternative no-punishment continuation equilibrium, i.e. the punishers cannot find a combination of transfers which, absent punishment, makes all of them better off relative to carrying out the punishment. Suppose, however, that condition (17) is satisfied with strict inequality. Then the slack could be exploited to induce individual i to make a transfer in addition to not expropriating others; which means that, following defection by i, there could exist an alternative continuation equilibrium where i avoids punishment and compensates the punishers with a transfer. Thus, a punishment expropriation level $\bar{\beta}_i$ is only consistent with renegotiation proofness if (17) is met with equality, otherwise a lower β_i would have to be invoked. Nevertheless, if (17) is slack and (16) binding, it will always be possible to find such β_i (given that the right-hand side of (16) is decreasing in s_i); so (17) is necessary and sufficient for characterizing a no-expropriation renegotiation-proof equilibrium.[47]

Proof of Proposition 8

In a large economy, where the number of individuals of each income type is greater than both ρ/γ and s_i for any possible value of s_i satisfying (16), we have $\Lambda^i = (\rho/\gamma)\max\{\beta(1)y^n - \gamma y^i, 0\}, \forall i$, and $y^i = y^1, i \leq s_i$. Moreover, $\beta''(s) < 0$ implies that $\beta(s)/s$ is decreasing in s, and therefore $\bar{s}^i$ is increasing in y^i. This means that the difference between the right- and the left-hand sides of (17) will always be smallest for the individuals with the smallest income, i.e. for $i = 1$. Then a sufficient condition for a no-expropriation outcome to be supported in equilibrium is that (17) be

[46]Here, we are implicitly assuming that $\gamma n'_{-i} < \bar{\beta}_i$, which implies that the actual income produced by a defector who fails to repent never falls short of $\bar{\beta}_i y^i$, the amount targeted for expropriation during punishment. If this condition were to be violated, then the repentance condition would also have to include the term $\max\{(\bar{\beta}_i - \gamma n'_{-i})y^i, 0\}$. In this case, maximizing the expropriation that is carried out against a defector would have the effect of reducing the opportunity cost of expropriation to the punishers. Hence, the number of punishers, s_i, in an efficient, renegotiation-proof punishment would also need to satisfy the condition $\bar{\beta}_i \leq \gamma n'_{-i}$. Also, note that punishment strategies so described can only be credible (subgame perfect) as part of a noncooperative equilibrium where no deviations occur, and where therefore the only potential deviations we need to consider are by single individuals. Given the technological constraints associated with expropriation, a joint deviation by a subset of individuals could not actually be punished in this way.

[47]If we consider multiperiod punishment, it can be shown that, for δ sufficiently large, there will exist renegotiation-proof punishment strategies that involve indefinite payment of a fine and dominate single-period expropriation. These continuation equilibria, however, imply a permanent switch to an asymmetric outcome, and can be ruled out by symmetry-preserving refinements such as the Consistent Bargaining Equilibrium refinement proposed by Abreu, Pearce, and Stacchetti (1993).

satisfied for $i = 1$. This condition can be rewritten as

$$\delta > \frac{\rho}{\bar{\beta}_1}\left(\frac{\beta(1)}{\gamma}\frac{y^n}{y^1} - 1\right). \tag{18}$$

This says that property rights are harder to establish the larger is $\beta(1)\rho/\gamma = \beta(1)\bar{n}$ relative to $\bar{\beta}_1 > \beta(1)$, and the larger is the gap between mean income and lowest income.

Proof of Proposition 9

Let $r^{ij} \geq 0$ denote the transfer from i to j in equilibrium. Condition (17) becomes

$$\Lambda^i + \sum_j r^{ij} \leq \delta\left(\bar{\beta}_i y^i + \sum_j r^{ji}\right), \quad \forall i. \tag{19}$$

A simultaneous increase in the transfer made and received by an individual increases the deviation gain relative to the punishment for that individual; we can thus restrict our attention to equilibria where individuals either receive transfers from others or make transfers to other, i.e. where $\sum_j r^{ij} \sum_j r^{ji} = 0, \forall i$. Let c^i denote the aggregate net transfer, made or received by individual i: $c^i = \sum_j r^{ij} - \sum_j r^{ji}$. Then, for a large a large economy, (19) can be rewritten as

$$\frac{\beta(1)\rho}{\gamma}y^n - \rho y^i + \max\{c^i, 0\} \leq \delta\left(\bar{\beta}_i y^i + \max\{-c^i, 0\}\right), \quad \forall i. \tag{20}$$

The net transfer c^i enters (20) asymmetrically, depending on its sign. This follows from the fact that, while a positive c^i (a positive transfer by i) is the result of an action by i (and therefore belongs to i's strategy space), a negative c^i (a transfer received by i) is the result of others' actions; hence c^i goes to zero following a deviation by i only if it is positive, and goes to zero as part of i's punishment only if it is negative. A combination of $c^i, \forall i$, with $\sum c^i = 0$, satisfying (20) for all individuals may or may not exist.

Proof of Proposition 10

Linked punishment will always be used in a full-information constrained-efficient equilibrium: deviation in contributions and expropriation choices, whether occurring in isolation from each other or jointly, will be best prevented by relying on the strongest available punishment. Given this, the deviations that we need to consider—the deviations that generate the largest one-shot gains for deviators—are joint deviations.

In a large economy, the conditions for the punishment strategies described in the text to be renegotiation-proof are

$$\max\{c^i, 0\} - \theta^i\left(v(g) - v(g - \max\{c^i, 0\})\right) + \Lambda^i - \delta\left((\beta_i + \eta_i(1-\beta_i))y^i - c^i\right) \leq 0, \tag{21}$$

(no defection);

$$\eta_i(1-\beta_i)y^i - \theta^i\left(v(g) - v(g - \eta_i(1-\beta_i)y^i)\right) + \Lambda^i - \delta\left((\beta_i + \eta_i(1-\beta_i))y^i - c^i\right) \leq 0, \tag{22}$$

(no postponement of repentance);

$$\beta_i y^i - \gamma \sum_{j \square s_i, j \neq i} y^j \geq 0, \tag{23}$$

(collectively rational punishment).

The levels of β_i and η_i in a constrained-efficient, renegotiation-proof equilibrium are determined as follows. If the defector were to voluntarily pay the whole penalty, y^i, the group would manage to save the costs associated with forcible expropriation; so it would be possible, in principle, to find an alternative continuation equilibrium at $t+1$ that Pareto dominates the one where punishment relies on expropriation. Nevertheless, if reduced reliance on expropriation increases the temptation to postpone repentance, it will reduce the contribution to which the defector can sustainability revert after punishment; for this reason, a continuation equilibrium that relies on expropriation is not Pareto dominated by one that does not.[48] Thus, equilibrium punishment strategies involving expropriation at $t+1$ are renegotiation-proof (even if they entail a deadweight loss). And if expropriation makes it possible to sustain higher levels of equilibrium contributions, it will be relied upon by constrained-efficient punishment strategies. On the other hand, if for a certain value $\beta_i{}' < \bar{\beta}_i$ the no-defection condition (21) also becomes binding, then raising β_i above $\beta_i{}'$ cannot be part of a renegotiation-proof equilibrium, since in that case it would be possible to recoordinate to an alternative equilibrium featuring less expropriation (given that in that case (21) is the binding constraint) and no less collective consumption: if $\eta_i = 1$, (21) is independent of β_i, and if $\eta_i < 1$ it is possible to lower β_i and increase η_i so as to relax (21).

Then, the maximum supportable tax for a certain individual i in a constrained-efficient, renegotiation-proof equilibrium is identified by one of three possible regimes: (i) the repentance constraint is the only binding constraint, the maximum collectively rational level of expropriation is applied against i in the punishment phase ($\beta_i = \bar{\beta}_i$), and i voluntarily pays all of her residual income in the punishment phase ($\eta_i = 1$); (ii) both the no-defection and the repentance constraints are binding, the maximum collectively rational level of expropriation is applied against i in the punishment phase ($\beta_i = \bar{\beta}_i$), and i voluntarily pays a fraction of her residual income in the punishment phase ($\eta_i < 1$); (iii) both the no-defection and the repentance constraints are binding, the level of expropriation applied against i in the punishment phase falls short of its maximum level ($\beta_i < \bar{\beta}_i$), and i voluntarily pays a fraction of her residual income in the punishment phase ($\eta_i < 1$).

Regime (i): If, for β_i at its maximum value $\bar{\beta}_i$, (22) remains the binding constraint, applying the same arguments used in Section 2 we conclude that the effect of a higher η_i on the repentance condition is ambiguous. Then, if a decrease in η_i tightens the second constraint, η_i will be set at its upper bound $\eta_i = 1$. In this case (regime (i)) the maximum sustainable level of taxation is identified by the condition

$$(1-\beta_i)y^i - \theta^i \left(v(g) - v(g-(1-\beta_i)y^i) \right) + \Lambda^i - \delta(y^i - c^i) = 0. \tag{24}$$

Regime (ii): In a situation where (22) is binding and a decrease in η_i from $\eta_i = 1$ relaxes (22) (and thus raises c^i), the tax will be maximized by a value η_i below unity. However, given that $v'(g-c^i) < 1$ and $v'(g-\eta_i(1-\beta_i)y^i) < 1$ (by assumption), for (22) to be binding when (21) is slack it must be the case that $c^i < \eta_i(1-\beta_i)y^i$. Thus, η_i can only decrease to the point where $\eta_i(1-\beta_i)y^i = c^i$ and both constraints are binding. In such a regime, with $\beta_i = \bar{\beta}_i$, and $\eta_i < 1$, the overall level of payment by a defector during the punishment phase will be less than y^i; the

[48] Suppose that expropriation does not take place at $t+1$; after giving an amount ϵ to the defector, the punishers still get y^i and save the expropriation cost; the expropriator's temptation not to repent is now increased by $\beta_i y^i$ and so the tax that can be supported at $t+1$ is reduced by $\beta_i y^i/\delta$, whose present value to nondefectors is $\beta_i y^i > \gamma \sum_{j \square s_j, s_j \neq i} y^j$ (since by assumption expropriation is profitable).

maximum sustainable level of taxation in this case (regime (ii)) is, for $c^i \geq 0$, identified by (21) in conjunction with the condition $\eta_i(1-\beta_i)y^i = \max\{c^i, 0\}$, which gives

$$\max\{c^i, 0\} - \theta^i \Big(v(g) - v(g - \max\{c^i, 0\})\Big) + \Lambda^i - \delta\beta_i y^i = 0. \tag{25}$$

A corner value $\eta_i = 0$ can only occur in this case if $c^i \leq 0$.

Regime (iii): It is not possible to have an interior solution where $c^i/((1-\beta_i)y^i) < \eta_i < 1$, i.e. where only (22) is binding. This can be shown by noting that, since the second derivative of the left-hand side of (22) with respect to η_i, is equal to $\theta^i v''(\min\{g, g_i^*\} - \eta_i(1-\beta_i)y^i) < 0$, if the first derivative with respect to η_i at $\eta_i = 1$ is positive, it remains positive for all values of η_i below unity. In a situation where $\beta_i = (\beta_i)' < \bar{\beta}_i$ and both constraints are binding, if $\eta_i < 1$, it is always possible to simultaneously increase both β_i and η_i so as to relax both constraints (for example, by increasing η_y and β_i in such a way as to hold $\eta_i(1-\beta_i)$ constant.) Hence, the tax cannot be at a maximum in a situation where η_i and β_i both lie below their respective upper bounds and both constraints are binding. If $\eta_i = 1$ reaches its upper bound first, with $\beta_i < \bar{\beta}_i$ and both constraints binding, the overall payment of the defector in the punishment phase is y^i, and expropriation is not used to its fully feasible extent. The maximum sustainable tax in this case (regime (iii)) is identified by the condition

$$\max\{c^i, 0\} - \theta^i \Big(v(g) - v(g - \max\{c^i, 0\})\Big) + \Lambda^i - \delta(y^i - c^i) = 0, \tag{26}$$

Regime (i), where β_i and η_i are at their respective upper bounds, is analogous to the no-expropriation regime of Section 2, in that the sustainable tax is identified here uniquely by the repentance constraint; the overall level of payment by a defector during the punishment phase will also in this case be y^i, but expropriation will be relied upon as a partial substitute for voluntary fines, as it relaxes the repentance constraint and raises the sustainable tax. In regime (ii) expropriation substitutes for voluntary fines, which are not used to their full available extent. Regime (iii) represents a situation where the ability of the group to credibly punish defectors with expropriation exceeds what can be plausibly promised, given that applying maximum expropriation would violate collective rationality (since it would be unnecessarily wasteful). In this latter regime, expropriation effectively makes the repentance constraint irrelevant; it is as though, in a no-expropriation scenario, the no-defection constraint were the binding one (given that β_i and η_i do not effectively appear in (21)).

The above discussion has simply assumed that g remains unchanged during punishment. In the presence of imperfect monitoring, this can be rationalized by the same argument given in the proof of Proposition 1. Note that the same argument does not rule out the use of expropriation in the punishment phase. In a large economy with imperfect monitoring, punishment will be triggered with probability approaching one, and punishment by expropriation will thus entail some ex-ante cost (given that it dissipates real resources). However, unlike punishment by $\underline{g}_i < g$, punishment by expropriation can raise the expected level of g in equilibrium, which can more than offset the expected social cost of expropriation.

Proof of Proposition 11

First note that $\bar{\beta}$ will be invariant with respect to an increase in n and to a decrease in per-capita income. Regime (i) can never be optimal in a large economy as long as $\beta > 1 - \delta$. In regime (ii), for n large, g/Y approaches $\delta\bar{\beta}$. In regime (iii), for n large, g/Y approaches $\delta/(1+\delta)$.

Proof of Proposition 12

We can prove this claim by way of example. Consider a scenario with $y^i = y, \theta^i = \theta, \forall i$, n large, and β at its upper bound in regime (i). Then, (22) can be written as

$$\eta(1-\beta)y - \theta\left(v\left(g\right) - v\left(g-(1-\beta)y\right)\right) + (\beta(1)\rho/\gamma - 1)y \leq \delta((\beta + \eta(1-\beta))y - g/n), \quad (27)$$

whereas the corresponding condition without expropriation is

$$y - \theta\left(v\left(g\right) - v\left(g-y\right)\right) \leq \delta(y - g/n). \quad (28)$$

For the left-hand side of (27) to be less than the left-hand side of (28) we must have

$$(\beta(1)\rho/\gamma - 1 - \beta)y - \theta\left(v\left(g-(1-\beta)y\right) - v\left(g-y\right)\right) < 0, \quad (29)$$

which is always satisfied if (18) is satisfied: if all individuals have the same income, (18) gives $\delta > 1/\left(\beta\left((\beta(1)\rho/\gamma) - 1\right)\right)$; for $\delta < 1$ this requires $\beta(1)\rho/\gamma - 1 - \beta < 0$. Proceeding in the same way for regimes (ii) and (iii), we find that if (18) is satisfied, the sustainable level of collective consumption is higher with income expropriation than without it.

Proof of Proposition 13

This can be shown by considering the following example. Let us go back to the scenario introduced earlier, where one-half of the population has income y' and the other half have an income of zero, and suppose that preferences are common across consumers ($\theta^i = \theta, \forall i$). Suppose also that $1 > \beta(1)\rho/\gamma > \delta^2\beta(s-1)$, and therefore (as discussed earlier), a no-expropriation equilibrium cannot be supported by the use of transfers in the absence of collective consumption. The level of collective consumption provision which maximizes aggregate (or expected) utility in this case is $g^*(n)$ s.t. $n\theta v'(g^*(n)) = 1$, but the minimum value of collective consumption that is consistent with Pareto efficiency (given the constraint that zero-income individuals cannot contribute to collective consumption or otherwise compensate contributors) is $g^*(n/2) < g^*(n)$ s.t. $(n/2)\theta v'(g^*(n/2)) = 1$. Consider an equilibrium with $g = g^*(n)$; then, the punishment to a zero-income individual for attempting expropriation consists of a temporary fall from $g^*(n)$ to $g^*(n/2)$ in the level collective consumption experienced. Then, in order for a zero-income individual to be persuaded not to expropriate, we must have that $(\beta(1)/\gamma)y' \leq \delta\theta(v(g^*(n)) - v(g^*(n/2)))$, and for a high-income individual to be persuaded to contribute $2g^*(n)/n$, we must have $2g^*(n)/n - \theta(v(g^*(n)) - v(g^*(n) - 2g^*(n)/n)) \leq \delta\left(y' - 2g^*(n)/n + \theta(v(g^*(n)) - v(g^*(n/2)))\right)$. Suppose the former condition is satisfied with equality. Then, for n large, the latter condition can be rewritten as $\beta(1)/\gamma > (1 + \delta - \theta v'(g^*(n)))(g^*(n)/(ny'/2)) - \delta$. One can verify that this condition can be satisfied even when $1 > \beta(1)/\gamma > \delta^2\beta(s-1)$, implying that taxes that are used to fund collective consumption can manage to prevent expropriation from occurring in situations where direct transfers cannot.

Proof of Proposition 14

Suppose, for example, that all individuals have the same income but there are $\underline{n}$ individuals with $\underline{\theta}$ and $\overline{n}$ individuals with $\overline{\theta} > \underline{\theta}$. Also suppose that individuals are not tempted to expropriate others ($\beta(1) = 0$) and group expropriation possibilities are not binding (regime (iii) in the previous section). Then a renegotiation-proof equilibrium such as the one described under (a), featuring a combination of contributions supported by the threat of expropriation, i.e. taxes, and "voluntary" contributions is identified by the conditions:

$$\underline{c} - \underline{\theta}(v(g) - v(g-y)) = \delta(y - \underline{c}); \quad (30)$$

$$y - \overline{\theta}(v(g) - v(g-y)) = \delta(y - \overline{c}); \tag{31}$$

$$g = \underline{nc} + \overline{nc}. \tag{32}$$

One can verify that for the above system of equations to admit a solution with $\overline{c} > \underline{c} \geq 0$, we must have $\overline{\theta} > (\delta/(1-\delta))\underline{\theta}$, i.e. $\overline{\theta}$ must be sufficiently larger than $\underline{\theta}$; hence, if preference types are similar, there is no scope for such a mixed equilibrium to arise.

The alternative equilibrium where only the $n/2$ higher-preference individuals pay taxes (labeled as (b) above), will be identified by the conditions

$$y - \overline{\theta}(v(g) - v(g-y)) = \delta(y - \overline{c}'); \tag{33}$$

$$g = \overline{nc}'. \tag{34}$$

This equilibrium can dominate or be dominated by the previous one depending on the preference parameters and on the relative number of high- and low-preference individuals. If $\underline{n}/\overline{n}$ is sufficiently small, the additional revenues that can be extracted from high-preference individuals by targeting them with stronger punishment dominate the loss of revenues from low-preference individuals, and the second type of equilibrium will therefore dominate the first. If $\underline{n}/\overline{n}$ is sufficiently large, on the other hand, then the first type of equilibrium will dominate the second.

Proof of Proposition 15

Consider a scenario with no collective consumption and suppose that there are two groups of individuals, $\underline{n}$ lower income individuals with income $\underline{y}$ and $\bar{n}$ lower income individuals with income $\bar{y}$. Higher income individuals face the same cost, $\bar{\gamma}$, of attempting expropriation, whereas lower income individuals, differ from each other in that they have different "expropriation strengths", reflected in different $\underline{\gamma}^i$s, and are ordered so that $\underline{\gamma}^i \leq \underline{\gamma}^{i+1} \leq \bar{\gamma}$. Suppose first that jailing is not feasible and that $\underline{\Lambda}^i > 0$ for all low-income individuals and $\bar{\Lambda} < 0$ for high-income individuals. The minimum transfers, $\underline{c}^i$, required to prevent low-income individuals from attempting expropriation are identified by the conditions $(\beta(1)\rho/\gamma^i)\bar{y} - \rho y^i = \delta((1-\underline{\beta})y^i - \underline{c}^i)$. The maximum transfer, $\bar{c}$, sustainable by high income individuals is identified by $\bar{c} = \delta\bar{\beta}\bar{y}$. If $\bar{n}\bar{c} + \sum_{i \square \underline{n}} \underline{c}^i < 0$, it will not be possible to prevent all individuals from attempting expropriation.

Now suppose that jailing is feasible and let q be the cost of jailing an individual in each period. Also assume that the cost of jailing is less than the amount expropriated by an offender, i.e. $q < (\beta(1)\rho/\gamma)\bar{y}$. Then, if it is not possible to prevent expropriation through transfers, a renegotiation-proof equilibrium under full information may involve pre-emptive jailing, which must be financed by viable contributions that are in turn supported by the threat of collective expropriation. Let i' be the minimum i for which $\bar{n}\bar{c} + \sum_{i>i'} \underline{c}^i - i'q \geq 0$. Then, under full information, it will be feasible, and collectively rational, and ex-ante efficient to pre-emptively jail all individuals $i \leq i'$.

Suppose next that γ^i is private information, and that individuals that are jailed receive a certain known payoff $\underline{u}$ which is less than the payoff they obtain by abstaining from expropriating others. Any individual who attempts expropriation is jailed indefinitely in all subsequent periods. Individuals who are jailed remain anonymous (so that their past compliance history beyond their more recent action does not directly reveal their type). Since γ^i is private information, transfers for low-income individuals must also be anonymous, i.e. equal to a common value $\underline{c}$ for all i. The common transfer required to prevent all low-income individuals from offending in the absence of jail will then be equal to $\underline{c}^1$ as defined above (the full-information transfer required for the

individual with the lowest γ^i). If $\bar{n}\bar{c} + \underline{nc}^1 < 0$, it will not be possible to prevent expropriation by transfers. If a low-income individual, i, anticipates that all expropriators face jail and faces a given transfer $\underline{c}$, attempting expropriation will be profitable if

$$\max\{\underline{c}, 0\} + ((\beta(1)\rho/\gamma^i)\bar{y} - \rho\underline{y} \geq (\delta/(1-\delta))(\underline{y} - \underline{c} - \underline{u}). \tag{35}$$

For a given $\underline{c}$, this identifies a borderline offender i'', such that individuals $i \leq i''$ offend and individuals $i > i''$ do not. Given that the transfer for the individual just above individual i'' must also be equal to $\underline{c}$, requiring a positive transfer by low income individuals cannot be part of a renegotiation-proof equilibrium, because lowering $\underline{c}$ to zero will reduce i''. On the other hand, positive contributions may be sustained by higher income individuals. Then if $\bar{n}\bar{c} + (\underline{n} - i')\underline{c} - i'q \geq 0$ jailing can be financed by sustainable contributions. And given that $\bar{n}\bar{c} + \underline{nc}^1 < 0$, it will not be possible to prevent further offences with transfers. This means that, individuals who attempt expropriation will be consistently identifiable as irredeemable offenders, who once identified could not be deterred by any sustainable form of punishment. Jailing offenders will therefore be sequentially rational and part of a renegotiation-proof punishment strategy. As long as jail punishment yields a sufficiently low payoff—$\underline{u} < \underline{y} - \underline{c} - ((\beta(1)\rho/\gamma^{i'})\bar{y} - \rho\underline{y})(1-\delta)/\delta$—we will have $i'' < i'$, i.e. jail acts as a deterrent.

If, however, jailing offenders with probability one makes offending unprofitable for all individuals, then jail cannot be part of a renegotiation-proof punishment strategy in a pooling equilibrium—since in that case, upon entering the punishment phase, it should be possible for the players to renegotiate to a cooperative equilibrium where those who just have offended are not jailed and do not offend again; but then, if offences are not punished by jail, all individuals $i \leq i'$ would offend by assumption; hence, a pooling equilibrium with no offenders cannot exist. A separating equilibrium may in this case be supported by a mixed punishment strategy where offenders are jailed with some probability less than one, ensuring that at least one individual actually offends and is jailed. Thus, actual offences are needed in every period for jail to be an effective deterrent. For this to be the case, some new individuals must join the population at any given period, so that unjailed offenders can be present in any period.

Finally, such an equilibrium will be ex-ante efficient—dominating, in terms of ex-ante utility, an equilibrium where jailing is not used—if the reduction in the deadweight loss associated with expropriation attempts is less than the social cost of jailing offenders. Let i''' be the minimum i for which $\bar{n}\bar{c} + \sum_{i>i'''} \underline{c}^i \geq 0$ (the borderline individual that cannot be prevented from offending in the absence of jail punishment). Then, the condition for ex-ante efficiency is $i''[\rho\underline{y} + (q+\underline{u})\delta/(1-\delta)] < i'''\rho\underline{y}/(1-\delta)$.

Proof of Proposition 16

An increase in ζ translates into an increase in $\beta(1)$ and a more than proportional increase in $\bar{\beta}_i$. Then, (18) is a sufficient (but not necessary) condition for c^i to increase in regimes (i) and (ii). If unilateral expropriation attempts are deterred by jail, a larger $\bar{\beta}_i$ always induces a larger c^i.

Proof of Proposition 17

We will focus on the limit of a large economy where the number of individuals of each type is large. As noted earlier, as long as $g < g^*$, c^i will approach a positive constant, whereas, the terms $v(g) - v(g - c^i)$ and $v(g) - v(g - \eta_i(1-\beta_i)y^i)$ in (21) and (22) will become arbitrarily small relative to c^i, which means that it can be neglected. Then, the tax in each of the possible three regimes will approach the following: (i) $c^i = [(\bar{\beta}_i + \delta - 1)y^i - \Lambda^i]/\delta$; (ii) $c^i = \delta\beta y^i - \Lambda^i$; (iii) $c^i = (\delta y^i - \Lambda^i)/(1+\delta)$.

For $c^i < 0$, the tax disappears from the left-hand side of the no-defection constraint, which means that regime (iii) cannot occur—because a negative c^i will be strictly less than $\eta_i(1-\beta_i)y^i$ for all values of η_i and β_i—and the negative tax required to prevent expropriation by i is always minimized in absolute value by a choice $\eta_i = 0$. For $c^i > 0$, regime (iii) can occur. For regime (ii) to be possible in this case, we must have $\bar{\beta}_i \leq (1+\Lambda^i/y^i)/(1+\delta)$, otherwise $c^i = \delta\bar{\beta}_i y^i - \Lambda^i = \eta_i(1-\bar{\beta}_i)y^i - \Lambda^i$ would require $\eta_i > 1$. Regime (iii) occurs when $\bar{\beta}_i > (1+\Lambda^i/y^i)/(1+\delta)$. One can verify that, for $\bar{\beta}_i < (1+\Lambda^i/y^i)/(1+\delta)$, the tax in regime (ii) is always greater than that in regime (i), so regime (i) can never be constrained-optimal.

Consider then individuals for whom $\Lambda^i > 0$. These can face positive or negative taxes. If both types of cases are present, the individuals facing positive taxes will be the ones with comparatively higher income, since the tax in regime (ii), $\delta\bar{\beta}_i y - \Lambda^i$, is increasing with y^i, and the tax in regime (iii) must be positive and regime (iii) requires a comparatively higher $\bar{\beta}_i$ (and therefore a higher y^i). In contrast, individuals for whom $\Lambda^i < 0$ always face a positive tax. Given that Λ^i is decreasing in y^i, these will be individuals whose income is comparatively higher relative to individuals for whom $\Lambda^i < 0$.

Then, as we move up the income distribution, we move through the following possible cases: (a) a negative tax rate $\delta\bar{\beta}_i - (\xi/y^i - \rho) < 0$ (where $\xi \equiv (\rho/\gamma)\beta(1)y^N$) and a marginal rate $\delta\bar{\beta}_i + \rho > 0$; (b) a positive tax rate $\delta\bar{\beta}_i - (\xi/y^i - \rho) > 0$ and a marginal tax rate $\delta\bar{\beta}_i + \delta y^i d\bar{\beta}_i/dy^i + \rho > 0$; (c) a positive tax rate $\delta/(1+\delta) - (\xi/y^i - \rho)/(1+\delta) > 0$ and a marginal tax rate $\delta/(1+\delta) + \rho/(1+\delta) > 0$; (d) (e) a positive tax rate $\delta\bar{\beta}_i > 0$ and a marginal tax rate $\delta\bar{\beta}_i + \delta y^i d\bar{\beta}_i/dy^i > 0$; (e) a positive tax rate $\delta/(1+\delta) > 0$ and a marginal tax rate $\delta/(1+\delta) > 0$. Cases (c) and (d) can never be simultaneously present (since increasing y^i cannot simultaneously make Λ^i zero and trigger a switch from regime (iii) to regime (ii)). The other cases need not all necessarily feature at the same time in an equilibrium. For example, if $\bar{\beta}_1$ (the maximum collectively rational punishment on the lowest income individual) is greater than $(1+\Lambda^i/y^i)/(1+\delta)$, (b) and (d) will not occur.

Moving through these various cases—and noting that in cases (c) and (e) we have $\bar{\beta}_i > (1+\Lambda^i/y^i)/(1+\delta)$—we conclude that the average rates of taxation will be (weakly) increasing in income and will always be at most $1/(1+\delta)$. This can be viewed as a natural upper bound to the rate of taxation that can be sustained by the threat of single-period, renegotiation-proof punishment strategies (for $\delta \leq 1$, it can never exceed $1/2$).

Marginal tax rates, on the other hand, will be positive but may be nonmonotonic with respect to income; even when they are always increasing in income, the rate at which they do so will eventually fall with income. For $\Lambda^i < 0$, in regime (iii) the marginal rate of taxation will be $1/(1+\delta)$. In regime (ii) (for comparatively lower income levels) the marginal rate of taxation will be $\delta\bar{\beta}_i + \delta\bar{\beta}_i y^i d\bar{\beta}_i/dy^i$. If the elasticity of $d\bar{\beta}_i/dy^i$ with respect to y^i is less than two in absolute value, this rate will be increasing with y^i; otherwise it will be decreasing. In the former case, if both regimes (ii) and (iii) are present, there must be a range of income levels, just below the critical level for which $\bar{\beta}_i = 1/(1+\delta)$, where marginal rates are greater than $\delta/(1+\delta)$ (since for $\bar{\beta}_i$ approaching $1/(1+\delta)$ the marginal rate approaches $\delta/(1+\delta) + \delta(d\bar{\beta}_i/dy^i)y^i > \delta/(1+\delta)$); this means that at $\bar{\beta}_i = 1/(1+\delta)$ the marginal tax rate will fall and the marginal tax rate schedule will therefore be nonmonotonic. If the marginal rate of tax is decreasing in regime (ii), it will be monotonically decreasing and will eventually settle down to the value $\delta/(1+\delta)$. For $\Lambda^i > 0$, the marginal rates in regimes (ii) and (iii) will involve an additional positive term—$\rho > 0$ in regime (ii) and $\rho/(1+\delta) > 0$ in regime (iii), respectively. Thus, marginal tax rates will be comparatively higher for individuals who face negative taxes.

Proof of Proposition 18

This follows directly from concavity. Suppose that a group of n_1 individuals have income $y+x/n_1$ and another group of n_2 individuals have income $y-x/n_2$, and consider the effect of a mean preserving income spread, represented here by an increase in x. The net revenues, $\hat{g}$, that can be sustained by other groups will be unaffected by this change. If we totally differentiate $g = \hat{g} + n_1 c(y+x/n_1) + c(y-x/n_2))$ with respect to x, $c''(y) < 0$ implies $dg/dx < 0$.

Proof of Proposition 19

Focus on a case where there are $\overline{n}$ voluntary contributors all having the same income y and preference parameter $\overline{\theta}$, and let $\bar{g}$ be revenues from those who are not voluntary givers. Then the sum, $\overline{c}$, of tax and voluntary contribution for each voluntary giver is identified by $y - \overline{\theta}\left(v(\bar{g}+\overline{nc}) - v(\bar{g}+\overline{nc}-y)\right) = \delta\left(y-\overline{c}\right)$, which gives $\partial(\overline{nc})/\partial\bar{g} = -1/(1-\delta/(n\overline{\theta}(v'(\bar{g}+\overline{nc}) - v'(\bar{g}+\overline{nc}-y)))) > -1$.

Appendix B

B.1 Self-enforcing Taxes with Unobservable Productivity

Consider the two-type scenario described in Section 5.1.

The repentance condition for the low-productivity individual—which is the binding one in this regime—in a constrained-efficient equilibrium can be written as

$$\begin{aligned}(1-\underline{l}^*)\underline{y} + h(\underline{l}^*) - \theta\left(v(g) - v(g-\underline{y})\right) \\ = \delta\left((1-\underline{l}^*)y + h(\underline{l}^*) - \underline{c}\right).\end{aligned} \tag{36}$$

An analogous condition identifies the sustainable tax for the high-productivity individual.

The no-mimicking constraint can be expressed as follows:

$$\left[\overline{y}(1-\overline{l}) + h(\overline{l}) - \overline{c}\right] - \left[\underline{y}(1-\underline{l}) + h\left(1 - \underline{y}(1-\underline{l})/\overline{y}\right) - \underline{c}\right] \equiv \Phi(\underline{c},\overline{c},\underline{l},\overline{l};\underline{y},\overline{y}) \geq 0. \tag{37}$$

There is another possibility to consider: the high productivity type must have no incentive to mimic the earnings of the low-productivity type *and* pay no taxes at the same time, thus behaving as a low-productivity defector (in a sense evading and avoiding at the same time). However, if the enforcement constraints for the low-productivity type are satisfied, mimicking a defecting low-productivity individual would reveal the defector as a high-productivity individual, and the appropriate punishment could then be administered; hence the standard no-mimicking constraint applies in conjunction with the relevant enforcement constraint(s)—depending on whether or not expropriation is feasible and on which optimality regime the individual falls under for the high-productivity type.

Any constrained-efficient contract will always feature an efficient level of leisure $\overline{l} = \overline{l}^*$, since any move from this level reduces welfare without relaxing the no-mimicking constraint nor the enforceability constraint. Furthermore, the enforceability constraint for the low-productivity type will always be binding at an optimum, since relaxing this constraint—via a decrease in $\underline{c}$ below the level which makes the low-productivity type's enforcement constraint binding—results directly in lower contributions and tightens the no-mimicking constraint, which then needs to be offset by a reduction in contributions by high-productivity individuals and/or by a (further) reduction in $\underline{l}$ below its efficient level $\underline{l}^*$.

It is, in principle, possible for the relevant enforcement constraints for both types to be binding at an optimum and for the no-mimicking constraint to be slack—as in the a full-information scenario; in this case contributions do not distort labour supply decisions. This, however, cannot occur if $\overline{y}$ and $\underline{y}$ are close to each other: for $\underline{l} = \underline{l}^*$, $\overline{y}$ approaching $\underline{y}$, and as long as the marginal rate of contributions out of earnings is positive, mimicking by the high-productivity type produces a positive gain for that type. Since we normally think of a continuum of productivity types—of which the discrete two-types model is only an abstract but convenient representation—a scenario where the self-selection constraint is slack is not practically relevant.

Then, at an optimum where the no-mimicking constraint is binding, the relevant low-productivity enforcement constraint will be binding–and will constrain the overall level of taxation in the economy—while the high-productivity enforcement constraint will be slack, with the high-productivity tax being identified by the mimicking constraint. Denote with $\underline{\Omega}(\underline{c}, \underline{l}; \underline{y}) \geq 0$ and with $\overline{\Omega}(\overline{c}, \overline{l}; \overline{y}) \geq 0$ the relevant enforcement constraints respectively for low- and high-productivity type individuals. Then, such equilibria will be characterized by $\underline{\Omega}(\underline{c}, \underline{l}; \underline{y}) = 0$, $\overline{\Omega}(\overline{c}, \overline{l}; \overline{y}) > 0$, and $\Phi(\underline{c}, \overline{c}, \underline{l}, \overline{l}; \underline{y}, \overline{y}) = 0$. To ensure that the no-mimicking constraint is not violated, the level of labour supply of low-productivity individuals will have to lie below the efficient level, thus making mimicking more costly for high-productivity individuals.

In the case of an optimum where taxes are determined by revenue-maximization, the condition identifying a sustainable, constrained-efficient, incentive-compatible combination $(\underline{c}, \overline{c})$ (together with the binding no-mimicking constraint and the relevant enforcement constraint for the low-productivity type) is

$$\underline{n}\frac{d\underline{c}}{d\overline{c}} + \overline{n} = 0, \tag{38}$$

where

$$\frac{d\underline{c}}{d\overline{c}} = \frac{\underline{\Omega}_{\underline{l}}\Pi_{\overline{c}}}{\underline{\Omega}_{\underline{c}}\Pi_{\underline{l}} - \underline{\Omega}_{\underline{l}}\Pi_{\underline{c}}}; \tag{39}$$

$$\frac{d\underline{l}}{d\overline{c}} = \frac{-\underline{\Omega}_{\underline{c}}\Pi_{\overline{c}}}{\underline{\Omega}_{\underline{c}}\Pi_{\underline{l}} - \underline{\Omega}_{\underline{l}}\Pi_{\underline{c}}}. \tag{40}$$

Condition (38) states that total revenues are maximized (a marginal increase in $\overline{c}$ leaves total revenues unchanged). In this regime, even in the presence of distortions, enforcement considerations alone—which here operate through a combination of self-enforcement and self-selection constraints—determine levels of taxation.

In the alternative case where an optimum is determined by equality between the marginal valuation of collective consumption and the marginal cost of the distortion associated with the no-mimicking constraint, the high-productivity contribution is not solely limited by enforcement incentives; as in conventional optimal tax models, it is identified as the level of tax that strikes an optimal balance between the need for increasing collective consumption provision and the cost that this entails in terms of distortions of labour-leisure choices. The condition identifying a sustainable, constrained-efficient, incentive-compatible combination $(\underline{c}, \overline{c})$ in this regime (together with the binding no-mimicking constraint and the relevant enforcement constraint for the low-productivity type) is

$$(n\theta v'(g) - 1)\left(\underline{n}\frac{d\underline{c}}{d\overline{c}} + \overline{n}\right) - \underline{n}(\underline{y} - h'(\underline{l}))\frac{d\underline{l}}{d\overline{c}} = 0. \tag{41}$$

with

$$\underline{n}\frac{d\underline{c}}{d\overline{c}} + \overline{n} > 0. \tag{42}$$

Condition (41) states that the marginal social surplus from the additional collective consumption that results from an increase in $\overline{c}$ must equal the marginal cost of the additional distortion induced by this increase on labour supply decisions. It should be stressed, however, that even in this regime, enforceability constraints—with respect to the low-productivity contribution only in this case—are a fundamental determinant of progressivity: in the absence of any enforceability constraints (and of redistributional concerns), requiring a common tax level from all productivity types would be efficient.

Whether enforcement constraints or the costs of labour-choice distortions are going to dictate the shape of income tax schedules will depend on the extent to which labour supply can respond. If labour supply is very elastic, then consideration of labour-choice distortions will dominate; however, if labour supply is less elastic, enforceability constraints will be binding at an optimum. One could then expect enforceability considerations to be more likely to dictate the shape of income tax schedules in economies characterized by comparatively less flexible responses (because of intrinsic behavioural characteristics or, equivalently, because of less flexible market institutions) than in economies where labour supply is more elastic.

B.2 Life-cycle Contributions in an OLG Economy

Consider an OLG framework where individuals live for three periods (young, adult, old) and all earn the same nonexpropriable income y in all periods. In each period $n/3$ old individuals die and $n/3$ new individuals are borne, thus maintaining a constant overall population size, n. Assuming that g and n are large enough that $\theta v'(g) < 1$, any individually rational strategy will involve zero contributions by old individuals, i.e. $c^3 = 0$. Adult individuals can be induced to make a positive contribution, $c^2 > 0$, under the threat of a reduction in contributions by others in the last period of their lives; we then have $c^2 - \theta\left(v(g) - v(g - c^3)\right) \leq \delta\theta\left(v(g) - v(\min\{g, g_3^*\})\right)$, where g_3^* is the level of provision that is collectively optimal for young and adult individuals. Note that it is collectively rational to apply the punishment because old individuals cannot be induced otherwise to make positive contributions. Suppose now that young individuals are to be induced to make a positive contribution, $c^1 > 0$. A reduction at t could then be followed by the requirement to pay a fine at $t+1$ followed by a reversion to normal play at $t+2$. Note that a reduction in contributions beyond the fine at $t+1$ and/or at $t+2$ by others is ruled out by the need to keep other adult individuals cooperating at c^2 during punishment (so as not to violate collective rationality). Then the maximum fine payable (satisfying the repentance constraint at $t+2$) is simply c^2. But then it is not possible to induce young individuals to add to total tax revenues: if c^2 is lowered, positive contributions by the young may be supported, but the associated increase in contributions from the young is less than the associated loss of revenue from the adults (because we must have $c^1 = \delta(c^2_{MAX} - c^2)$). So, a constrained-efficient equilibrium will feature $c^1 = 0$, $c^2 > 0$, $c^3 = 0$.

B.3 Politics as an Equilibrium Coordination Mechanism under Private Information

Consider an economy where all individuals have the same income, y, but where individuals are of different preference types, θ^i. For the sake of simplicity, we shall assume that income expropriation is not feasible. Suppose that θ^i is private information. Furthermore, suppose that—unlike in our

earlier discussion—the distribution of types is also unknown. Then coordination to the self-enforcing contract that is ex-ante constrained efficient for each realization of the distribution of types must rely on individuals truthfully revealing their type. To be more specific, assume that there are only two preference types, $\underline{\theta}$ and $\overline{\theta} > \underline{\theta}$, and there are two possible realizations of the type distribution: under distribution 1, there are $\underline{n}$ and $\overline{n}$ individuals of the low and high preference types respectively, while under distribution 2, the corresponding numbers are $\underline{n}-1$ and $\overline{n}+1$, and each distribution can occur with probability equal respectively to π_1 and π_2, where $\pi_1 + \pi_2 = 1$. In other words, the uncertainty over the distribution of types is limited here to the preference type of a single individual.

Suppose that individuals truthfully reveal their type under each realization; then it will be ex-ante optimal to coordinate on an arrangement whereby the strategies played by each player are the equilibrium strategies corresponding to the self-enforcing contract that is constrained-efficient for that particular realization depending on what players reveal; these contracts will be denoted as $\mathcal{C}_w$ $(w = 1, 2)$. This course of action will also be individually rational, i.e. it will be consistent with a noncooperative equilibrium: if, depending on players revealing realization w $(w = 1, 2)$, all individuals except i adopt the equilibrium strategies corresponding to contract $\mathcal{C}_w$, then, as long as players' announcements can be taken as truthful, adopting the equilibrium strategy that $\mathcal{C}_w$ assigns to i will be a best response for i. (The problem of deriving such constrained-efficient, constrained-efficient equilibria can be formally characterized as an implementation problem.)

What remains to be seen is whether, given the strategies adopted ex post in dependence of the other players' initial announcements, players will find it individually rational to reveal their true type truthfully, i.e. whether they it will be a best response for each of them to announce their true type when all other players also announce their true type, and when the type distribution so revealed leads to the adoption of the self-enforcing contract that is ex-ante constrained-efficient for that realization.

It can be shown that, in this case, truthful revelation can generally not be secured. Consider the choice of an individual i who is a high-preference individual, when all other individuals truthfully reveal their type. If only $\overline{n} - 1$ of the other individuals have announced that they are high-preference types, then there is no scope for misrepresentation by i, since it is common knowledge that a realization with $\overline{n} - 1$ high-preference individuals occurs with probability zero. However, if $\overline{n}$ of the other individuals have announced that they are high-preference types, then i, being a high-preference individual, could misrepresent her type. If the high-preference individual also announces truthfully, the resulting equilibrium is $\mathcal{C}_2$, otherwise it is $\mathcal{C}_1$. In the former, $\underline{n} - 1$ players contribute a certain level $\underline{c}_2$ and $\overline{n} + 1$ players contribute $\overline{c}_2 > \underline{c}_2$; in the latter, $\underline{n}$ players contribute a certain level $\underline{c}_1$ and $\overline{n}$ players contribute $\overline{c}_1 > \underline{c}_1$. Given that all players other than i believe the other players announcements to be truthful in each case, they will behave as prescribed by the corresponding strategies. Thus, if i, being a high-preference type individual, misrepresents her type as being low-preference, all other $\overline{n}$ high-preference players will contribute $\overline{c}_1$ and i will be able to contribute $\underline{c}_1$ without incurring any punishment.

Then i will have an incentive to misrepresent her type in this way if

$$\overline{c}_2 - \underline{c}_1 - \overline{\theta}(v(\underline{g}_2) - v(\underline{g}_1)) > 0. \tag{43}$$

Note that, since $\underline{g}_2 > \underline{g}_1$, we will have (by (4)) that $\overline{c}_2 < \overline{c}_1$) and $\underline{c}_2 < \underline{c}_1$. Thus, $\underline{g}_2 - \underline{g}_1 = \overline{n}(\overline{c}_2 - \overline{c}_1) + (\underline{n} - 1)(\underline{c}_2 - \underline{c}_1) + \overline{c}_2 - \underline{c}_1 < \overline{c}_2 - \underline{c}_1$. Then (43) will be satisfied as long as $\underline{g}_1$ is high enough that $\overline{\theta}v'(\underline{g}_1) < 1$, since in that case the unilateral gain from the reduced contribution will exceed the i's valuation of the reduction in collective consumption.

Low-preference individuals will have no incentives to misrepresent their type when $\underline{n} - 1$ of the other individuals (truthfully) announce that they are low-preference individuals. If they

do, less than the expected level of contributions will be received (since they will not follow up their announcement with payment of the higher contribution). Punishment will then have to follow, even if individuals as a whole can infer from the deviation that misrepresentation took place. Repentance will require payment of y, and expecting for this payment to be made by the individual who misrepresented her type can be an equilibrium, assuming that everyone else (including future incarnations of that individual) keep announcing truthfully in all subsequent rounds. Since i is punished anyway, it will be a best response for her to contribute zero. Then, the level of provision at t will be $\underline{g}_2 - \overline{c}_2 < \underline{g}_1$), and i will then experience a net gain, from misrepresenting her type, equal to

$$(1+\delta)\underline{c}_1 - \delta y - \overline{\theta}(v(\underline{g}_1) - v(\underline{g}_2 - \overline{c}_2)) \leq$$

$$(1+\delta)\underline{c}_1 - \delta y - \overline{\theta}(v(\underline{g}_1) - v(\underline{g}_1 - \underline{c}_1)) \leq 0, \tag{44}$$

(since $\underline{g}_2 - \overline{c}_2 < \underline{g}_1 - \underline{c}_1$ and given that the second inequality coincides with (3)).

There are then two possible alternatives: either the equilibrium $\mathcal{C}_1$ is adopted in all cases, or, rather than coordinating to either $\mathcal{C}_1$ or $\mathcal{C}_2$ in dependence of the players' announcements, players coordinate to alternative continuation equilibria, selected in such a way that misrepresentation by high-preference individuals does not occur. Specifically, rather than associating the ex-ante constrained-efficient equilibrium $\mathcal{C}_1$ to realization 1, players can associate an alternative contract, $\mathcal{C}_1' = (\underline{c}_1', \overline{c}_1')$, for which

$$\overline{c}_2 - \underline{c}_1' - \overline{\theta}(v(\underline{g}_2) - v(g_1')) \leq 0, \tag{45}$$

thus ensuring truthful revelation by high-preference individuals. Such alternative contract will have to feature lower aggregate contributions because of lower contributions by high-preference individuals, $\overline{c}_1'$ in comparison with the maximum supportable level $\overline{c}_1$, so as to increase the loss $\overline{\theta}v(\underline{g}_2) - v(g_1'))$ relative to $\overline{c}_2 - \underline{c}_1'$, as well as higher contributions by low-preference individuals, $\underline{c}_1'$ (consistently with a lower g_1'). Such a combination $(\mathcal{C}_1', \mathcal{C}_2)$ supported by truthful revelation will constitute an incentive-compatible, ex-ante constrained-efficient arrangement.[49]

Note that anonymous revelation—whereby the distribution of individuals' announcements is publicly observable, but where announcements cannot be mapped back to the individuals making then—will be sufficient to support such an arrangement, since the selection mechanism described uses announcements in an anonymous way. Indeed, nonanonymous announcements could generally not be expected to lead to truthful revelation if the associated mechanisms uses them nonanonymously, e.g., by conditioning an individual's punishment on their announcement. Suppose, for example, that there is income expropriation and that there are also unobservable preferences, and suppose that you have a single high-preference individual. If y is sufficiently large, and unless this individual values collective consumption enough that her marginal valuation is not below unity at g, it will not be possible to sustain a higher contribution by her in comparison with low-preference individuals, whether or not the distribution of types is known (i.e. with or without truthful, anonymous revelation). On the other hand, if this individual where to reveal herself fully as high preference, this information could be used to support a higher contribution by her under threat of expropriation. If punishment strategies are conditioned nonanonymously on nonanonymous announcements in this way, they are incompatible with truthful revelation, and we should not therefore expect nonanonymous signals to be used.

[49]This problem belongs to the class of repeated games with communication and imperfect private monitoring, as described by Compte (1998).

How will individuals signal their type under such an arrangement? Any set of signals will do as long as their meaning is part of an accepted set of conventions. In particular, anonymous voting over alternatives can play this role, if the various alternatives over which individuals vote are each conventionally associated to a different individual type.

Is there a natural way of associating policy alternatives to types in this way? Suppose that we think of $\mathcal{C}_1'$ and $\mathcal{C}_2$ as being the two alternatives over which voting occurs. Can we naturally associate $\mathcal{C}_1'$ with low preference and $\mathcal{C}_2$ with high preference in the sense of each being the alternative favoured by each type? Not generally. It is possible for both types to be better off at $\mathcal{C}_2$ in comparison with $\mathcal{C}_1'$, assuming that each is adopted in correspondence of the true realization of the preference distribution. Thus, there does not seem to be a natural mapping from policy alternatives to signals; therefore, if the only role of voting is facilitating coordination to a self-enforcing social contract through information pooling, there does not seem to be any reason to expect voting to take place directly over policy alternatives.

A more natural way of signalling preferences anonymously could be by voting over candidate types, which provides a direct mapping from types to signals. Then, elected policymakers would simply serve as a publicly observable signal, whose characteristics and identity convey information to individuals about which particular self-enforcing equilibrium to coordinate on, even though they possess no independent policy making powers (beyond the power to enact the policies they represent). Nevertheless, as discussed earlier, the need to achieve information pooling through voting will generate divergence in the policies that each of these candidate types represents—the kind of ideological tension we associated with politics. In order to secure truthful revelation, the self-enforcing policies associated with each candidate type may have to be skewed away from the self-enforcing policies that are ex-ante constrained efficient in each realization. In our previous discussion this translated into a bigger provision gap across policy alternatives combined.

We can generalize the above construction to the case where positive probabilities π_j are attached by individuals (with common beliefs) to all possible pairs $j = (\underline{n}, \overline{n})$ such that $\underline{n} + \overline{n} = n$. As discussed above, for any of such combinations, a high-preference individual would be tempted not to reveal her type truthfully if the constrained efficient self-enforcing contract is adopted in conjunction with it. Thus, truthful revelation will require the adoption of alternative self-enforcing contracts under each realization, such that truthful revelation will be consistent with a Bayesian Nash equilibrium, i.e. it will be a best response "on average" for high-preference individuals. Such a best response is formally characterized by the condition

$$\sum_{j \neq (0,n)} \pi_j \left(\overline{c}'_{j+1} - \underline{c}'_j - \overline{\theta}(v(\underline{g}_2) - v(g'_1)) \right) \leq 0, \tag{46}$$

where (abusing notation), $j = (\underline{n}, \overline{n})$ and $j+1$ denotes the distribution $(\underline{n} - 1, \overline{n} + 1)$. Again, this will require broader gaps between levels of collective consumption under each realization as well as contributions by high- and low-preference types that are closer to each other.

Note, however, that in the above story it is not only the identity of the candidate elected that conveys information about the distribution of types; it is the actual distribution of votes. This means that the collective choices that will prevail in each case will not just depend on the candidate elected by majority (as it does, for example, in a citizen-candidate model of collective choices): in this interpretation, the same candidate will adopt different policies in dependence of the voting outcome. Such prediction is consistent with the observation that the actual policies put in place by policymakers do reflect the degree of political support received, i.e. the strength of the political mandate matters: a policymaker elected by a narrow margin will typically tend to adopt more centrist policies expecting that more extreme policies will be met by stronger opposition (i.e. they will hit against enforceability constraints).

The above story does not rule out that politics may play an additional information-aggregation role with respect to other types of information that may be relevant to collective choices. Suppose, for example, that the characteristics of the contract that is ex-ante efficient depend on the prevailing state of the world, and that individuals have different prior beliefs regarding the likelihood of each state of the world prevailing. Then voting could serve as an information-aggregation device (along the lines described in the literature, e.g. by Feddersen and Pesendorfer, 1996) that could make it possible for the group to converge to a (more accurate) collective prior, and, on the basis of this aggregation, to a suitable self-enforcing contract. As time unravels and new information is gathered—some of it private but much of it public—there will a need for updating the prior on the basis of this new information, which in turn may require further pooling. For example, different individuals may acquire new pieces of (private) information and they may also be uncertain about what new information others may have acquired, which means that there will no longer be a common understanding of which continuation equilibrium everyone should be coordinating to; so, a new, single, commonly observable, announcement, replacing the various different signals privately observed by each individual, will be required to achieve such coordination.

Voting could then take place on an ongoing basis—which may be prohibitively costly—or, alternatively, if the rate of arrival of new information is not too fast, it may occur at certain specified intervals; and, in the periods intervening between consecutive voting rounds, some individual or restricted group of individuals—the appointed policymakers—will be relied upon to update collective beliefs on the basis of information they can gather themselves. Delegating information updating to an individual or a subset individuals in this way is more likely to be preferable to information-pooling through voting when the new information can be accessed individually by everyone, albeit at a cost. So, one could view private information about individual preferences as representing one polar case—where informational asymmetry requires pooling—and of public but costly information as the other polar case.

Delegates, in turn, may be subjected to the temptation to misrepresent any information they gather. To induce truthful revelation by policymakers, it may then be necessary to skew continuation equilibria in their favour, which will confer some informational rents to them;[50] nevertheless, doing so may still be ex-ante preferred to incurring the costs associated with ongoing voting.[51] In addition, truthful revelation may be supported by the application of some suitable punishment, which must itself be credible and plausible (i.e., renegotiation-proof).[52] This could consist

[50]E.g., a regime where delegates have ample (albeit not limitless) latitude to skew the outcome in their favour, thus receiving significant rents, would be consistent with a *dictatorship*.

[51]Buchanan discusses the related idea of *rational ignorance* on the part of voters in comparison with policymakers, describing policymakers as specialized information gatherers, and voters as individuals who rationally choose not to incur the cost of gathering information, preferring instead to defer decisions to the policymaker. As has been already mentioned, this choice may stem from the fact that pooling information on a continuous basis is too costly; or they may be even more compelling reasons for delegation. Suppose, for example, that there is a collective interest that certain information should not be revealed to outsiders (e.g. another competing group of individuals), and suppose that revealing this information to group members also reveals it to outsiders (confidential information). Then an ex-ante optimal arrangement may involve the (arbitrary) appointment of certain individuals who have exclusive access to this information and must be relied upon by others to send out suitable coordinating signals.

[52]This is not unlike the idea of a self-enforcing contract between voters and elected policymakers as described by Acemoglu (2002); the key difference being that policymakers do not have here unlimited latitude with respect to the choice of policies (they are effectively restricted to the self-enforcing policies that do not clash with people's priors, although in the case of one-off policies, the enforceability constraint will not apply).

of switching to a Pareto-undominated continuation equilibrium that penalizes the policymaker given her type: in other words, a policymaker who has been found to exhibit undue bias in her choice of policy announcement may have to pay for this bias by a subsequent swing in collective choices in the opposite direction, possibly even personally sanctioning this change while in office. Finally, the identity of the delegate and her preferences will affect her misrepresentation incentives and the information rents that have to be surrendered to her by an ex-ante constrained-efficient system of state-contingent, self-enforcing equilibria. The delegate's type should then be selected accordingly. This means that the ex-ante efficient delegation choice will involve delegation to a policymaker whose type is least likely, on average, to be tempted to misrepresent information. Such a choice, in combination with a suitable set of state-contingent self-enforcing equilibria (inducing truthful revelation) will minimize the expected informational rents that have to be surrendered to the delegate.[53]

[53]This idea can be formalized as follows. Let θ summarize preferences, and let the continuation payoff for an individual of type θ in state s is the continuation equilibrium $\mathcal{C}$ is selected be denoted by $U(\mathcal{C}; s, \theta)$. Let $(\mathcal{C}_s(\theta'), \forall s) \equiv \mathcal{R}(\theta')$ be the profile self-enforcing state-contingent continuation equilibria (referring to all possible states) that are ex-ante constrained-efficient and induce truthful revelation in each possible state if the delegate is of type θ' (the profile $(\mathcal{C}_s, \forall s)$. which, given delegate θ', maximizes the expectation $E_{s\theta}[U(\mathcal{C}_s; s, \theta)]$ subject to $U(\mathcal{C}_s; s, \theta')] \geq U(\mathcal{C}_z; s, \theta')], \forall z \neq s)$. Then, a constrained-efficient choice of delegate θ' is that which maximizes the expectation $E_{s\theta}[U(\mathcal{C}_s(\theta'); s, \theta)]$. Delegates with extreme preferences are on average going to be more likely to be tempted to misrepresent the true state of the world (since continuation equilibria in all states are selected ex-ante to benefit the average type). However, given that the relationship between individual types, misrepresentation incentives, and the ex-ante costs associated with ex-post informational rents is generally nonlinear, the ex-ante optimal delegate θ' need not necessarily coincide with the median nor with the mean type. The above story in itself does not require voting. All individuals would agree ex-ante that θ' is the best delegate. Suppose, however, that we also have information-pooling about individual preference to take care of and we do this by voting over candidates. So a state of the world is now represented by a couple (s, d) where d refers to a particular distribution of preferences in the population. Suppose then, that the vote must reveal individuals' preferences and also result in a selection of delegate (consistently with the ex-ante selection of overall tax constitution, whereby we all agree that a particular delegate $\theta'(d)$ is best under distribution d). Also suppose that $\theta'(d)$ happens to be the median under d. Then voting for a certain candidate could be conventionally taken as an anonymous signal that the voter is of that type, and at the same type the median vote would actually be for $\theta'(d)$. (Truthful revelation may require some further skewing of continuation equilibria to induce voters to announce truthfully (as discussed earlier) beyond the skewing required to induce truthful revelation by candidates; this can be readily formalized but is omitted here for the sake of brevity.) If, however, the median under d is not the optimal delegate $\theta'(d)$, then it could still be the case that the distance between $\theta'(d)$ and the median under d is independent of d—say equal to x—which means that casting a vote for a candidate θ'' could be taken as a signal that the voter is of type $\theta'' - x$ while ensuring that the median vote is for candidate $\theta'(d)$ under all possible type distributions. There will be situations, however, where x is not independent of d; nevertheless voters could in principle compute the expectation $E_d(x(d))$ and vote for a type $\theta'' - E_d(x(d))$ if they are of type θ''; this will result in the election of a delegate that is of type $\theta'(d) - [x(d) - E_d(x(d))]$, which may result in a biased choice of candidate in some states. Such bias will nevertheless be preferable to selecting a candidate independently of the distribution d (i.e. without voting), which would mean selecting a θ', independent of d, that maximizes the ex-ante expectation of welfare taken over all possible realizations of d.

References

Abreu, D. (1988). "On the theory of infinitely repeated games with discounting," *Econometrica* 56, 383-396.

Abreu, D., D. Pearce, and E. Stacchetti (1993) "Renegotiation and symmetry in repeated games," *Journal of Economic Theory* 60, 217-240.

Acemoglu, D. (2002) " Why not a political Coase Theorem? Social conflict, commitment and politics," National Bureau of Economic Research Working Paper No. 9377.

Andreoni, J. (1990) "Impure altruism and donations to public goods: a theory of warm-glow giving?" *Economic Journal* 100, 464-77.

Becker, G. (1968) "Crime and punishment: an economic approach," *Journal of Political Economy* 76, 169-217.

Bergstrom, T. (2002) "Evolution of social behavior: individual and group selection," *Journal of Economic Perspectives* 16, 67-88.

Bergstrom, T., L. Blume, and H. Varian (1986) "On the private provision of public goods," *Journal of Public Economics* 29, 25-49.

Bernheim, B., Peleg, B., and Whinston, M. (1987) "Coalition-proof Nash equilibria: concepts," *Journal of Economic Theory* 42, 1-12.

Bevan, D. (2004) "Optimum income taxation when employers cannot accurately observe productivity and effort." Mimeo, University of Oxford.

Bird, R. (2003) "Taxation in Latin America: reflections on sustainability and the balance between equity and efficiency." Mimeo, University of Toronto.

Boyd, R., and P. Richerson (1994) "The evolution of norms: an anthropological view," *Journal of Institutional and Theoretical Economics* 150, 72-87.

Brennan, G., and J. Buchanan (1980) *The power to tax: analytical foundations of a fiscal constitution*, Cambridge University Press, Cambridge.

Buchanan, J. (1975) *The limits to liberty*. Chicago: University of Chicago Press.

Buchanan, J., and Flowers, M. (1960) *The public finances: An introductory textbook*. Homewood, Ill.: Irwin.

Compte, O. (1998) "Communication in repeated games with imperfect private monitoring," *Econometrica* 66, 597-626.

Dal Bo, P. (2001) "Social norms, cooperation and inequality," UCLA Economics Working Paper No. 802.

Chen, Z., and L. Epstein (2002) "Ambiguity, risk, and asset returns in continuous time ," *Econometrica* 70, 1403-1443.

Farrell, J., and E. Maskin (1989) "Renegotiation in repeated games," *Games and Economic Behavior* 1, 327-60.

Feddersen, T., and W. Pesendorfer (1996) "The swing voter's curse," *American Economic Review* 86, 408-424.

Greif, A. (1994) "Cultural beliefs and the organization of society: a historical and theoretical reflection on collectivist and individualist societies," *Journal of Political Economy*102, 912-950.

Greif, A. (1998) "Historical and comparative institutional analysis," *American Economic Review* 88, 80-84.

Greif, A., P. Milgrom, and B. Weingast (1994) "Coordination, commitment, and enforcement: the case of the merchant guild," *Journal of Political Economy* 102, 745-776.

Grossman, H. (2001) "The creation of effective property rights," *American Economic Review* 91, 347-352.

Hayek, F. (1967) *Studies in philosophy, politics and economics*. Chicago: University of Chicago Press.

Hobbes, T. (1651) *Leviathan*, edited by C. Macpherson. London: Penguin Classics, 1986.

Kandori, M. (1992) "Repeated games played by overlapping generations of players," *Review of Economic Studies* 59, 81-92.

Kreps, D., and R. Wilson (1982) "Reputation and imperfect information," *Journal of Economic Theory* 27, 253-79.

Laffont, J.-J. (2000) *Incentives and political economy* Oxford and New York: Oxford University Press.

Locke, J. (1690) *Two treatises on government*. Cambridge: Cambridge University Press, 2001.

Mankiw, N., D. Romer, and D. Weil (1992) "A contribution to the empirics of economic growth," *Quarterly Journal of Economics* 107, 407-437.

McMillan, J. (1979) "Individual incentives in the supply of public inputs," *Journal of Public Economics* 12, 87-98.

Mirrlees, J. (1971) "An exploration of the theory of optimum income taxation," *Review of Economic Studies* 38, 175-208.

Muthoo, A. (2003) "A model of the origins of basic property rights," University of Essex Econ-Paper No. 546.

North, D. (1990) "Institutions, institutional change and economic performance," Cambridge: Cambridge University Press.

Olson, M. (1965) *The logic of collective action: public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Ratto, M., R. Thomas, and D. Ulph (2004) "Tax compliance as a social norm and the deterrent effect of investigations," mimeo, University of Bristol and Inland Revenue.

Rawls, J. (1971) *A theory of justice*. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Robson, A. (2001) "The biological basis of economic behavior," *Journal of Economic Literature* 39, 11-33.

Robson, A., and H. Kaplan (2003) "The evolution of human life expectancy and intelligence in hunter-gatherer economies," *American Economic Review* 93, 150-69.

Rousseau, J.-J. (1762) *The social contract* translated by M. Cranston. London: Penguin Classics, 1968.

Van Damme, E. (1989) "Renegotiation-proof equilibria in the repeated Prisoners' Dilemma," *Journal of Economic Theory* 47, 206-17.

Weingast, B. (1997) "The political foundations of democracy and the rule of law," *American Political Science Review* 91, 245-263.

Wicksell, K. (1934) *Lectures on political economy*. London: Routledge, 2003.