Self-Enforcing Voting and Delegation in International Organizations

Christina J. Schneider*
Department of Political Science, University of California – San Diego

Branislav L. Slantchev
Department of Political Science, University of California – San Diego

January 16, 2011

Abstract. The traditional approach to explaining international cooperation treats it as a "within-group" problem: all actors can benefit from it but have individual incentives to free-ride, which ensure that cooperation is underprovided. Collective action, however, often creates negative externalities and splits the population of actors into supporters and opponents, all of whom can invest resources toward their preferred outcome. Cooperation becomes a "between-groups" problem which is especially severe when actors have private information about their preferences. We study how actors can communicate these preferences through voting in an environment where they are not bound either by their own vote or the outcome of the collective vote. We identify two organizations with endogenous enforcement — coalitions of the willing and universal organizations, and find that the optimal voting rule can be *ex post* socially efficient. We also analyze a non-coercive organization where actors delegate execution to an agent. Even though this institution is costlier, it does not depend on the shadow of the future, and thus is implementable when the others are not.

*Corresponding author (cjschneider@ucsd.edu). Schneider gratefully acknowledges financial support from the Hellman Foundation. Slantchev gratefully acknowledges financial support from the National Science Foundation (Grant SES-0850435).

1 Introduction

A widespread approach to explaining international cooperation that has emerged over the last twenty-five years is based on insights from the analysis of repeated games.[1] This "cooperation theory" typically assumes that the underlying preferences of governments have the structure of a Prisoners' Dilemma (which makes defection from any agreement the dominant strategy for each single interaction), and then shows how cooperative behavior can be sustained in the long run despite the absence of an agent that can enforce agreements.[2] The answer this theory provides is invariably the same: reciprocal threats to punish deviations from the desired behavior can be used to coerce the cooperation of the actors. To study cooperation in this framework essentially means to study ways of making these threats credible. Research has identified various conditions that can make them more so: sufficiently long "shadow of the future," conditional sanctions that are neither so severe that their execution would be problematic nor so lenient that they would be toothless, and relatively effective monitoring arrangements to detect deviations, whether they are from the desired cooperative pattern or from the costly sanction that enforces it.[3] The PD configuration of preferences is an extreme ideal type where cooperation is presumably most difficult to achieve because each actor strictly prefers not to cooperate regardless of what the others do. In this setting, individual members of the collective have incentives to free-ride on the efforts of others, and the problem of cooperation is one of the collective disciplining individual members that attempt to take advantage of others. Cooperation is a "within group" problem that the collective solves by appropriate group enforcement against individual members.

Although such an approach might explain how cooperation can emerge "spontaneously" under anarchy and make agreements self-enforcing, it is poorly suited as a guide to understanding many interesting cases of international collective action. One reason for this is that international action rarely affects the members of the international system in the same way.[4] Most collective actions of any consequence generate both positive and negative externalities, which split the population of actors into "supporters" and "opponents" of that particular collective endeavor. Cooperative behavior within each group means collaborating toward the commonly desired goal but the disagreement about the goal between the two groups means that actors might find themselves in a highly conflictual situation overall. Whereas free-riding incentives might still arise within each group, a second important problem with respect to the collective action is that of one group overcoming the opposition of the other. In other words, in a world where actors can dedicate costly effort toward their goals, the success of collective action depends on the relative efforts groups contribute

[1] Stein (1982); Axelrod (1984); Keohane (1984); Axelrod and Keohane (1985).

[2] Although there has been some work on problems of coordination and mixed-motive situations where actors agree on the need to coordinate but disagree on the terms they will coordinate on, most research analyzing international cooperation is based on Prisoners' Dilemma (PD)-like situations (Larson, 1987; Rhodes, 1989; Downs and Rocke, 1990, 1995; Evangelista, 1990; Martin, 1992; Fearon, 1998; Downs, Rocke, and Barsoom, 1998; Gilligan, 2004; Voeten, 2005; Svolik, 2006). For a discussion of the coordination dilemma see, for example, Krasner (1991); Garrett (1992); Sebenius (1992); Morrow (1994).

[3] Snidal (1985); Oye (1985); Martin and Simmons (1998); Koremenos, Lipson, and Snidal (2001); Rosendorff and Milner (2001).

[4] Gruber (2000).

toward or against its realization. Cooperation is a "between groups" problem that the collective must solve by an appropriate distribution of benefits to the groups of supporters and opponents of the collective action.

In this paper, we argue for conceiving of international cooperation in terms of competition between groups that are differently affected by the collective action and we show that this provides us with a unified framework for analyzing different organizational forms to solve collective action problems. We develop a theoretical model in which actors can disagree about the desirability of the collective action and can choose to spend their resources either in its support or in opposition. We assume that the individual actors' preferences can change over time so that the coalitions of supporters and opponents are constantly shifting. We further assume that these preferences are private information that actors must somehow communicate to each other in order to identify others who share them (through, for example, voting). This setting reveals an interesting problem of collective action: since its success requires that supporters devote resources sufficient to overcome any potential opposition, there are incentives to vote insincerely in order to increase the chances of the preferred outcome or in order to minimize one's costs.[5]

We show that it possible to ensure truthful revelation of preferences (and thus sustained cooperation) in the usual way: a coercive mechanism that punishes acts contrary to one's statements of intent. We explore two alternative organizational forms. In both, players agree on a voting rule — a minimum number of supporting votes that must be cast before the action is implemented — and then in each period, they vote after observing privately their current preferences. If the vote fails to clear the support threshold, no collective action is taken and players consume privately. The organizations differ in what happens when the quota is met. In the first, which we call a *coalition of the willing*, only players who vote in support of the action contribute to its implementation. In the second, which we call a *universal organization*, everybody does.[6]

We find that in the coalitions of the willing, the equilibrium voting quota is at the complete information social optimum (the minimum number of supporters for the action to become socially desirable) as long as the probability that there will be sufficient support for the action is not too high. This rule, which is what the actors would have agreed upon had they had complete information, is socially efficient even *ex post* (after the uncertainty is resolved by the vote). However, as the probability for support increases, the incentives for supporters to free-ride by falsifying their vote become stronger. The only way of keeping them honest is to increase the risks of such attempts by requiring a larger quota, which drives the equilibrium voting rule away from the social optimum. It is because of this inefficiency that we consider universal organizations, where both supporters and opponents agree to contribute to the common action. We find that the equilibrium quota for this orga-

[5]Opponents could pretend to be supporters and then undermine the action by failing to contribute what the other supporters expect them to. Supporters, on the other hand, could pretend to be opponents (if they believe the quota will be reached without them) to avoid paying towards the collective action.

[6]Empirically, coalitions of the willing can be observed in United Nations Framework Convention on Climate Change, the European Monetary Union, the Concert of Europe, and the US-led alliances in the two Persian Gulf Wars. In this paper 'universal' refers to the idea that both supporters and opponents contribute to the organization, not to a geographical definition. Some examples for this organizational form are the World Trade Organization, regional trade agreements, and the International Whaling Commission.

nization is always at the social optimum regardless of uncertainty, so it does not suffer from the inefficiency inherent in a coalition of the willing. This might make the latter a puzzling phenomenon but we show that there are circumstances under which players might prefer to organize in coalitions of the willing *ex ante* anyway.

Although these coercive organizations can endogenize cooperation by making it self-enforcing, they have the same problems as the institutional cooperative solutions in repeated-PD settings: they make heavy informational demands on monitoring, and they are plagued with credibility problems which only become more severe as the benefits of cooperation increase (because punishment in form of indefinite exclusion from the institution means even greater losses).[7] Most importantly, both of these organizational forms require long shadows of the future for successful implementation. This suggests that we might wish to explore an organizational form that does not rely on coercive threats to maintain cooperation. In this type of organization, which we call an *agent-implementing organization*, the players hire an agent who is neutral with respect to the outcome of the action, and agree on a voting rule and on the pre-determined contributions to make each period. If the vote clears the support threshold, the agent implements the action, and if it does not, it disburses the contributions (net its operating costs) back to the actors.[8] We show that this organizational form, which is independent of the shadow of the future, could be quite attractive and players might be willing to spend very large portions of their endowments to maintain it when none of the alternatives are viable. This is so even though we assume no special informational or expertise advantages for the agent over the players.

Overall, then, we show that there exist circumstances that make each of the three organizational forms better than the other two, which might help account for the fact that empirically we observe all three. Conceiving of international cooperation as a between-groups problem thus provides a unified framework for analyzing alternative forms of cooperation with new insights into the organization of international collective action. First, the theory provides a novel rationale for delegation. The prevalent view treats agents as reinforcing the positive effects of repeated interaction by increasing information flows (to help with monitoring), reducing transaction costs (to help with credibility), or providing expertise (to help with the cooperative benefit) in order to provide a solution to a PD-like problem.[9] Whereas all of this is doubtless important, we uncover a very different reason for delegation: it can help achieve stable cooperation precisely because such an organization does not require enforcement threats. That is, instead of seeing agents as facilitating the use of such threats, we see them as obviating the need to rely on coercive measures to sustain cooperation. Because the success of the collective action does not depend on sanctioning undesirable behavior, there is no need to monitor participants very closely. Hence, one need not assume that the

[7]For example, the European Monetary Union (EMU) does not have formal rules that allow members to exclude countries that break the fiscal rules of the Union despite the impact such deviations have on other members' economies and social policies. Even if Greece were to be expelled from the EMU, as some of its members argued because of Greece's persistent breach of EMU's fiscal rules, long-term economic gains from exchange-rate stability would most likely prevent them from doing so indefinitely.

[8]Some examples of organizations that share most of these features are the International Monetary Fund, United Nations peacekeeping, and multilateral and regional aid institutions such as the World Bank or the African Development Bank.

[9]Pollack (1997); Abbott and Snidal (1998); Epstein and O'Halloran (1999); Nielson and Tierney (2003); Hawkins et al. (2006).

agent should have any informational advantages. The finding that cooperation with delegation is independent of the shadow of the future also sharply contrasts with the widespread emphasis on the importance of that shadow. Second, we show that one of the major driving forces behind the form organizations take is the need to ensure that players reveal their preferences truthfully through voting. In other words, our approach also rationalizes the use of voting in IOs, something that is quite puzzling if one treats IOs as merely implementing informal institutions. In particular, we show that it is possible to make this voting meaningful even in an environment where the members themselves are not bound by the outcome of the vote.

2 Institutions and Organizations

The following discussion is primarily intended to motivate the assumptions of our model by substantiating three major claims. First, the uneven distribution of externalities from international action means that international "cooperation" to implement it often has both supporters and opponents. Second, these groups of supporters and opponents can "invest" resources either to facilitate that action or hinder its implementation. Third, the changing international environment alters the distribution of externalities and thus the preferences of the actors over time, so the memberships in these two groups can be unstable.

International cooperation theory has concerned itself mainly with the dilemmas of cooperation that arise in PD-like situations within the group of supporters. While focusing exclusively on the behavior of the group that is interested in cooperating, this approach assumes that everyone would be better off if all cooperated, but that each individual actor is strictly better off if it free-rides on the efforts of others. As a result, the collective good is frequently under-provided.[10] In this framework, the main problem of cooperation is how to decrease the incentives to free-ride among supporters. Despite the significant merits of this approach, its main short-coming is that it neglects the simple fact that "cooperation" among those that want the action to take place might well mean "conflict" from the perspective of those that do not.[11] Collective action often creates positive externalities for some actors but negative externalities for others. For example, increasing trade cooperation through enlargement of the World Trade Organization (WTO) might mean very different things to different existing WTO members. Some states (e.g., those with strong import or export interests in the newcomers) gain from the enlargement. Others (e.g., those that experience stiffening of export competition to major export markets after accession) might well lose from that cooperative action.[12] When China was applying for membership in the 1990s, Mexico expected large negative externalities from the opening of US markets to Chinese products. Since Mexico and China have very similar export structures, Mexico feared a significant increase in competition for exports to the US. Such negative externalities are not restricted to international economic cooperation. In the realm of collective security, opposition to international peacekeeping (a form of cooperation) can emerge if these missions

[10] Larson (1987); Rhodes (1989); Downs and Rocke (1990, 1995); Evangelista (1990); Martin (1992); Fearon (1998); Downs, Rocke, and Barsoom (1998); Gilligan (2004); Voeten (2005); Svolik (2006).

[11] Gruber (2000).

[12] Kraft (2006); Neumayer (2010). This does not even consider political considerations that might impinge on one's preferences regarding a particular economic action.

produce negative externalities for some governments. For example, when the UN Security Council discussed intervention in Rwanda, the Rwandan nonpermanent representative to the Security Council was part of the extremist Hutu government that had planned the genocide. It therefore had strong objections to a UN-led intervention.[13] More generally, governments with strategic interests in the target of intervention (or those who prefer another form of international pressure because they disagree with the leaders of the action) might well experience negative externalities should the action take place, and as a result can end up in active opposition to it.

The uneven distribution of externalities from collective action can lead to conflict between supporters and opponents, and this conflict might be quite costly. One illustration of what can happen when actors fail to avoid that costly confrontation is furnished by the attempts to regulate trade of genetically modified organisms (GMOs).[14] Since the 1990s major food importers have demanded international regulation on the increasing trade of GMOs. One of the main supporters of international regulations of GMOs is the European Union. The EU calls for the international adoption of the "precautionary principle" which legitimizes the restriction of GMO trade even without compelling scientific evidence about possible health and safety hazards. Major exporters of GMO products, on the other hand, fear tremendous costs from the implementation of this principle because it implies a *de facto* embargo on GMO imports in states that decide to regulate GMO trade. The unilateral regulation of GMOs in the EU since the 1990s has led to a decline of American GM corn imports from about $211 million in 1997 to merely $0.5 million in 2005. Similarly, GM soybean exports fell from $2.3 billion in 1997 to $0.51 billion in 2005 (Peterson, 2009). These losses are expected to continue particularly after the EU negotiated the Cartagena Protocol on Biosafety which created an opportunity for states to pursue collective action according to the precautionary principle outside of the WTO framework. As of December 2010, 159 states have signed or ratified the Protocol, and that includes all members of the EU, major importers of agricultural products like China, Egypt, India, and Japan, and many African counties that import food or receive food aid from the United States.

In an attempt to avoid these negative externalities, the United States (together with some other GMO exporters) lobbied in favor of keeping the existing "sound science principle" (which rejects biosafety regulation without compelling scientific evidence), and invested heavily in sabotaging the efforts of the European Union. The US vetoed the adoption of regulations within the WTO and other international bodies. When the EU attempted to mobilize support for its position, the US heavily criticized its actions, and even accused the EU of having threatened developing countries with denial of aid should they accept GMO food aid.[15] Most drastically, the US government initiated a trade dispute within the WTO to stop GMO regulation, and then started itself to put serious pressure on African countries to abide by that position. Many of these countries had implemented some sort of regulation on GMO trade and had requested that aid donors send non-GMO food or, if that were not possible, to give cash instead of in-kind aid. Although many donors complied with these requests, the United States did not. Instead, it used its preeminent position as a major food

[13] Barnett (2003).

[14] For an excellent summary of international regulation of GMOs see Pollack and Shaffer (2009).

[15] Clapp (2004).

donor in the region to threaten to cut off aid completely unless the recipients abandoned the regulations. Faced with famine, many of these African states complied and waived the restrictions on GMO trade they had implemented. The US also went further in trying to generate African support for its WTO complaint. When the Egyptian government, which had initially supported the US, decided to withdraw from the complaint, the US retaliated by pulling out of the free trade agreement talks. Thus, fighting against the negative externalities that the EU policy would have imposed cost the US both political and economic capital.

The European Union itself had to spend significant resources in the fight against the US position. It had to invest heavily in institution-building projects in African developing countries to offset the potential loss of American aid. It also threatened to ban imports of agricultural products from countries that used GMOs in an attempt to rally support for the precautionary principle and overcome American opposition.

Thus, the conflict between the United States and the European Union about the desirability of the international collective action regarding trade in GMOs proved quite costly to both sides. Part of the problem that prevented a workable compromise was the uncertainty about the preferences of many countries. Although many African states were concerned about GMO trade, they were also highly dependent on American food aid. It was by no means clear either to the US or the EU which of these factors would weigh most heavily on their final position. Moreover, uncertainty can also arise from the variability of future preferences.[16] For example, if a country allocates more land towards GMO production, it will become less supportive of the Cartagena Protocol. Similarly, changes in public opinion, which might not be easy to predict, can affect the degree of support for international cooperation. For instance, the increasing public awareness of climate change has arguably affected the willingness of governments to support climate change agreements on the international level.

In the situation described above, the success of international cooperation does *not* depend on the effectiveness of enforcement mechanisms, but on the ability of supporters to overcome the opposition to cooperation in a situation in which preferences are private information (e.g. the U.S. expected Egypt to back it up against Cartagena, but Egypt withdrew from the WTO complaint). In other words, actors must identify each other's preferences through some form of communication in order to organize into groups of supporters and opponents that can then coordinate on some policy. Once these groups are identified, one can use its superior resources to impose a solution on the other. If the opponents prevail, the collective action will not take place, and if the supporters do, then it will. Conflict usually entails significant waste as resources are dissipated on the fight rather than on the implementation of the desired action. In the GMO case, both the EU and the US had to spend resources on ensuring sufficient third-party support for their position — one had to spend in order to overcome the opposition created by the spending of the other. Clearly, this type of imposition is very burdensome for both sides and there are strong incentives to find a way to avoid paying its costs. The GMO case, then, shows what can happen when actors fail to organize themselves in order to avoid the costs of conflict. It is against this background — of what will happen without a mechanism to avoid the "brute-force" imposition — that actors must consider how to coordinate on an issue where international cooperation has

[16]Downs, Rocke, and Barsoom (1998).

uneven distributional consequences.

The GMO case motivates the assumptions we make in our model about the "anarchic" context in which actors must organize. We study three ways in which they can coordinate the conflictual and cooperative aspects of the situation to achieve outcomes that are Pareto superior to the type of fighting that the GMO case exemplifies. We must make clear that we do not attempt to explain under what conditions actors might fail to coordinate on one of these organizational forms; indeed, this is an important and fascinating topic for future study.[17] Instead, we seek to show what kind of alternatives they might consider and explore the conditions under which they might prefer one to the other two.

We start with the fundamental uncertainty about the distribution of preferences regarding the collective action, which necessitates some form of communication that actors can use to reveal them. We focus on voting as the simplest form of communication that can indicate support or opposition for an action. We find that in an environment where one is not bound either by the collective outcome of the vote (one can still act contrary to what the required majority has indicated) or by the implication of one's own vote (one can act contrary to the support or opposition indicated with one's vote), making this communication meaningful can be difficult. The institutional design must ensure both that voting is sincere and that the subsequent action implied by the vote is self-enforcing. It is precisely to the study of the the ways of doing so that we now turn.

3 The Stage Game

There are N players, each endowed with 1 unit of resource, who might want to take a collective action.[18] The action produces a discrete public outcome, $a \geq 2$, and players differ in their valuation of that outcome. The action succeeds only if at least $\theta > 1$ resources are dedicated to it, and it fails (if attempted) otherwise. If the action is taken, the individual payoff is:

$$u_i = 1 - x_i + \pi v_i a,$$

where $x_i \in [0, 1]$ is i's spending in support or opposition of the action, $v_i \in \{-1, 1\}$ is i's valuation of the benefit, and π is the probability that a is produced. Observe that if $v_i = -1$, the individual prefers that the action is not taken, so we shall call him an *opponent*, and if $v_i = 1$, he prefers that it is, so we shall call him a *supporter*. Since individuals might have opposing preferences over the desirability of the action, they might choose to dedicate their resources either in support of its success or against it. We assume a simple "technology of conflict" in which whether the action will succeed depends on the difference between the resources dedicated in its support and the resources dedicated against it. Let $\mathcal{S}$ be the set of supporters and $\mathcal{O}$ be the set of opponents. Then $X = \sum_{i \in \mathcal{S}} x_i$ are the resources devoted to

[17] However, our model does suggest one possibility: short shadows of the future that make the two coercive mechanisms impossible to implement and large transaction costs that put the third, delegated, solution out of reach. We think that overt conflict of the type that occurred in the GMO case should be relatively rare because most international interaction would be successfully organized to avoid it. Still, we can think of at least one other similar situation: the conflict over the Kyoto Protocol. Schneider and Urpelainen (2010) study the causes of such conflicts in general and provide an extended discussion of the Cartagena Protocol case in particular.

[18] A mathematical appendix with all proofs is available from the authors.

action by its supporters, and $Y = \sum_{i \in \mathcal{O}} x_i$ are the resources devoted against the action by its opponents. Given any positive integer $\theta \in \mathbb{Z}$, the probability that a is produced is:

$$\pi = \begin{cases} 1 & \text{if } X - Y \geq \theta \\ 0 & \text{if } X - Y \leq \theta - 1 \\ {}^1\!/_2 & \text{otherwise.} \end{cases}$$

The idea is that if supporters out-spend opponents enough to meet the minimum costs of action, then the action will take place.[19] With this specification and the assumption that $a \geq 2$, it always pays for an individual to spend his entire resource if doing so meant he would obtain the preferred outcome on a with certainty. Assume that $\theta \leq N$ or else the action is infeasible because it is beyond the means of the entire collective.

We will consider situations in which players can (costlessly) coordinate their actions so that supporters and opponents can act as groups. Let S denote the number of supporters and $N - S$ denote the number of opponents. The supporters can *impose* the action if they have enough resources to pay for it and overcome any opposition, which will be the case when $S - (N - S) \geq \theta$, or when

$$S \geq \left\lceil \frac{N + \theta}{2} \right\rceil \equiv S_c.$$

Since $S_c \geq \theta$, the action is feasible when it can be imposed. When $S < S_c$, then opponents can *impose* the status quo on the supporters. There is always an equilibrium in which nobody invests anything and the action does not take place: since no supporter can unilaterally implement the action, if no supporter is expected to contribute, then no supporter would be willing to deviate by contributing (uselessly). Since the status quo remains in place at no cost, no opponent would deviate to costly defensive measures.

Whether there are other equilibria and what form they take depends on the timing of investments. It is not difficult to see that when it comes to committing one's resources, the two groups are in a war of attrition: each wants the other one to commit first so that it can tailor its own contribution appropriately.[20] We have several substantive reasons for choosing the model in which supporters commit before opponents do. First, it might make sense that since in the absence of action the status quo (that opponents like) prevails, the onus on changing this is on the supporters who would have to take the initiative. Second, if the action takes time to occur, then opponents can always wait for supporters to make their irreversible investment before committing their resources one way or the other. While supporters want opponents to move first, the premium on taking the action might compel them not to wait

[19] The first-past the post contest-success function is common, but we introduced the additional assumptions to ensure that there are optimal solutions. If, for instance, we assume that $\pi = 1$ when $X - Y \geq \theta$ and 0 otherwise, there will be no optimal way to block the action (although there will be an optimal way to take it). That's because contributions are costly, and opponents can get arbitrarily close to θ from below while blocking the action and reducing their own contributions. The same problem crops up with $\pi = 1/2$ when $X - Y = \theta$ only. Then, if there is any slack in resource contribution on either side, an arbitrarily small increase in contribution would discontinuously tip the scales in favor of that side. With our specification, there is a flat region in the neighborhood around θ where uncertainty prevails.

[20] More details available from the authors. With simultaneous moves, there is no coalition-proof pure-strategy Nash equilibrium in any non-trivial environment.

too long (especially if the action becomes ineffective past some deadline). Third, as we shall see, the issue becomes rather moot when there is uncertainty over the preferences of players: before any group can move, its membership would have to be ascertained. Thus, we now focus on the model where the supporters move first.

This single-shot interaction has two types of pure-strategy subgame perfect equilibria, of which only one is coalition-proof when $S \geq S_c$. Recall that a coalition-proof Nash equilibrium is one where there exists no proper subset of players that can benefit by deviating as a group, and where the group itself is not vulnerable to sub-group deviations. Let's start with $S < S_c$, in which case opponents have a credible threat (as a group) to block the action. In the subgame after supporters have invested some X, opponents can always form some group that can block it. Any subset of $\mathcal{O}$ with size $G_o \leq N - S$ and the property that $G_o \leq X + \theta - 1$ would work and would be self-enforcing if all members spend exactly $x = (X + \theta - 1)/G_o$ each and all non-member opponents spend nothing. The minimum-cost of such coalition is, of course, the grand coalition of opponents. Any deviation by a contributor would increase the probability of action from 0 to ½, and any contribution by a non-contributor would not affect the zero-probability. Given that the action will always be blocked by some coalition, supporters have no incentive to spend anything, in coalitions or otherwise. The unique symmetric coalition-proof SPE is thus for nobody to spend anything along the path, and for the grand coalition of opponents to block any positive spending off the path.

The zero-spending equilibrium is also subgame-perfect when $S \geq S_c$ but is not coalition-proof. To see this, observe that if no supporter spends anything, no individual can implement the action by himself, so there is no incentive to contribute either. Given the zero contribution by supporters, consuming privately is a best response for the opponents. Off the path, opponents can credibly threaten to block any action such that $X - (N - S) < \theta$ (by forming a successful blocking coalition in the manner described above), and would consume privately otherwise (because there is no sense in spending against an action that will take place regardless). This means that no coalitional deviation that fails to invest more than that amount can succeed. However, since $S \geq S_c$ implies that, at the very least, the grand coalition of supporters can impose the action, it follows that there exists some subset of $\mathcal{S}$ of size $G_s \leq S$, that can profitably deviate by spending exactly $X = \theta + (N - S)$ and ensuring success. Moreover, this coalition is self-enforcing if each member contributes precisely $[\theta + (N - S)]/G_s$. Thus, the zero-contribution SPE is not coalition-proof. Conversely, any SPE in which some group of size G_s spends this amount is coalition-proof: if any member contributed less, the action will fail, and any non-contributor benefits from the action for free. The only possible coalitional deviation would involve some subset of contributors, but they are already spending the minimum necessary for the action. Since players are symmetric, there is no reason to expect that anything but the grand coalition of supporters should form. Thus, we shall focus on the minimum-cost coalition-proof SPE where all supporters contribute their equal shares.

PROPOSITION 1. *The stage game has a unique symmetric coalition-proof subgame perfect equilibrium. If $S < S_c$, then every player consumes privately and the status quo prevails. If $S \geq S_c$, then each supporter spends $x_c = (N + \theta)/S - 1$, opponents consume privately, and the action takes place.* □

Proof. The optimality of on-the-path strategies is established in the text. Off the path, any positive spending by supporters in the private spending equilibrium is successfully blocked by a minimum-cost coalition of opponents. To impose the action, the minimum-cost coalition with symmetric contributions must spend $x_c S - (N - S) = \theta$, which yields the value of x_c specified in the proposition. Note that when $S \geq S_c$, this spending is feasible. Any lower investment by supporters in the implementation equilibrium is successfully blocked by a minimum-cost coalition of opponents. ∎

This result illustrates the "brute-force" resolution of the problem of collective action, and is the type of "solution" that can arise when actors do not coordinate to avoid it (e.g., the GMO case). This behavior is socially inefficient: (i) the action fails to take place when it should from social welfare perspective, and (ii) when it does take place, resources are wasted on deterring opponents. It is easy to see the latter because supporters must spend enough to deter potential blocking coalitions even though these do not form on the equilibrium path. The other problem is that the status quo might persist in situations when it would be socially beneficial to implement the action. To see this, suppose there existed a planner who simply maximized social welfare and who could implement the action at cost while (costlessly) enforcing his decision. Since he can always maintain the status quo, society is guaranteed the income from private consumption whenever he chooses not to implement the action. The social welfare then will be N.

When would he implement the action? The planner could choose to tax either supporters only or everyone, at a flat rate that collects just enough resources to pay for the action. Social welfare from implementation will be the same, $N + a(2S - N) - \theta$, in either case.[21] The planner will act when doing so is at least as good as remaining with the status quo, or whenever:

$$S \geq \left\lceil \frac{N + \theta/a}{2} \right\rceil \equiv \mathfrak{S}.$$

Since $S_c - \mathfrak{S} = \theta(1 - 1/a)/2 > 0$, the coalition size required to implement the action in equilibrium is always strictly greater than what the societal optimum would be. Moreover, this difference increases in the costs of the action, θ, and in the magnitude of its externalities, a. The one thing that the SPE has going for it is that at least the action *can* take place when there is sufficient support. This, however, turns out to be a consequence of the complete information assumption.

4 The Problem of Voting: Why Have It and How to Make It Meaningful?

The assumption of common knowledge of preferences is, of course, very strong. More realistically, a player might learn his own preferences about taking the action in a particular period but the actual distribution of preferences among the rest of the players would not be common knowledge; players can only estimate it from the priors. To model this, let each

[21] If only supporters are taxed, the rate would be $x = \theta/S$, and social welfare will be $S(1 + a - x) + (N - S)(1 - a) = N + a(2S - N) - \theta$. If everyone is taxed, the rate would be $x = \theta/N$, and social welfare will be $S(1 - a - x) + (N - S)(1 - a - x) = N + a(2S - N) - \theta$ as well. Of course, if supporters must shoulder the costs by themselves, it is also necessary that $\mathfrak{S} \geq \theta$ or else the action would be infeasible. This constraint does not arise if all players are taxed because $\theta < N$ by assumption.

player privately observe v_i, and only know that individual realizations are randomly and independently drawn from a distribution with $p \in (0, 1)$ being the probability of $v = 1$, and $1 - p$ being the probability of $v = -1$. This distribution is common knowledge. From any player's perspective, the probability that there are exactly k supporters among the remaining $N - 1$ players is:

$$f(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

Since the private values are independently drawn, learning one's own value tells a player nothing about the other players.

Let us begin with a solution without communication among the players. Since the identity of supporters (and opponents) is not known, players cannot coordinate in the way they could before. In particular, we now have to assume that everybody moves simultaneously. Thus, in each period players learn their preferences privately, and then simultaneously decide how much to spend. As with complete information, there is always the equilibrium where nobody spends anything regardless of their private realizations. An opponent is happy when the action will not take place and has no reason to spend on deterrence given the equilibrium expectations. A supporter cannot implement the action unilaterally, and since he expects that nobody else will contribute, he does not contribute either. Since players cannot communicate, they cannot identify a coalitions of supporters that can impose the action, and as a result cannot coordinate to do so. The following result shows that this is, in fact, the only equilibrium of the game.

LEMMA 1. *The unique equilibrium without communication is for everybody to consume privately, and the outcome is always the status quo.* □

We conclude that with uncertainty the action will never take place in the absence of communication. Since players cannot coordinate without knowing about the distribution of preferences, they need to be able to reveal what their privately-known preferences are. Because the relevant bit of information is whether the player supports or opposes the action, we will consider the simplest possible form of communication: players simultaneously announce whether they support the action or oppose it, i.e., they vote. That is, we have a straightforward reason to consider voting if we conceptualize it as *a method of communicating privately-known preferences*.[22]

Consider then the model with a voting stage. Players observe their preferences privately and simultaneously vote whether to implement the action or not. Then they play the stage game as before. To maintain symmetry with the original specification, the group that voted in support moves first, followed by the group that voted in opposition. These groups may not necessarily coincide with the groups of actual supporters and opponents. (The timing implies that if everyone votes for the action or everyone votes against it, all players move

[22]Since the voting outcome is not binding in our environment, we will assume that players can decide on their allocations regardless of the outcome of the vote (they cannot be forced to abide by it). In this sense, voting in this environment is not for preference aggregation that would result in some sort of social choice. Rather, it is a primitive form of communication. As we shall see later, the voting outcome *can* be made meaningful through endogenous enforcement schemes.

simultaneously.) The equilibrium from the game without communication exists here as well (players simply do not condition behavior on the voting outcome). The question is what they can do if they do pay attention to the voting. More to the point, can the voting reveal their preferences? The following result shows that it cannot.

LEMMA 2. *Sincere voting cannot be sustained in equilibrium for any $S_c < N$.* □

The crux of the problem with the voting scheme in the single-shot interaction is that when voting outcomes are not binding on the collective, there is no cost for casting one's vote one way and then acting contrary to it.[23] Thus, whereas there is a dire need to make preferences common knowledge through voting, there is no way of doing so in this environment. The lack of enforcement creates incentives to falsify one's vote. However, since acting contrary to one's vote is observable, it could be subject to possible retaliation in the future if the interaction is repeated. We now show that the traditional approach to overcoming such problems through endogenous enforcement that relies on punishment strategies can be employed to ensure that voting is sincere.

4.1 Coalitions of the Willing

The first institution we examine parallels the solution under complete information: whenever an action is to be implemented, only the (self-identified) supporters contribute toward it; it is a "coalition of the willing." Consider the super-game which begins with players choosing a quota, Q, which will be the minimum number of supporting votes before an action can take place. The game then continues indefinitely with repeated plays of the original stage game with non-binding voting and asymmetric information. The realization of preferences is independent between periods. Let $\delta \in (0, 1)$ be the common discount factor. Players maximize the discounted sum of per-period payoffs.

Recall the private consumption equilibrium of the stage game exists even when voting is possible. This means that the repeated game has an SPE in which players always consume privately (that is, they always play the stage-game SPE). We shall use this as the reversion equilibrium so it functions as punishment that might enforce sincere voting. Since each player expects a per-period payoff of 1 in this equilibrium regardless of preferences, the SPE threat payoff is $1/(1 - \delta)$. The following proposition establishes the necessary and sufficient conditions under which the threat of reverting to this SPE can sustain minimum-cost implementation in SPE with sincere voting.

PROPOSITION 2. *Fix some Q ($\theta \le Q \le N$), let $x(q) = \theta/q$, and let*

$$\underline{\delta}_{\mathrm{w}}(Q) = \frac{a}{a + \zeta_{\mathrm{w}}(Q)}, \tag{1}$$

where $\zeta_{\mathrm{w}}(Q) = p\,(a - x(Q))\,f(Q-1) + \sum_{k=Q}^{N-1} \left[(2p-1)a - px(k+1)\right] f(k)$. *The*

[23] The only situation where sincere voting can work is the special case with $S_c = N$, which implies $\theta = N$; that is, all players must spend everything just to pay for the action. This is such a knife-edge condition that we will ignore.

following strategies constitute an SPE for all $\delta \geq \underline{\delta}_{\mathrm{w}}(Q)$ *if, and only if,* $\zeta_{\mathrm{w}}(Q) > 0$ *and*

$$\underbrace{af(Q-1)}_{\textit{benefit of sincerity}} \geq \underbrace{\sum_{k=Q-1}^{N-1} x(k+1)f(k).}_{\textit{cost of sincerity}} \tag{SC}$$

In each period players vote sincerely; if there are $q \geq Q$ *votes in favor of the action, supporters spend* $x(q)$ *each and opponents consume privately; otherwise everyone consumes privately. If the action ever fails when it is supposed to take place (because some player who voted for it fails to contribute or because it is blocked by opponents) or is imposed when it is not, players revert to the unconditional SPE with private consumption. If the action is attempted and the opponents can block it, then they block it and the game continues as before. The equilibrium period payoff is* $1 + \zeta_{\mathrm{w}}(Q)$*.* □

As we shall see, it is always possible to find a quota that can satisfy all conditions. Repeated interaction can coerce sincere voting by threatening retaliation for acting contrary to one's vote. Although this institution can support cooperation, it has at least two problems even in the highly permissive environment which ignores monitoring and coordination costs. First, the institution must guard against two types of deviations: opponents derailing the action at the implementation stage, and supporters trying to free-ride by pretending to be opponents and enjoying the benefits without incurring the costs. The first deviation is observable, so players can implement the conditional punishment to deter it. Given an optimal quota, the high discount factor guarantees that the punishment will be sufficiently severe to keep opponents in line. Thus, the informal institution can protect against such the first type of deviation provided future cooperation is sufficiently valuable. The other type of defection is far harder to guard against because it cannot be observed: the supporter votes against the action and then consumes privately regardless of the outcome. This behavior is identical with that of opponents, so there is no way to implement conditional punishments to protect against it. There is no threat-based solution for this problem: condition (SC) in Proposition 2, the "sincerity constraint", shows what is necessary and sufficient for it not be so severe as to derail cooperation. To see this more clearly, note that if an opponent votes sincerely, the only benefit of doing so is that the action will be implemented if he turns out to be pivotal because without his vote it will then fail. In all other cases his vote does not affect the probability of the outcome but it does affect the costs to the player should the action be voted for implementation. Since the player must contribute when he supports the action (failing to do so will derail the action immediately for sure), the cost of sincerity is the expected contribution he will have to make for all those situations when the quota will be met. Thus, sincere voting by a supporter can be sustained if, and only if, its benefits outweigh its costs, which is what (SC) ensures.

As it turns out, the sincerity constraint can be severely binding, especially when the probability of being a supporter is moderate to high. To show this, we now examine the optimal quota, which maximizes $U(Q) = 1 + \zeta_{\mathrm{w}}(Q)$ under the two constraints that $Q \geq \theta$ and (SC).

LEMMA 3. *The optimal quota for a coalition of the willing is* $Q_w = \max\{\theta, \mathfrak{S} + n(p)\}$, *where* $n(p) \geq 0$ *is the smallest integer such that* $\mathfrak{S} + n(p)$ *satisfies the sincere voting constraint in* (SC)*. The stepping function* $n(p)$ *is non-decreasing.* □

When p is sufficiently low, the optimal quota is at the societal optimum, $\mathfrak{S}$, provided a group of that size can implement the action, or at the smallest such group that can, θ, otherwise. (This constraint would also bind if the social planner taxed supporters only.) However, as p increases, so does the optimal quota, in a stepwise manner with discontinuous jumps. In these cases, condition (SC) binds and forces the quota up and away from the social optimum, as illustrated in Figure 1. The vertical line marks the smallest value for p for which the sincerity constraint binds: for all p less than that value $n(p) = 0$, so $Q_w = \mathfrak{S}$.[24]

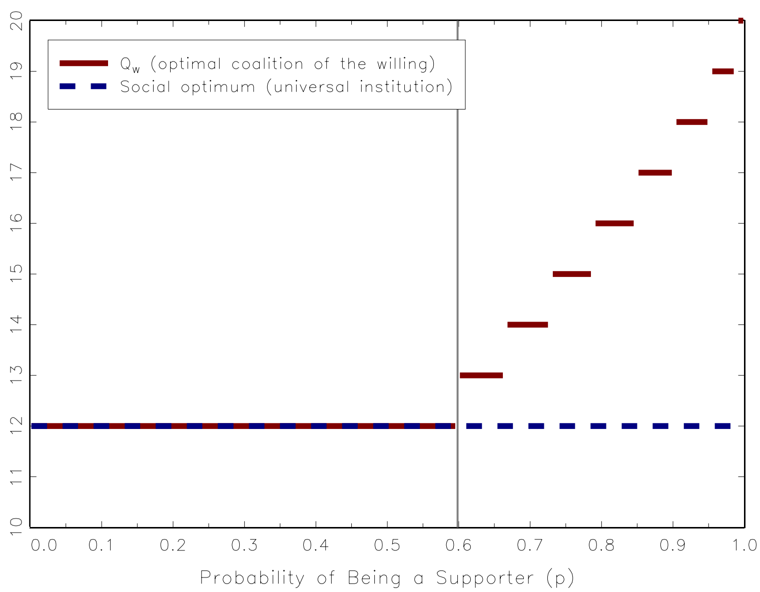


Figure 1: Coalitions of the Willing and the Social Optimum ($N = 20$, $a = 3$, $\theta = 11$).

Why does this happen? As the probability of support increases, the likelihood that any one player would be pivotal for any given quota decreases. This increases the temptation to free-ride because it reduces the expected cost of voting against the action. The only way to overcome this problem is to increase the quota: doing so reduces the expected benefit of free-riding because it decreases the probability that the action would take place without one's vote. This restores the incentive to vote sincerely but as p increases further, the problem re-appears and the quota must be adjusted again. In this way, the sincerity constraint drives the optimal quota further away from what is socially desirable. Somewhat paradoxically, as the number of actors that might be supportive of the action increases, the institution, in which only the coalition of the (self-identified) willing contributes to the action, becomes ever less socially efficient.

[24]This can be easily computed by solving (SC) for p as an equation at $\mathfrak{S}$.

This social inefficiency and its causal source, the incentive for supporters to free-ride, suggest that it might be beneficial to organize cooperation differently. The first problem is that concentrating the costs on the group of cooperators precludes socially desirable outcomes because doing so puts expensive actions out of reach. The second problem is that a supporter might have incentives to distort his vote in attempt to conserve his resources. An institution with universal contributions might help with both problems: it spreads the costs among all players, and since one has to contribute whenever the action is voted to take place regardless of whether one voted for it or against it, there should be no incentive to distort a supporting vote.

4.2 Universal Decentralized Contributions

We now consider an institution with universal contributions: one, where each player — supporter and opponent alike — is supposed to contribute whenever the agreed-upon quota is met. This changes nothing in the single-shot interaction: there is no reason for opponents to abide by the outcome if they can block it, and there is no reason for supporters not to impose it if they identify themselves. This destroys the incentives for sincere voting and private consumption remains the sole equilibrium outcome. However, since every relevant deviation is now observable, it can be subject to collective punishment when the interaction is repeated.

Consider then the same model but with different strategies. Actors agree on a quota, Q, and promise to spend amount, $x = \theta/N$ each, if the action garners $q \geq Q$ votes in favor. As before, in each period, they observe their private values, vote, and then act depending on the outcome of the vote. If any player fails to contribute when $q \geq Q$ or if the supporters force the action even though $q < Q$, then they revert to the stage-game Nash equilibrium where everyone always consumes privately. The following result establishes the necessary and sufficient conditions to support a cooperative SPE with sincere voting and universal contributions.

PROPOSITION 3. *Fix some quota Q $(1 \leq Q \leq N)$, let $x = \theta/N$, and let*

$$\underline{\delta}_{\mathrm{u}}(Q) = \frac{a + x}{a + x + \zeta_{\mathrm{u}}(Q)}, \tag{2}$$

where $\zeta_{\mathrm{u}}(Q) = p(a - x) f(Q - 1) + [(2p - 1)a - x](1 - F(Q - 1))$. The following strategies constitute an SPE for any $\delta \geq \underline{\delta}_{\mathrm{u}}(Q)$ if, and only if, $\zeta_{\mathrm{u}}(Q) > 0$. In each period players vote sincerely; if there are $q \geq Q$ votes in favor of the action, then each player spends x and it gets implemented, otherwise everyone consumes privately. If some player fails to contribute what they are supposed to or if the action gets implemented when $q < Q$, players revert to the unconditional SPE with private consumption. If the action is attempted with $q < Q$ but opponents can block it, then they block it and the game continues as before. The equilibrium period payoff is $1 + \zeta_{\mathrm{u}}(Q)$. □

Observe that there is no analogue to (SC); that is we no longer need a special condition to exclude free-riding by supporters. The reason is simple: a supporter who votes against the action lowers the probability of implementation (by the probability that he is pivotal) but does not save on his contribution for all those cases where the action will go forward regardless of his vote. Thus, the only relevant comparison for him involves private consumption by derailing the action and getting it implemented, and we know that in such situations supporters always prefer implementation. The universal institution can alleviate the free-riding problem. Since this was the cause of the sincerity constraint to bind with the coalitions of the willing, there should be no such lower bound on the optimal quota here. Furthermore, since everyone contributes once the action is voted for implementation, there is no constraint implied by its costliness. In other words, there should be nothing to force the quota of the universal institution away from the social optimum. Indeed, this turns out to be the case, as the following result shows.

LEMMA 4. *The optimal quota for the universal institution is $Q_{\mathrm{u}} = \mathfrak{S}$ regardless of p, and is always socially optimal even* ex post. □

The optimal quota for a universal institution is not just independent of the uncertainty, it is socially optimal even after the uncertainty is removed by the act of voting. It is worth emphasizing this finding because asymmetric information usually induces serious *ex post* inefficiencies (as indeed it does with the coalition of the willing). Lemma 4, however, shows that a universal institution does not have to suffer from this problem. The intuition is that in that institution the quota is selected to maximize the difference between the private consumption outcome and the expected outcome when everyone chips in to pay for the action. In the latter, each player expects to pay the cost when the action is taken, removing any incentive to consider the likelihood of being a supporter. Since the collective will bear the costs, the only relevant consideration is how many members will find the action beneficial. In other words, precisely what the value of $\mathfrak{S}$ gives us.

Does this mean that players would always opt for the universal institution over a coalition of the willing? The answer, it turns out, is negative.

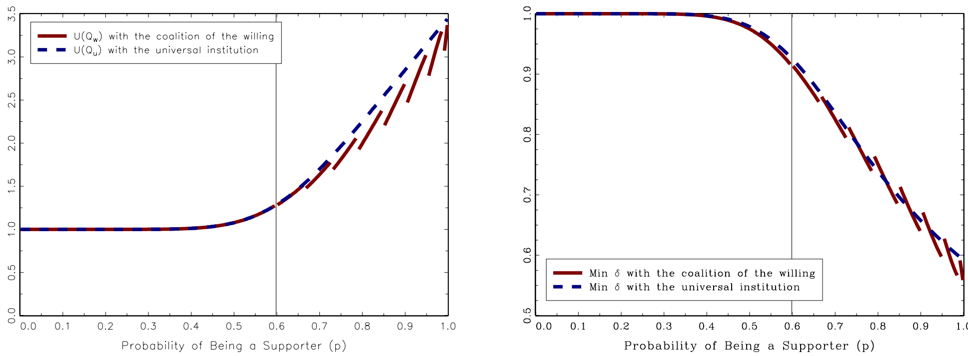
4.3 The Organization of Coercive Cooperation

Since the universal institution is socially optimal *ex post* and because players are symmetric *ex ante*, one might think that they would never choose to organize as coalitions of the willing. Indeed, when it comes to the expected payoff, the universal institution is always at least as good as the coalition of the willing, and often strictly better. However, it requires a longer shadow of the future to implement. Thus, when players are indifferent between the two institutions in the short-term, only the coalition of the willing might be feasible.

We develop this argument in several steps. First, we show that if the quotas are the same, both institutional arrangements yield the same expected payoff.

PROPOSITION 4. *The expected payoff in both institutions is the same when they implement the same quota.* □

Recall from Lemma 3 that the social optimum $\mathfrak{S}$ can be supported in a coalition of the willing whenever the cost and sincerity constraints do not bind. Since this is the equilibrium quota for the universal institution, Proposition 4 immediately implies that players will be indifferent between the two in these circumstances. Since in all other situations the coalition of the willing requires a quota that is worse than the unconstrained social optimum but the universal institution does not, it follows that the latter must be strictly better. Thus, when it comes to expected payoffs, the universal institution is never worse than the coalition of the willing and is often strictly preferable. Figure 2(a) illustrates this.[25]



(a) Per-period Equilibrium Payoffs.

(b) The Shadow of the Future.

Figure 2: Coalitions of the Willing and Universal Institutions ($N = 20$, $a = 3$, $\theta = 11$).

The only reason to opt for a coalition of the willing, then, must be that the universal institution is infeasible given how much actors discount the future. Observe now that

$$\underline{\delta}_{\mathrm{u}}(\mathfrak{S}) > \underline{\delta}_{\mathrm{w}}(Q_{\mathrm{w}}) \Leftrightarrow \zeta_{\mathrm{w}}(Q_{\mathrm{w}}) > \left(\frac{a}{a+x}\right)\zeta_{\mathrm{u}}(\mathfrak{S}),$$

where $x = \theta/N$. If $Q_{\mathrm{w}} = \mathfrak{S}$, then $\zeta_{\mathrm{w}}(\mathfrak{S}) = \zeta_{\mathrm{u}}(\mathfrak{S})$ by Proposition 4, which immediately implies that the inequality holds. Thus, in these situations the discount factor required to sustain the universal institution is strictly higher than what is required to sustain a coalition of the willing.[26] When players are indifferent among the two institutions, only the coalition of the willing might be feasible if players do not care about the future sufficiently to implement the more demanding universal institution. Moreover, this problem might crop up even when the universal institution is strictly preferable: It is entirely possible for the inequality to hold even though $\zeta_{\mathrm{w}}(Q_{\mathrm{w}}) < \zeta_{\mathrm{u}}(\mathfrak{S})$ because the coefficient on the latter is less than 1. As Figure 2(b) shows, the relationship between the two thresholds can be quite involved once p forces Q_{w} away from the social optimum: for some values of p the coalition of the willing is easier to implement, and for others it is the universal institution. The overall picture, however, is clear: if players are patient enough, then the universal institution is the way to go, especially if the probability of support is not too low.

[25]The payoff in the coalition of the willing is discontinuous in p because the optimal quota increases in discrete jumps, as shown in Figure 1. Recall that when the sincerity constraint binds and the quota increases by one unit, the constraint will not bind again within some range of values for p. Over this range, the payoff increases with p until p gets to so high that the constraint binds again. This results in another sharp (discontinuous) drop, and the process repeats itself.

[26]Since we used grim trigger strategies to support cooperation, these discount factors are the least demanding; any other strategy would require a longer shadow of the future.

4.4 The Limits of Self-Enforcement

We have now identified two solutions to the problem of meaningful communication. They both make sincere voting self-enforcing with the threat to abandon cooperation if any player deviates in his actions from the way he is supposed to behave given the voting outcome. This, of course, is the traditional method of coercing cooperative behavior that might be suboptimal in the short term. As such, these solutions suffer from the familiar host of problems associated with this approach to endogenizing enforcement. We have essentially stacked the model in favor of cooperation because we (1) neglected monitoring and coordination costs, (2) assumed deviations are identifiable, and (3) ignored the fact that reversion to the private consumption SPE is not renegotiation-proof. We now note how these might affect our results.

First, we have assumed away transaction costs. The costless coordination might not be an issue in the complete information setting but it does become a tougher assumption when we move to the asymmetric information settings. Players can vote, observe voting outcomes, monitor each other's compliance, and then coordinate their contributions (or, if need be, punishments off the equilibrium path), all without paying any transaction costs. Introducing any of these considerations in the model will make the institutions harder to sustain because they will lower the expected payoff from participation. Since the reversion payoff is constant at 1 (private consumption), this will make cooperation less valuable overall, and therefore harder to sustain.

Second, and related to the first, we have assumed that actions (e.g., contributions) are perfectly observable, that there is no noise, and that the action succeeds whenever actors contribute enough to it. These permit players to identify those who attempt to free-ride or purposefully derail implementation. If the outcome of the attempt to implement the action is uncertain — i.e., there is some positive probability that it will not succeed even if all required contributions have been made — and if players can only imperfectly monitor the effort they put forward, then deviations will be harder to detect. This will increase the incentives for supporters to free-ride on the efforts of their fellows, and might even tempt opponents to derail the action hoping that the bad outcome would be attributed to the inherent riskiness of the action itself. The institution would have to account for these problems by relaxing the trigger somewhat but keeping it tight enough to prevent players from taking advantage of the noisy environment. It is not *a priori* clear whether the overall impact on the expected value of the institutions would be detrimental: on one hand, the institutions would deliver less to those who do what they are supposed to (pay when they are supporters and refrain from blocking when they are opponents and the vote goes against them) because cooperation will fail when it would not have had in the noiseless environment. On the other hand, however, the occasional bump from secret deviations from these strategies might be beneficial. At any rate, the institutions would have to be far more involved if they are to be able to handle these problems, which in turn would increase the transaction costs and make them less valuable.

Third, we used a grim trigger strategy to sustain cooperation. This type of punishment might be too severe for the other players to execute. If any opponent yields to temptation and blocks the action in one period, the game reverts to the worst possible equilibrium where the action never takes place. When the probability of being a supporter is sufficiently high (as required by the cooperative SPE), this punishment is extremely painful because the expected benefits from the cooperative solution are quite high. Thus, the players that are supposed to punish the deviant might be sorely tempted to ignore the deviation. In other words, the grim trigger might not be renegotiation-proof. The players will either have to agree to abandon cooperation or, more likely, agree to stop the punishment. The latter tactic, of course, makes the temptation to defect even stronger because it undermines the severity of the threat. It might be possible to support cooperative outcomes with renegotiation-proof punishment strategies that are not as grim but those would require players to agree to complicated schemes that allow some of them to deviate in order to balance out deviations from the equilibrium path. Moreover, any strategy of punishments milder than the grim trigger would require higher discount factors to sustain, making it more likely that the lower bounds we identified become binding.

The fundamental problem with these solutions is that they all require players to be sufficiently patient. Transaction costs, monitoring and noise, the credibility of punishment strategies — all of these issues require investments or behaviors that reduce the expected value along the cooperative path. Since compliance is enforced with threats to revert to private consumption, the lower value of cooperation makes it harder to sustain the institutions because they require even longer time horizons to deter deviations. Ultimately, *any* institution that coerces cooperation with conditional threats of future punishment would be vulnerable in this way. Thus, we want to know if it is possible to sustain cooperation without coercion: if it can be done, then there would be no need for threats, and no need to worry about how valuable the future is.

5 An Institution without Coercion: Delegation to an Agent

We now analyze whether it is possible to maintain cooperation regardless of the players' time preferences. To this end, we begin with the single-shot game: if we can find a way to obtain a cooperative equilibrium here, then we automatically obtain the result in the repeated setting by simply having players choose the stage-game equilibrium unconditionally in each period.

We propose the following formal organization. The players agree on a quota, Q, and hire an agent whose wage is $W > 0$. The agent's wage is exogenous (possibly determined by outside opportunities for employment), and each player contributes $w = W/N$ toward it. Assume that the agent has no preferences regarding the action and cannot use any resources (other than his wage) entrusted to him for private gain. Before players learn their preferences, they contribute a portion of their resources, $x_0 \in (w, 1]$, to the agent. If any player contributes less, the agent returns the contributions and the game continues as it would without him. After learning their preferences, players vote whether to implement the action. The agent invests $R = (x_0 - w)N$ toward action if the number of votes in support is at least Q, and returns $x_0 - w$ to each player otherwise. Since the action must be feasible at the maximum that the players can contribute toward it, we require that $(1 - w)N > \theta$, which we can express as $w < \overline{w}$, where $\overline{w} = 1 - \theta/N$, or else the combined cost of the formal organization and the action exceed the total resources available to the players (i.e., the organization is not feasible). To keep things consistent with the preceding analysis, the timing is as follows: after the vote takes place, the agent moves first, followed by the group that voted in support of the action, followed by the group that voted against it.

Several things about this scenario are worth noting. First, all players contribute to the agent's war chest. Since *ex ante* they are all the same, we focus on symmetric contributions. Second, the voting outcome is still not binding for the players, only for the agent. Since the agent is assumed to have no preferences for the action, he can commit to invest according to the agreed-upon voting outcome. Players, on the other hand, can still choose how to spend their resources. In particular, supporters can increase their contribution, and opponents can still try to block the action. Third, the assumption that the agent returns the contributions (net his fee) if the action fails to garner the minimum required support stacks the model *against* sincere voting because it might allow the players to use the information obtained at the voting stage after a failed vote to force the action with the resources they obtain. Fourth, we have not assumed any special expertise or informational advantages for the agent relative to the other players. That is, none of the usual rationales for delegation apply here. All the agent has to do here is comply with the voting outcome.

5.1 The Agent-Implementing Equilibrium

The first feature of this organization we must decide upon is whether players should contribute anything over their initial investment when the vote goes in favor of implementing the action. Let $x(q)$ denote that additional contribution. Claim 1 in Appendix A shows that any strategy profile with sincere voting where $x(q) > 0$ for some $q \geq Q$ is vulnerable to opponent deviations of the type that destroyed any chance for sincere voting in Lemma 2. The intuition is that whereas falsely voting in favor of the action triggers an attempt to implement it that would not have otherwise occurred (that is, when there are $Q - 1$ actual supporters), this cost can be offset by successfully derailing the implementation after all tallies that require supporters to invest a positive amount to ensure that the action takes place: they would under-invest, mistakenly believing that their number is larger, and the opponents will then have enough to derail the implementation. Thus, we shall restrict attention to equilibria in which $x(q) = 0$ for all $q \geq Q$: supporters (and opponents) consume privately their remaining resources when the action takes place.

When there are no additional contributions after the vote, it must be the case that the resources the agent controls are sufficient to overcome any opposition that might arise. If this were not so, then the group of self-identified opponents can coalesce to derail the action in spite of the favorable vote. The largest opponent group that needs to be deterred from doing so occurs at the quota Q, so it is sufficient to ensure that the agent's resources can overcome their opposition. Since there is no need to give the agent any more resources than absolutely necessary for that, it follows that $x_0(Q)$ must solve $R - (1 - x_0)(N - Q) = \theta$, which pins down the optimal initial "no-blocking" contribution (NBC), to:

$$x_0(Q) = \frac{(1 + w)N - Q + \theta}{2N - Q}. \tag{NBC}$$

Note that $x_0(Q) \leq 1$ for any $w \leq \overline{w}$, so this contribution is feasible whenever the organization itself is. Since the initial investment is sufficient for the action to take place after the affirmative vote, we shall call this an *agent-implementing* equilibrium.

Although $x_0(Q)$ is sufficient to ensure that opponents would not attempt to block the action whenever it is supposed to take place, we must also make sure that supporters do not attempt to impose the action whenever it is not supposed to take place. There should exist no $q < Q$ such that the self-identified group of q supporters can impose the action using the resources that the agent returns to the collective after the failed vote. Given any quota Q, the largest such group is $Q - 1$: if this group can be deterred from imposing the action after reimbursement, then all smaller groups will be deterred as well. Since the agent always keeps his fee, the "no-imposition" constraint (NIC) is $(1-w)(Q-1) - (1-w)(N-(Q-1)) \leq \theta - 1$, which we can rewrite as:

$$Q \leq \left\lceil 1 + \left(\frac{1}{2}\right)\left(N + \frac{\theta - 1}{1 - w}\right)\right\rceil \equiv \overline{Q}_{\mathrm{a}}. \tag{NIC}$$

Thus, any quota that does not exceed $\overline{Q}_{\mathrm{a}}$ will be such that the remaining opponents can always successfully block imposition attempts by the supporters who are not numerous enough to get the agent to implement the action.[27] This means that together (NBC) and (NIC) guarantee that both opponents and supporters will abide by the outcome of the vote and will consume privately whatever resources they have after the agent moves. The following proposition shows that this is also sufficient to guarantee that they vote sincerely without coercion, so an equilibrium with delegation exists.

PROPOSITION 5. *For any $Q \leq \overline{Q}_{\mathrm{a}}$, there exists an agent-implementing subgame-perfect equilibrium. Players contribute $x_0(Q)$ each and vote sincerely. The agent invests toward the action if there are at least Q votes in favor, and reimburses the players otherwise. Players always consume privately the resources they have after the agent's move. If supporters attempt to impose the action, the grand coalition of opponents blocks them.* □

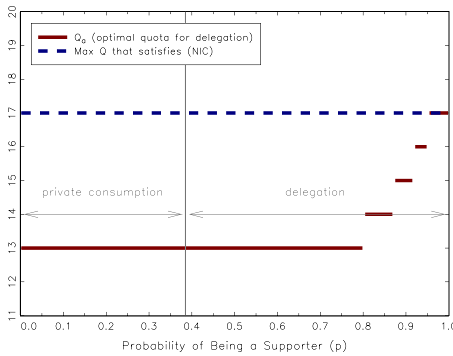
Although this result tells us that there exists an SPE with delegation, it says nothing about the optimal quota players would use, and indeed nothing whatsoever about whether they would even choose to delegate. The next step, then, is to show that there exists a unique optimal quota, which maximizes the payoff from delegating to the agent.

LEMMA 5. *There exists a unique $Q_{\mathrm{a}}(w, p)$, which maximizes the delegation payoff. Moreover, this optimal quota is non-decreasing in p.* □

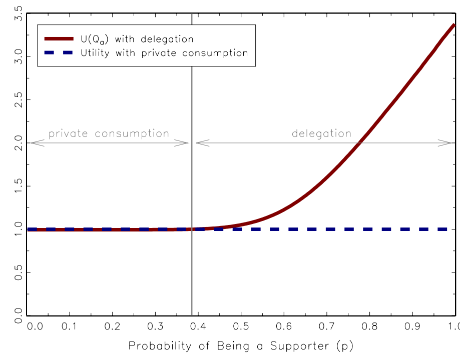
Although we could not obtain a closed-form solution for this optimal delegation quota, finding it numerically is straightforward.[28] To see whether players would choose to delegate, we need to consider the alternative that they do not. We know what happens in that case: the action will never take place (Lemma 1) because sincere voting cannot be supported (Lemma 2). The best alternative to no delegation in the single-shot interaction is private consumption with a payoff of 1. This implies that players would choose to delegate if, and only if, doing so gives them something better.

[27] How demanding is (NIC)? Note that $\overline{Q}_{\mathrm{a}}$ is decreasing in w and that $\lim_{w \to 0} \overline{Q}_{\mathrm{a}} = \lceil (N + \theta + 1)/2 \rceil \geq S_c$. Therefore, any $Q \leq S_c$ is *sufficient* to ensure that (NIC) is satisfied regardless of the wage.

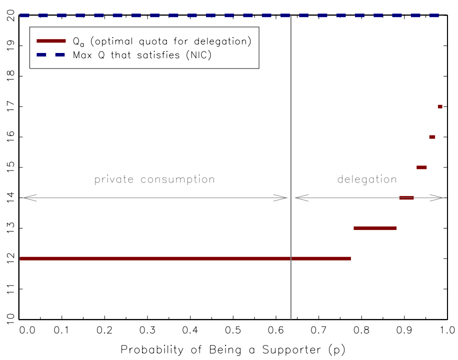
[28] The optimal delegation quota is the smallest integer such that the left-hand side of (13) is less than the right-hand side (the payoff from the next quota higher up is strictly smaller).



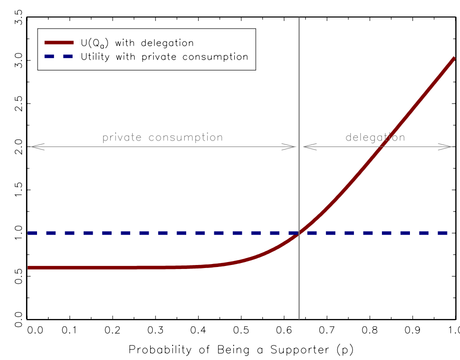
(a) The Voting Rule, $w = 0.005$.



(b) Equilibrium Payoffs, $w = 0.005$



(c) The Voting Rule, $w = 0.4$



(d) Equilibrium Payoffs, $w = 0.4$

Figure 3: Delegation and Private Consumption ($N = 20$, $a = 3$, $\theta = 11$).

Figure 3 shows when delegation is preferable to private consumption for two organizational cost scenarios: relatively modest costs (each actor pays 0.5% of his resource endowment in agent fees, shown in the top row) and somewhat exorbitant ones (each actors pays 40% of his resource endowment in agent fees, the bottom row). All the other parameters are held at the values we used in the previous figures for the coalition of the willing and the universal institution. The vertical lines separate the values of p for which private consumption is preferable from those for which delegation is.

As we know from Lemma 5, the optimal quota is a non-decreasing discontinuous step-function of p. Figures 3(a) and 3(c) show a pattern reminiscent of the optimal quota for the coalition of the willing in Figure 3, which might be surprising. Recall that in coalitions of the willing the quota is forced upward by the sincerity constraint. Delegation is more like the universal institution in that respect: since all players contribute, there is no incentive to vote insincerely when one is a supporter. Something else must be pushing the optimal quota up here.

To understand what this is, note that the upward pressure on the quota under delegation comes from the unconstrained optimization itself: the quota increases because doing so produces better expected payoffs, not because it must or else an equilibrium condition would fail. Setting aside the *ex ante* probability that the quota is met for a moment, it is clear that players prefer *larger* quotas: a large quota means that when it is met, the lingering opposition group will be small, which in turn means fewer resources must be wasted on deterring its potential attempt to undermine the collective decision to implement the action. In other words, fewer resources have to be committed to the agent in excess of what is necessary to implement the action at cost. Formally,

$$\frac{\mathrm{d}x_0(Q)}{\mathrm{d}Q} = \frac{\theta - (1-w)N}{(2N-Q)^2} < 0,$$

where the inequality follows from $w < \overline{w}$. Since the amount contributed to the agent in itself does not affect the probability that the action takes place in equilibrium, it follows that players prefer to conserve as much as possible for private consumption. Thus, for any given probability of the action taking place, players would prefer the largest possible quota in order to minimize excess spending on deterrence. The ceiling on how high this quota can be, of course, comes from the fact that for any given probability of being a supporter, larger quotas means a *lower* probability that the action will take place. This decreases the expected benefits, especially when players expect to be supporters with high probability. The trade off players face, then, is that the lower cost of implementation must come at the expense of its lower probability. As p increases, the probability that any given quota will be met increases as well, which makes the trade off less and less salient. At some point, it becomes beneficial to increase the quota and get the lower implementation cost because the probability that it will be met is high enough. Continuing in this way, we can see that the optimal quota will increase in step-wise fashion as p increases. Even though the behavior of the quota under delegation is superficially the same as its behavior in a coalition of the willing, the causes are radically different. This is clearly seen in Figures 3(b) and 3(d), which show that the expected equilibrium payoff is strictly increasing in p under delegation, whereas it is non-monotonic in the coalition of the willing, as revealed by Figure 2(a).

Another noteworthy result is that there are circumstances under which delegation appears to be preferable even when the agent demands an excessive fee — in this case, up to 45% of each player's budget! The wastage reflected by this fee does reduce the expected payoff from delegation, as one can see by comparing Figure 3(d) with Figure 3(b). This in turn means that delegation will not become attractive until p is sufficiently high. The plots do, however, raise a question: is it the case that delegation is always preferable if p is high enough no matter how large the organizational costs (provided, of course, they retain the feasibility of implementation)? As it turns out, the answer is affirmative. We now establish a *sufficient* condition that ensures that players will prefer to delegate in an agent-implementing equilibrium whenever the action is feasible given the agent fee.

LEMMA 6. *If the probability of being a supporter is sufficiently high, then players strictly prefer to delegate for* any *feasible agent fee.* □

As Figure 3 shows, delegation becomes optimal (relative to private consumption) at much lower values of p than the sufficient condition in Lemma 6 might suggest. This is especially pronounced when the agent's wage is not too high. At any rate, we have shown that there exist conditions under which players would delegate even though it is costly. They prefer to create a formal organization that would enable them to cooperate even in a single-shot interaction even though voting in such an organization is non-binding and even though they must pay organizational costs and dissipate additional resources (to ensure that whatever opposition to implementation remains, it cannot block the action).

5.2 Why Organize with Delegation with Repeated Play?

Consider now the repeated game with the possibility of delegation. Since the choice of Q is made before players learn their preferences for the period, the game is stationary with respect to the voting rule. Thus, we lose no generality if we focus on equilibria in which players set some Q for the duration of the game, and each period begins with their contributions to the agent. We now show that delegation is easily supported in SPE in the repeated game whenever it can be supported in equilibrium of the single-shot game.

PROPOSITION 6. *If delegation is preferable in the single-shot interaction at Q_{a}, then the following strategies constitute a SPE of the repeated game regardless of the discount factor: players choose delegation with Q_{a}, and use the stage-game equilibrium strategies from Proposition 5 in every period of the game.* □

This is an important result because it suggests one way in which players can overcome the limits of self-enforcement inherent in organizing as coalitions of the willing or in an universal institution. These institutional arrangements are very attractive because they can implement the action at cost and avoid the wastage inherent in the delegated environment where the resources must be sufficient to overcome any lingering opposition.[29] However, these coercive environments are fundamentally constrained by the shadow of the future: it has to be long enough so that the long-term costs of failing to cooperate today outweigh any gains that players might obtain by doing so. This requirement could be quite severe, as Figure 2(b) shows. For $p \approx 0.4$, for instance, the minimum discount factor required to enable cooperation is close to 1. It stays above 0.8 for p up to 0.75. In other words, even when there is a 75% chance of each player being supportive of the action, the coercive institutions require them to discount the future by no more than 20% or else cooperation will be impossible.

The great advantage of the non-coercive environment is that cooperation requires no threats of future punishment, which makes the shadow of the future irrelevant. For instance, as Figure 3(b) shows, cooperation with delegation is preferable to private consumption for any $p > 0.4$, which means that it can be implemented in situations where the tough demands of the discount factor would make the other arrangements impossible. Even with the relatively exorbitant agency fees in Figure 3(d), delegation might work where nothing else would (e.g., at p around 0.65). The fact that delegation is non-coercive has other positive implications: there is no need to monitor compliance, and the problems of noise, involuntary defection, and renegotiation do not arise. Deviations are simply ignored: the play continues as if nothing has happened and there is no need to destroy cooperation if someone defects (or is believed to have defected). Finally, note that there is no special

[29] Even this must be tempered somewhat: including transaction costs, which are very similar to the sunk cost of the agent's fee, will reduce the attractiveness of the non-delegated arrangements. Similarly, one could delegate such that implementation is at cost and the lingering opponents are deterred with threats of future retaliation. This, in effect, is equivalent to incorporating transaction costs in the model of the universal institution.

expertise required of the agent when it comes to the action. Delegation here does not occur because the agent can implement the action at lower cost or because he knows something others do not. This, in fact, strengthens our result: any of the traditional reasons to delegate would increase the value of this organizational form relative to the two others, making it even more likely that players would choose it.

6 Conclusion

This paper provides an integrated theoretical framework for analyzing different organizational forms of international cooperation. Our model is premised on two fundamental aspects of collective action. First, and as usual, such action might be difficult to achieve because of incentives to free-ride on the efforts of others. Second, and innovatively, whereas international cooperation might be beneficial for some, it might involve negative externalities for others. This can give rise to highly conflictual situations where significant resources can be wasted on imposing one group's preferred outcome on the other. Whereas most existing work focuses on ways of overcoming the contributor dilemma, we focus on the problem of avoiding dissipation in attempts to implement some collective action.[30]

This perspective of collective action is not meant as a contradiction to extant approaches but as a refinement, an extension that can provide new insights for our understanding of international cooperation. First, we offer an analysis of the rationale for diverse organizational forms for cooperation in a unified theoretical framework. The most important advantage of coalitions of the willing or universal organizations is their ability to avoid conflict and implement the action without any dissipation at all. Delegation, on the other hand, involves sunk cost in form of the agent fee and considerable waste in periods when there are many more supporters than the quota. The great advantage of agent-implementing organizations, however, is that they do not require a long shadow of the future and do not depend on coercive threats to function.

Second, we uncover a novel rationale for delegation. The traditional explanation of why states delegate relies almost exclusively on the assumption that the agent has better information or superior expertise in implementing the collective action (all enforcement problems in the PD scenario can be resolved through repeated interaction). Our model does not require any such asymmetry between the agent and the other actors. Instead, delegation eliminates the need for coercive enforcement mechanisms and works even when the shadow of the future is not long enough to render coercive solutions effective.

Third, our model helps explain why members of international organizations tend to vote on what actions the organization should take, and how the voting rule interacts with the structure of the organization itself. We show that the need to ensure that players truthfully reveal their preferences through voting can be a major driving force behind alternative organizational solutions to collective action problems. Voting only serves to make preferences common knowledge, *not* to impose outcomes on unwilling minorities. However, since voting outcomes are not binding in the anarchic international environment, actors might have

[30] It is important to note, that we do *not* argue that the "Prisoner's Dilemma is not it" or that the "Principal-Agent relationship is not it". Our view of international organizations just does not emphasize these issues in order to get leverage on the question of why states want to create different forms of organizations and why voting in international organizations actually matters.

strong incentives to falsify their own vote or not abide by the outcome of the collective vote. Whereas we identify two coercive mechanisms that can overcome these problems (and in this we are consistent with the existing literature's emphasis on endogenous enforcement), we also identify a solution that requires no threats. We also show that the different solutions have varying pros and cons, and that there exist circumstances that make each of them preferable to the other two. This helps rationalize the empirical existence of all three types.

Fourth, these findings can shed more light on the question of why states tend to comply with international agreements.[31] Whereas most of the literature either treats compliance as epiphenomenal (members comply because they want to and would have done so even without the organization) or attributes it to the enforcement capabilities of the organization (members comply because they are punished if they do not), we show that it is possible to design an institution in which neither is the case. In the agent-implementing organization, members contribute even though they would have not done so without it, but are not punished by others for deviating from prescribed behavior (beyond the failure to take action, that is).

The conceptual shift toward analyzing international cooperation as a phenomenon that involves, at least in part, between-group conflict is merely the first step toward a theory of international organizations. We have abstracted away from many important aspects of international interaction that are highly relevant for the outcomes we observe. For instance, it would be important to relax the assumption of symmetry in resource endowments and valuation of the action for the players. Doing so would introduce more interesting voting rules (e.g., some types of weighted voting) in the design of international organizations. Bringing in security concerns might well introduce the need to admit veto power for some members of the collective. The model is readily adaptable to these types of extensions. In the delegated solution we have also ignored the possibility that the agent might have preferences regarding the action, and in doing so we have not considered the usual agency problems directly although bureaucratic politics and agency slippage are partially reflected in the agent "fee" that members must bear for having access to the agent; the more pronounced these problems, the higher the sunk costs they would have to pay, and the less attractive the delegated solution will be.[32] The model can be extended to consider how agents with preferences about the action itself can be disciplined for failing to comply with the outcome of the collective vote (e.g., by replacing it or by cutting its wage) although doing so would necessarily move us back into the dynamic setting, bring the shadow of the future back into play.

[31]Downs, Rocke, and Barsoom (1996) argue that states enter international organizations if their interests are in line with the policies of the international organization.

[32]See, *inter alia*, Barnett and Finnemore (1999); Nielson and Tierney (2003); Hawkins et al. (2006) on agency costs.

References

Abbott, Kenneth W., and Duncan Snidal. 1998. "Why States Act through Formal International Organizations." *Journal of Conflict Resolution* 42(1): 3–32.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, Robert, and Robert O. Keohane. 1985. "Achieving Cooperation under Anarchy: Strategies and Institutions." *World Politics* 38(oct): 226–254.

Barnett, Michael N. 2003. *Eyewitness to a Genocide: The United Nations and Rwanda*. Ithaca: Cornell University Press.

Barnett, Michael N., and Martha Finnemore. 1999. "The Politics, Power, and Pathologies of International Organizations." *International Organization* 53(4): 699–732.

Clapp, Jennifer. 2004. "WTO Agricultural Trade Battles and Food Aid." *Third World Quarterly* 25(8): 1439–1452.

Downs, George, and David Rocke. 1990. *Tacit Bargaining, Arms Races, and Arms Control*. Ann Arbor: University of Michigan Press.

Downs, George W., and David M. Rocke. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton: Princeton University Press.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50(3): 379–406.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1998. "Managing the Evolution of Multilateralism." *International Organization* 52(2): 397–419.

Epstein, David, and Sharyn O'Halloran. 1999. *Delegating Powers. A Transaction Cost Politics Approach to Policy Making Under Separate Powers*. Cambridge: Cambridge University Press.

Evangelista, Matthew. 1990. "Cooperation Theory and Disarmament Negotiations in the 1950s." *World Politics* 42: 502–528.

Fearon, James D. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52(2): 269–305.

Garrett, Geoffrey. 1992. "International Cooperation and Institutional Choice: the European Communities Internal Market." *International Organization* 46(2): 533–560.

Gilligan, Michael J. 2004. "Is There a Broader-Deeper Trade-Off in International Multilateral Agreements." *International Organization* 58: 459–484.

Gruber, Lloyd. 2000. *Ruling the World: Power Politics and the Rise of Supranational Institutions*. Princeton: Princeton University Press.

Hawkins, Darren L., David A. Lake, Daniel L. Nielson, and Michael J. Tierney. 2006. *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press.

Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(4): 761–799.

Kraft, Christian. 2006. "Third-Market Competition and the Politics of Strategic Delay in WTO Enlargement." Unpublished Working Paper.

Krasner, Stephen D. 1991. "Global Communications and National Power: Life on the Pareto Frontier." *World Politics* 43: 336–366.

Larson, Deborah. 1987. "Crisis Prevention and the Austrian State Treaty." *International Organization* 41: 27–60.

Martin, Lisa L. 1992. "Interests, Power, and Multilateralism." *International Organization* 46: 765–792.

Martin, Lisa L., and Beth A. Simmons. 1998. "Theories and Empirical Studies of International Institutions." *International Organization* 52(4): 729–757.

Morrow, James D. 1994. "Modeling the Forms of International Cooperation: Distribution Versus Information." *International Organization* 48(3): 387–423.

Neumayer, Eric. 2010. "Strategic Delaying and Concessions Extraction in Accession Negotiations to the World Trade Organization." Unpublished Working Paper.

Nielson, Daniel L., and Michael J. Tierney. 2003. "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform." *International Organization* 57(2): 241–276.

Oye, Kenneth. 1985. "Explaining Cooperation under Anarchy." *World Politics* 38(1): 1–24.

Peterson, MJ. 2009. "The EU-US Dispute Over Regulation of Genetically Modified Organisms, Plants, Feeds, and Foods – Case Summary." International Dimensions of Ethics Education in Science and Engineering Case Study Series.

Pollack, Mark. 1997. "Delegation, Agency, and Agenda Setting in the European Community." *International Organization* 51(1): 99–134.

Pollack, Mark A., and Gregory C. Shaffer. 2009. *When Cooperation Fails: The International Law and Politics of Genetically Modified Foods*. New York: Oxford University Press.

Rhodes, Carolyn. 1989. "Reciprocity in Trade: The Utility of a Bargaining Strategy." *International Organization* 43: 273–300.

Rosendorff, B. Peter V., and Helen Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55(4): 829–857.

Schneider, Christina J., and Johannes Urpelainen. 2010. "Distributional Conflict between Powerful States and International Treaty Ratificiation." Unpublished Working Paper.

Sebenius, James K. 1992. "Challenging Conventional Explanations of International Cooperation: Negotiation Analysis and hte Case of Epistemic Communities." *International Organization* 46: 323–366.

Snidal, Duncan. 1985. "Coordination versus Prisoners' Dilemma: Implications for International Cooperation." *American Political Science Review* 79: 923–42.

Stein, Arthur. 1982. "Coordination and Collaboration: Regimes in an Anarchic World." *International Organization* 36(2): 299–324.

Svolik, Milan. 2006. "Lies, Defection, and the Pattern of International Cooperation." *American Journal of Political Science* 50(4): 909–925.

Voeten, Erik. 2005. "The Political Origins of theUN Security Council's Ability to Legitimize the Use of Force." *International Organization* 59: 527–557.

A Mathematical Appendix

Proof (Lemma 1). If supporters are spending $x = 0$ in equilibrium, then the unique such equilibrium is where $y = 0$ as well. We now show that there is no (symmetric) equilibrium in which supporters spend a positive amount. The proof proceeds in several steps. We first show that if opponents are spending a positive amount, then there exists some number of supporters that are exactly sufficient to implement the action without any slack from their investment. We then show that if supporters are spending, then opponents must be spending as well. Next, we show that the opponents cannot be spending less than their entire resource, and complete the proof by showing that spending the entire resource is not an equilibrium.

STEP 1: in any equilibrium with $x > 0$ and $y \geq 0$, there exists $S_x \in \mathbb{Z}$ such that $xS_x - y(N - S_x) = \theta$. This asserts that the equilibrium positive contributions for supporters must be such that there exists a number of supporters, S_x, that are exactly necessary to implement the action given the spending of the other players. Clearly, any $S \geq S_x$ will impose the action. Note that $S_x \geq 2$ because no single player can implement the action. Solving for S_x yields

$$S_x = \frac{\theta + yN}{x + y}, \tag{3}$$

and the claim is that $S_x \in \mathbb{Z}$; that is, it is a valid group size. To prove this, suppose to the contrary that it is not, which means that the minimum number of supporters required to implement the action with this spending is $\hat{S}_x = \lceil S_x \rceil > S_x > \lfloor S_x \rfloor = \hat{S}_x - 1$, where the inequalities follows from the fact that S_x is not an integer. Consider now the strategy for a player who learns that he is a supporter. If he spends x, as required by the equilibrium strategy, his expected payoff would be:

$$U_s(x) = \sum_{k=0}^{\hat{S}_x - 2} (1 - x) f(k) + \sum_{k=\hat{S}_x - 1}^{N-1} (1 + a - x) f(k) = 1 - x + a \sum_{k=\hat{S}_x - 1}^{N-1} f(k).$$

Note now that since S_x is *decreasing* in x, there exists $x' < x$ such that $S_x < S_{x'} < \lceil S_x \rceil = \hat{S}_x$. That is, player i can reduce his spending without affecting the size of the coalition necessary to implement the action. But then

$$U_s(x') = 1 - x' + a \sum_{k=\hat{S}_x - 1}^{N-1} f(k) > U_s(x),$$

so deviating to that smaller contribution would be profitable, contradicting the equilibrium supposition. Therefore, in equilibrium where supporters spend a positive amount $x > 0$, it must be that $S_x \in \mathbb{Z}$. That is, they spend just enough for some subset of the players to be able to implement the action without any remaining slack in the amount they spend.

STEP 2: in any equilibrium, $x > 0 \Rightarrow y > 0$; that is, if supporters spend a positive amount in equilibrium, then so must opponents. To see this, suppose to the contrary that there is an equilibrium in which $x > 0$ but $y = 0$. From (3) and the first step, we obtain

$$S_x = \frac{\theta}{x} \in \mathbb{Z}.$$

Consider now the strategy for a player who realizes he is an opponent. If he sticks with his supposed equilibrium strategy and spends nothing, $y = 0$, against the action, his expected payoff would be:

$$U_o(y) = \sum_{k=0}^{S_x-1} (1) f(k) + \sum_{k=S_x}^{N-1} (1-a) f(k) = 1 - a \sum_{k=S_x}^{N-1} f(k).$$

If he deviates and spends some $y' > 0$ against the action, S_x will no longer be able to implement it because $xS_x - y' = \theta - y' < \theta$. This implies that the minimum number of supporters that can impose action is now $S \geq \lceil (\theta + y')/x \rceil \geq S_x + 1$, where the second inequality follows from $S_x \in \mathbb{Z}$. Thus, such a deviation would, at the very least, undermine the action when there are S_x supporters among the remaining players, in which case the deviating opponent would get $1 - y' - a/2$. The deviation would yield:

$$\begin{aligned} U_o(y') &= \sum_{k=0}^{S_x-1} (1-y') f(k) + \left(1 - y' - \frac{a}{2}\right) f(S_x) + \sum_{k=S_x+1}^{N-1} (1-a-y') f(k) \\ &= 1 - y' - \left(\frac{a}{2}\right) f(S_x) - a \sum_{k=S_x+1}^{N-1} f(k). \end{aligned}$$

We now claim that this payoff is strictly better for y' small enough:

$$\lim_{y'\to 0} \left[U_o(y') - U_o(y)\right] = \lim_{y'\to 0} \left[\left(\frac{a}{2}\right) f(S_x) - y'\right] = \left(\frac{a}{2}\right) f(S_x) > 0,$$

where the inequality follows from $f(S_x) > 0$. Thus, there always exists a profitable deviation, contradicting the equilibrium supposition. We conclude that if supporters spend a positive amount in equilibrium, then it must be that opponents spend a positive amount as well.

STEP 3: in any equilibrium, $x > 0 \Rightarrow y = 1$. This strengthens the result in the previous step by showing that in any equilibrium in which supporters spend a positive amount, opponents must be spending their entire resource against the action. To see that, suppose to the contrary that $x > 0$ and $y \in (0, 1)$. By (3) and the first step, we know that $S_x \in \mathbb{Z}$. But then the logic of the second step carries through: any opponent can spend $y' \in (y, 1)$ and ensure that S_x would no longer be sufficient for imposing the action, which would now require $S \geq S_x + 1$. This means that the difference between this deviation and spending y is $y - y' + af(S_x)/2$, but since $\lim_{y'\to y} [y - y' + af(S_x)/2] = af(S_x)/2 > 0$, this deviation would be profitable. This contradicts the equilibrium supposition, and establishes the claim.

STEP 4: there is no equilibrium with $x > 0$ and $y = 1$. We have established that if supporters spend positive amount in equilibrium, then opponents must be spending everything. We now show that if they are supposed to spend everything, then there exists a profitable deviation. Suppose, to the contrary, that there is an equilibrium with $x > 0$ and $y = 1$. From the first step and (3) we know that

$$S_x = \frac{\theta + N}{x + 1} \in \mathbb{Z}.$$

Consider now the behavior of a player who realizes he is an opponent. If he spends $y = 1$ as he is supposed to, his expected payoff will be:

$$U_o(y) = \sum_{k=0}^{S_x-1} (1-y) f(k) + \sum_{k=S_x}^{N-1} (1-y-a) f(k) = 1 - y - a \sum_{k=S_x}^{N-1} f(k).$$

Suppose now that the player deviates and invests $y' \in (0,1)$ instead of y against the action. The action will be implemented if $xS - y(N-S-1) - y' \geq \theta$, or if

$$S \geq S_x - \frac{1-y'}{x+1} > S_x - 1,$$

where the second inequality follows from $(1-y')/(x+1) < 1$. This means that the smallest group of supporters necessary to impose the action will remain at S_x. But then the payoff from this deviation would be

$$U_o(y') \sum_{k=0}^{S_x-1} (1-y') f(k) + \sum_{k=S_x}^{N-1} (1-y'-a) f(k) = 1 - y' - a \sum_{k=S_x}^{N-1} f(k) > U_o(y),$$

making the deviation profitable and contradicting the equilibrium supposition. Thus, it cannot be the case that opponents are spending all their resources against the action.

Collectively, these steps establish that everyone consuming privately is the unique (symmetric) equilibrium of the game without communication. ■

Proof (Lemma 2). Suppose there is an equilibrium in which all supporters vote to implement the action, and all opponents vote against it. In such an equilibrium, the game after the vote would be identical to the game of complete information, so Proposition 1 tells us what will happen there: the action will get implemented if the (self-identified) group of supporters is large enough to overcome the largest possible (self-identified) opposing coalition.

Consider the behavior of a player who learns that he opposes the action. If he follows his supposed equilibrium strategy and votes against the action, then the action will be imposed if there are at least S_c supporters among the remaining players. By supposition, $S_c \leq N-1$, so his expected payoff will be:

$$U_o = \sum_{k=0}^{S_c-1} (1) f(k) + \sum_{k=S_c}^{N-1} (1-a) f(k) = 1 - a \sum_{k=S_c}^{N-1} f(k). \tag{4}$$

If he deviates and votes *for* the action but then invests *against* it, then the action will not be attempted if there are $S_c - 2$ supporters among the remaining players, and will be attempted otherwise. However, for any $S \geq S_c - 1$, supporters will incorrectly believe that there are S of them, and will spend x_c each, when in fact there will be $S-1$ because the opponent will not invest for the action. Since every other player is supposed to have voted sincerely, the opponent would know precisely how many supporters there are after the vote: $S-1$. He can tailor his opposition spending to this knowledge and ensure that the grand coalition of remaining opponents would be able to block the action. To see this, note that $S-1$ players would spend x_c each for the action, the opponent will spend $y>0$ against it, and the remaining $N-(S-1+1)=N-S$ players can spend up to 1 against it (because they move after observing the move by the self-identified supporters). We thus have:

$$(S-1)x_c - y - (N-S) \leq \theta - 1 \Leftrightarrow y \geq 2 - \frac{N+\theta}{S} \equiv \hat{y}(S).$$

Since $S \leq N$, it follows that $(N+\theta)/S > 1$, which means that $\hat{y}(S) < 1$ for all S, so such an investment is feasible. Furthermore, observe that $\hat{y}(S_c) = 0$ and $\hat{y}(S) < 0$ for any $S < S_c$, which means that any $y = 0$ would be sufficient to ensure that the action would be blocked in these cases. This implies that the opponent can ensure that the action is blocked for any $S \leq S_c$ at no cost to himself (just by voting insincerely), and that is is blocked for any $S > S_c$ at minimal cost to himself by spending $\hat{y}(S)$. The expected payoff from this deviation is:

$$\hat{U}_o = \sum_{k=0}^{S_c}(1)f(k) + \sum_{k=S_c+1}^{N-1}(1-\hat{y}(k))f(k) = 1 - \sum_{k=S_c+1}^{N-1}\hat{y}(k)f(k). \tag{5}$$

We now obtain

$$\hat{U}_o - U_o = af(S_c) + \sum_{k=S_c+1}^{N-1}\left[a - \hat{y}(k)\right]f(k) > 0,$$

where the inequality follows from $f(S_c) > 0$ because $S_c \leq N-1$, and $a \geq 2 > 1 \geq \hat{y}(S) > 0$ for any S. The deviation is profitable, which means that sincere voting cannot be sustained in equilibrium. ∎

Proof (Proposition 2). Fix Q and consider the *ex ante* per-period equilibrium payoff for some player i:

$$\begin{aligned} u_i(\sigma) = & \underbrace{\sum_{k=0}^{Q-2}(1)f(k)}_{\text{no action regardless of } i\text{'s vote}} + \underbrace{\left[p(1+a-x(Q)) + (1-p)(1)\right]f(Q-1)}_{\text{action occurs only if } i \text{ votes in favor}} \\ & + \underbrace{\sum_{k=Q}^{N-1}\left[p(1+a-x(k+1)) + (1-p)(1-a)\right]f(k)}_{\text{action occurs regardless of } i\text{'s vote}}, \end{aligned}$$

which simplifies to $u_i(\sigma) = 1 + \zeta_{\mathrm{w}}(Q)$, where $\zeta_{\mathrm{w}}(Q) < a$ is defined in the proposition. The equilibrium payoff from this strategy is $u_i(\sigma)/(1-\delta)$.

Consider first the implementation stage. Suppose first that $q \geq Q$ so the action should take place. Any supporter who deviates from $x(q)$ will cause the action to fail, making this unprofitable. Furthermore, there is no need to contribute more than the minimum necessary to implement it. Since this is an at-cost implementation, any opponent who invests against the action some y arbitrarily close to zero can derail it but then the game will revert to the unconditional SPE. Doing so would not be profitable if

$$1 - y + \frac{\delta(1)}{1-\delta} \leq 1 - a + \frac{\delta u_i(\sigma)}{1-\delta}$$

for $y \to 0$. We can rewrite this as $(1-\delta)a \leq \delta\zeta_{\text{w}}(Q)$. The necessary condition for this inequality to work is $\zeta_{\text{w}}(Q) > 0$. This condition is also sufficient to ensure that there exists δ high enough to satisfy the inequality. In that case any $\delta \geq \underline{\delta}_{\text{w}}(Q)$, where the latter is defined in (1), will work. Note in particular that $a + \zeta_{\text{w}}(Q) > 0$, and that $\zeta_{\text{w}}(Q) > 0$ ensures that $\underline{\delta}_{\text{w}}(Q) < 1$, so solutions exist.

Suppose now that $q < Q$ so the action is not supposed to take place. If $q < S_c$ then the action cannot be imposed and if the supporters attempt to do so, the opponents will block it. Since the game will continue as before, the only payoff from such an attempt for supporters would be the wasted resources in the attempt, which makes it unprofitable. As usual, opponents do not need special incentive to block an action when they can do so. If, on the other hand, $q \geq S_c$, then the (self-declared) supporters can impose the action if they wish to but doing so would result in the reversion to the unconditional SPE. This deviation will not be profitable if:

$$1 + a - x(q) + \frac{\delta(1)}{1-\delta} \leq 1 + \frac{\delta u_i(\sigma)}{1-\delta},$$

which we can rewrite as $(1-\delta)(a - x(q)) \leq \delta\zeta_{\text{w}}(q)$. Recall now that the condition that prevents the deviation of an opponent is $(1-\delta)a \leq \delta\zeta_{\text{w}}(q)$. Thus, if the opponent will not deviate, then a supporter certainly would not do so in the implementation phase.

We now turn to the voting stage. Consider now a player who learns that he opposes the action. If he votes sincerely, then his expected payoff in this period will be:

$$u_o(\sigma) = \sum_{k=0}^{Q-2}(1)f(k) + (1)f(Q-1) + \sum_{k=Q}^{N-1}(1-a)f(k) = 1 - a(1 - F(Q-1)).$$

If he votes, falsely, in support of the action and then behaves as a supporter (so the action gets implemented), his payoff in this current period will be

$$\sum_{k=0}^{Q-2}(1)f(k) + (1 - a - x(Q))f(Q-1) + \sum_{k=Q}^{N-1}(1 - a - x(k+1))f(k) < u_o(\sigma).$$

Since this deviation will not be detected (and would not have been punished if it had), the game will continue as before. Thus, this deviation cannot be profitable. If he votes for the action but then derails it with some y arbitrarily close to zero, his payoff for the current period will be:

$$\sum_{k=0}^{Q-2}(1)f(k) + (1-y)f(Q-1) + \sum_{k=Q}^{N-1}(1-y)f(k) = 1 - y(1 - F(Q-1)).$$

However, this deviation is observable and will be punished. This deviation will not be profitable for any $y \to 0$ if $1 + \delta/(1-\delta) \leq 1 - a(1 - F(Q-1)) + \delta u_i(\sigma)/(1-\delta)$. This reduces to $(1-\delta)a(1 - F(Q-1)) \leq \delta\zeta_{\mathrm{w}}(Q)$. However, since $(1 - F(Q-1))a < a$, this condition will be satisfied whenever the condition that prevents an opponent (who has voted sincerely) from derailing the implementation. (This makes sense: an insincere vote will increase the probability of having to derail the action, and thus the probability of the sanction relative to a sincere vote against it followed by derailing.)

Finally, consider a player who learns that he supports the action. If he votes sincerely, then his expected payoff will be:

$$u_s(\sigma) = \sum_{k=0}^{Q-2} (1) f(k) + (1 + a - x(Q)) f(Q-1) + \sum_{k=Q}^{N-1} (1 + a - x(k+1)) f(k).$$

If he deviates and votes insincerely and then does not derail the action (he has no incentive to vote insincerely and derail it), his payoff would be

$$u_s(\sigma') = \sum_{k=0}^{Q-2} (1) f(k) + (1) f(Q-1) + \sum_{k=Q}^{N-1} (1 + a) f(k).$$

Since this deviation will go undetected, the game continues as before. Thus, the necessary and sufficient condition for this deviation to be unprofitable is $u_s(\sigma) - u_s(\sigma') \geq 0$, or

$$(a - x(Q)) f(Q-1) \geq \sum_{k=Q}^{N-1} x(k+1) f(k),$$

which we can rewrite as (SC). This exhausts the possible deviations and completes the proof. ∎

Proof (Lemma 3). We begin by showing that unconstrained maximization selects the complete information social optimum; that is $Q_{\mathrm{u}} = \mathfrak{S}$. The payoff function will be increasing at Q if, and only if, $U(Q+1) - U(Q) = \zeta_{\mathrm{u}}(Q+1) - \zeta_{\mathrm{u}}(Q) > 0$, and decreasing if the difference is negative. We now obtain:

$$\begin{aligned}
&\zeta_{\mathrm{u}}(Q+1) - \zeta_{\mathrm{u}}(Q) \\
&\quad = p\left(a - \frac{\theta}{Q+1}\right) f(Q) - p\left(a - \frac{\theta}{Q}\right) f(Q-1) - \left[(2p-1)a - \frac{p\theta}{Q+1}\right] f(Q) \\
&\quad = (1-p) a f(Q) - p\left(a - \frac{\theta}{Q}\right) f(Q-1).
\end{aligned}$$

Thus, $\zeta_{\mathrm{u}}(Q+1) - \zeta_{\mathrm{u}}(Q) > 0 \Leftrightarrow (1-p) a f(Q) > p\left(a - \frac{\theta}{Q}\right) f(Q-1)$. The latter inequality is:

$$(1-p)a\binom{N-1}{Q}p^{Q}(1-p)^{N-Q-1} > p\left(a-\frac{\theta}{Q}\right)\binom{N-1}{Q-1}p^{Q-1}(1-p)^{N-Q}$$
$$a\binom{N-1}{Q} > \left(a-\frac{\theta}{Q}\right)\binom{N-1}{Q-1}$$
$$\frac{a}{Q} > \frac{a-\theta/Q}{N-Q},$$

which yields

$$Q < \frac{N+\theta/a}{2} \equiv \widetilde{Q}.$$

Thus, the payoff is strictly increasing for all $Q < \widetilde{Q}$, and strictly decreasing for all $Q > \widetilde{Q}$, which implies that the unconstrained optimum is at $Q_{\mathrm{u}} = \lceil \widetilde{Q} \rceil = \mathfrak{S}$. Clearly, if $\theta \leq \mathfrak{S}$, then the first constraint will not be binding; otherwise, $Q_{\mathrm{u}} = \theta$ as long as the second constraint is not binding. We now turn to investigating the conditions under which it will.

We can rewrite (SC) as

$$\frac{a}{\theta} \geq \sum_{k=0}^{N-Q}\left[\frac{(N-Q)!(Q-1)!}{(Q+k)!(N-Q-k)!}\right]\left(\frac{p}{1-p}\right)^{k} \equiv T(p,Q). \tag{6}$$

Note that $a/\theta > 0$, but since

$$\frac{\partial T}{\partial p} = \sum_{k=0}^{N-Q}\left[\frac{(N-Q)!(Q-1)!}{(Q+k)!(N-Q-k)!}\right]\left[\frac{kp^{k-1}}{(1-p)^{k+1}}\right] > 0,$$

the inequality must be violated for p sufficiently high ($\lim_{p\to 1} T(p,Q) = \infty$ for any $Q < N$). On the other hand, $\lim_{p\to 0} T(p,Q) = 0$, and the inequality is satisfied for any Q.

Take now $Q_{\mathrm{u}} = \max\{\mathfrak{S}, \theta\}$ so that the first constraint is satisfied. For p sufficiently low condition (SC) will be met (with $n(p) = 0$), but as we increase p, it must eventually fail. Since $T(p,Q)$ is continuous in p, there must exist some $\hat{p}$ where (6) is satisfied with equality, so that the condition will fail for any $p > \hat{p}$. We now show that it is necessary to increase Q to restore the condition. First, note that $T(p,Q)$ is strictly decreasing in Q. Since Q changes in discrete jumps, we can rewrite $T(p,Q+1) - T(p,Q) = D(p,Q)$ as:

$$D(p,Q) = \sum_{k=0}^{N-Q}\left[\frac{(N-Q-1)!(Q-1)!}{(Q+k+1)!(N-Q-k)!}\right]\left(\frac{p}{1-p}\right)^{k}[Q-(k+1)N] < 0,$$

where the inequality follows from the fact that the first two terms in the summation are positive but the third is negative for any $k \geq 0$.

We now show that it is possible to satisfy (6) at $p > \hat{p}$ by choosing some $Q > Q_{\mathrm{u}}$. For this, it is sufficient to establish that there exists $\varepsilon > 0$ such that $T(\hat{p}+\varepsilon, Q_{\mathrm{u}}+1) < T(\hat{p}, Q_{\mathrm{u}})$. Since $T(p,Q) = T(p,Q+1) - D(p,Q)$, we can write this as:

$$T(\hat{p}+\varepsilon, Q_{\mathrm{u}}+1) - T(\hat{p}, Q_{\mathrm{u}}) = T(\hat{p}+\varepsilon, Q_{\mathrm{u}}+1) - T(\hat{p}, Q_{\mathrm{u}}+1) + D(\hat{p}, Q_{\mathrm{u}}).$$

But since $\lim_{\varepsilon \to 0} [T(\hat{p} + \varepsilon, Q_u + 1) - T(\hat{p}, Q_u + 1)] = 0$ but $D(\hat{p}, Q_u) < 0$, the fact that this difference is continuous in ε implies that that there exists $\hat{\varepsilon} > 0$ such that $T(\hat{p}+\varepsilon, Q_u + 1) - T(\hat{p}, Q_u + 1) + D(\hat{p}, Q_u) < 0$ for all $\varepsilon < \hat{\varepsilon}$. In other words, (6) must be satisfied at $T(\hat{p}+\varepsilon, Q_u+1)$. Thus, the optimal quota for these values of p will be Q_u+1, or $n(p) = 1$. Continuing in this way, we find that as p increases, $n(p)$ must increase by one unit in a step-wise manner as well until the quota reaches unanimity, in which case the condition will be satisfied regardless of the value of p because then $T(p, N) = 1/N < a/\theta$. ∎

Proof (Proposition 3). Fix Q and consider the voting phase assuming that players will contribute if the quota is met. With everyone contributing when they have to there is no incentive not to vote sincerely. If a supporter votes against the action, it will fail if he happens to be pivotal, and he will contribute if it gets implemented even without his vote. Clearly such a deviation cannot be profitable. If an opponent votes for the action, he will only cause it to be implemented if he happens to be pivotal, an unprofitable deviation. Thus, it is only necessary to ensure that the contribution is properly enforced.

Consider now the phase in which players have voted and there are $q \geq Q$ in support so the action should take place under the equilibrium strategies. Since $x = \theta/N$, any player who fails to contribute will derail the action. The consequences of not contributing x are the same regardless of how one has voted, so we can analyze the deviation in this phase of the stage game without reference to the vote of the player. It is easy to see that if an opponent can be induced to contribute, then a supporter will surely do so: the continuation game is the same for both and the current payoff from the equilibrium strategy is lower for the opponent. Thus, it is sufficient to provide an incentive to the opponent. If he does not contribute, the action will fail to take place, and the game will revert to the non-cooperative equilibrium. Thus, the expected payoff from this deviation is 1 (private consumption) in each period, starting with the current one, so the total payoff will be $1/(1-\delta)$. If the player follows the equilibrium strategy σ and contributes x, the action will take place now and in every future period in which the quota is met. To calculate the latter, we need the *ex ante* expected payoff to an arbitrary player (i.e., the expected payoff before he learns his preferences). Since the action takes place for any $q \geq Q$, the per-period expected payoff is:

$$u_i(\sigma) = \underbrace{\sum_{k=0}^{Q-2} (1) f(k)}_{\text{no action regardless of } i\text{'s vote}} + \underbrace{[p(1+a-x) + (1-p)(1)]\, f(Q-1)}_{\text{action occurs only if } i \text{ votes in favor}}$$
$$+ \underbrace{\sum_{k=Q}^{N-1} [p(1+a) + (1-p)(1-a) - x]\, f(k)}_{\text{action occurs regardless of } i\text{'s vote}},$$

which simplifies to:

$$u_i(\sigma) = 1 + p(a-x) f(Q-1) + \sum_{k=Q}^{N-1} [(2p-1)a - x]\, f(k).$$

Thus, the condition for an opponent to follow the equilibrium strategy and invest for the action today is:

$$1 - a - x + \frac{\delta u_i(\sigma)}{1-\delta} \geq 1 + \frac{\delta(1)}{1-\delta},$$

which we can rewrite as $\delta u_i(\sigma) \geq \delta + (1-\delta)(a+x)$, or $\delta\zeta_{\mathrm{u}}(Q) \geq (1-\delta)(a+x)$. Since $a + x + \zeta_{\mathrm{u}}(Q) > 0$, this yields $\delta \geq \underline{\delta}_{\mathrm{u}}(Q)$, with $\underline{\delta}_{\mathrm{u}}(Q)$ defined in (2).To ensure that $\underline{\delta}_{\mathrm{u}}(Q) < 1$, we require that $\zeta_{\mathrm{u}}(Q) > 0$, as stated.

Finally, we need to consider $q < Q$ when the action will not take place. Clearly, no opponent would contribute anything if the supporters follow the equilibrium strategy, so we only need to make sure that the supporters do so. If they attempt to impose the action but $q < S_c$, then the opponents would block it and the game will continue as before. Regardless of whether opponents do so now, the game continues as if nothing has happened. (There is no need to provide opponents with an incentive to block the action when they can do so.) Thus, this deviation will result in wasted spending and no action, so it cannot be profitable. The only possibly tempting deviation is for them to impose the action, which they can do when $q \geq S_c$ because then opponents would not be able to block it (recall that they move second).[33] In this case, the action can take place now (with opponents consuming privately) but the game will revert to the non-communicative Nash equilibrium of the stage play from the following period. The condition for supporters to follow their equilibrium strategy and not impose the action today is:

$$1 + \frac{\delta u_i(\sigma)}{1-\delta} \geq 1 + a - x(q) + \frac{\delta(1)}{1-\delta},$$

which simplifies to $\delta\zeta_{\mathrm{u}}(Q) \geq (1-\delta)(a - x(q))$. Since this inequality must hold for all realizations of $q < Q \leq N$ and because the RHS is increasing in q (since $x(q) = \theta/q$ is decreasing), it is necessary that it be satisfied at $q = N$. Thus, we end up with $\delta\zeta_{\mathrm{u}}(Q) \geq (1-\delta)(a-x)$. Recalling that the condition that prevents deviation by opponents is $\delta\zeta_{\mathrm{u}}(Q) \geq (1-\delta)(a+x)$, we conclude that whenever the latter is satisfied, the supporters will have no incentive to impose the action either. ■

Proof (Lemma 4). Since $U(Q) = 1 + \zeta_{\mathrm{u}}(Q)$, the payoff function will be increasing at Q if, and only if, $U(Q+1) - U(Q) = \zeta_{\mathrm{u}}(Q+1) - \zeta_{\mathrm{u}}(Q) > 0$, and decreasing if the difference is negative. We now obtain:

$$\begin{aligned}
&\zeta_{\mathrm{u}}(Q+1) - \zeta_{\mathrm{u}}(Q) \\
&\quad = p(a-x)f(Q) - p(a-x)f(Q-1) + [(2p-1)a - x]\,(F(Q) - F(Q-1)) \\
&\quad = p(a-x)f(Q) - p(a-x)f(Q-1) - [(2p-1)a - x]\,f(Q) \\
&\quad = (1-p)(a+x)f(Q) - p(a-x)f(Q-1) \\
&\quad = (1-p)(a+x)\binom{N-1}{Q}p^{Q}(1-p)^{N-Q-1} - p(a-x)\binom{N-1}{Q-1}p^{Q-1}(1-p)^{N-Q} \\
&\quad = p^{Q}(1-p)^{N-Q}\left[(a+x)\binom{N-1}{Q} - (a-x)\binom{N-1}{Q-1}\right]
\end{aligned}$$

[33] For instance, suppose $Q = N$ so they have agreed to require unanimity for the action. Then, if $q = N-1$ vote in favor, they should have no incentive to force it against the one holdout.

$$= p^Q(1-p)^{N-Q}\left[(a+x)\frac{(N-1)!}{Q!(N-Q-1)!} - (a-x)\frac{(N-1)!}{(Q-1)!(N-Q)!}\right]$$
$$= \left[\frac{p^Q(1-p)^{N-Q}(N-1)!}{(Q-1)!(N-Q-1)!}\right]\left[\frac{a+x}{Q} - \frac{a-x}{N-Q}\right].$$

Since the first bracketed term is always positive, it follows that

$$\zeta_{\mathrm{u}}(Q+1) - \zeta_{\mathrm{u}}(Q) > 0 \Leftrightarrow \frac{a+x}{Q} - \frac{a-x}{N-Q} > 0.$$

Solving the second inequality yields $(a+x)N > 2aQ$, which, after substituting $x = \theta/N$ ends in:

$$Q < \frac{N+\theta/a}{2} \equiv \widetilde{Q}. \tag{7}$$

Thus, if $Q < \widetilde{Q}$, then $U(Q+1) > U(Q)$, and the payoff function is increasing; but if $Q > \widetilde{Q}$, then $U(Q+1) < U(Q)$, so it is decreasing. Since for any $Q < \widetilde{Q}$ we would pick $Q+1$ for a higher payoff, it follows that the best possible payoff is at $Q_{\mathrm{u}} = \lceil \widetilde{Q} \rceil = \mathfrak{S}$. ■

Proof (Proposition 4). We need to show that $U_{\mathrm{w}}(Q) = U_{\mathrm{u}}(Q) \Leftrightarrow \zeta_{\mathrm{w}}(Q) = \zeta_{\mathrm{u}}(Q)$. We can rewrite this equation as:

$$p\left(a - \frac{\theta}{Q}\right) f(Q-1) + \sum_{k=Q}^{N-1}\left[(2p-1)a - \frac{p\theta}{k+1}\right] f(k)$$
$$= p\left(a - \frac{\theta}{N}\right) f(Q-1) + \sum_{k=Q}^{N-1}\left[(2p-1)a - \frac{\theta}{N}\right] f(k),$$

which simplifies to:

$$\sum_{k=Q}^{N-1}\left(\frac{1}{N} - \frac{p}{k+1}\right) f(k) = p\left(\frac{1}{Q} - \frac{1}{N}\right) f(Q-1). \tag{8}$$

We need to prove (8) for an arbitrary Q, which we now do by induction. First, we show that it holds for $Q = N$. Since the summation term is zero (the lower bound exceeds the upper bound), it is sufficient to show that the right-hand side is zero too:

$$p\left(\frac{1}{N} - \frac{1}{N}\right) f(N-1) = 0.$$

For the inductive step, assume that (8) holds for some $Q > 1$. We now prove that the claim holds for $Q-1$ as well. Rewriting the claim at $Q-1$ yields:

$$p\left(\frac{1}{Q-1} - \frac{1}{N}\right) f(Q-2) = \sum_{k=Q-1}^{N-1}\left(\frac{1}{N} - \frac{p}{k+1}\right) f(k)$$
$$= \left(\frac{1}{N} - \frac{p}{Q}\right) f(Q-1) + \sum_{k=Q}^{N-1}\left(\frac{1}{N} - \frac{p}{k+1}\right) f(k),$$

and since the claim is assumed to hold at Q, we substitute the second term using (8):

$$\begin{aligned} &= \left(\frac{1}{N} - \frac{p}{Q}\right) f(Q-1) + p\left(\frac{1}{Q} - \frac{1}{N}\right) f(Q-1) \\ &= \left(\frac{1-p}{N}\right) f(Q-1). \end{aligned}$$

Using the definition of the probability mass function, we can rewrite this as:

$$\left(\frac{1}{Q-1} - \frac{1}{N}\right)\binom{N-1}{Q-2} p^{Q-1}(1-p)^{N-Q+1} = \left(\frac{1}{N}\right)\binom{N-1}{Q-1} p^{Q-1}(1-p)^{N-Q+1}$$

which, after canceling the probability terms on both sides, yields

$$\left[\frac{N-Q+1}{N(Q-1)}\right]\left[\frac{(N-1)!}{(Q-2)!(N-Q+1)!}\right] = \left(\frac{1}{N}\right)\left[\frac{(N-1)!}{(Q-1)!(N-Q)!}\right]$$

and since $(N-Q+1)! = (N-Q+1)(N-Q)!$, and $(Q-1)(Q-2)! = (Q-1)!$, cancellations on both sides yield

$$\frac{1}{(Q-1)!(N-Q)!} = \frac{1}{(Q-1)!(N-Q)!},$$

or simply

$$1 = 1,$$

so the claim holds at $Q-1$. By induction, it must hold for all $Q = 1, 2, \ldots, N$. ■

CLAIM 1. *Strategy profiles with sincere voting where $x(q) = 0$ for any $q \geq Q$ are least vulnerable to deviations by opponents.* □

Proof. In equilibrium with sincere voting, they would not spend more than absolutely necessary to ensure the action would succeed: $R + qx(q) - (1 - x_0)(N - q) = \theta$, which gives us

$$x(q) = \max\left\{\frac{\theta + (1 - x_0)(N - q) - R}{q}, 0\right\},$$

where we note that the numerator would be negative for any

$$q \geq \left\lceil N - \frac{R - \theta}{1 - x_0} \right\rceil \equiv Q_0.$$

This means that in any equilibrium, supporters would contribute $x(q) > 0$ for any $q \in [Q, Q_0 - 1]$, and nothing for any $q \geq Q_0$. Note that it might well be the case that $Q = Q_0$ or, if that is not the case, $Q_0 > N$ (in which case supporters would always have to contribute positive amounts to implement the action). We now argue that any $Q_0 > Q$, which can be expressed as $Q_0 \geq Q + 1$, is vulnerable to a deviation by opponents who could vote insincerely for the action and then derail it afterwards.

Consider first the expected payoff for an opponent from a sincere vote:

$$U_o = \sum_{k=0}^{Q-1} (1-w) f(k) + \sum_{k=Q}^{N-1} (1 - x_0 - a) f(k) = 1 - w - (a + x_0 - w) \sum_{k=Q}^{N-1} f(k). \quad (9)$$

If he deviates and votes insincerely, then the agent will attempt implementation when there are $q \geq Q - 1$ supporters among the remaining players (because with the false vote from the opponent there would be Q votes in favor). On the other hand, having observed $q + 1$ votes in favor, supporters would invest $x(q + 1) > 0$ for any $q + 1 \leq Q_0 - 1$, and would invest nothing for $q + 1 \geq Q_0$. However, since there are actually only q of them, this investment would be insufficient to implement the action. To see this, let the deviating opponent's spending against the action be $y \geq 0$, so then $R + qx(q + 1) - (1 - x_0)(N - q - 1) - y = \theta - q[x(q) - x(q + 1)] + 1 - x_0 - y$, where the first equality follows from $R - (1 - x_0)(N - q) = \theta - qx(q)$. But now we get $\theta - q[x(q) - x(q+1)] + 1 - x_0 - y < \theta \Leftrightarrow 1 - x_0 - y < q[x(q) - x(q + 1)]$. Since any $y \leq 1 - x_0$ is possible, this inequality can always be satisfied for sufficiently large y because the left-hand side will get arbitrarily close to zero whereas the right-hand side is strictly positive. Thus, the deviating player will always be able to derail the implementation of the action in any period in which supporters are spending a positive amount. At the very least, he can ensure that it gets implemented only with probability $1/2$.[34] Thus, his payoff from a deviation would be:

$$\begin{aligned}
\hat{U}_o &= \sum_{k=0}^{Q-2} (1-w) f(k) + \sum_{k=Q-1}^{Q_0-1} \left(1 - x_0 - y(k) - \frac{a}{2}\right) f(k) + \sum_{k=Q_0}^{N-1} (1 - x_0 - a) f(k) \\
&= 1 - w - \sum_{k=Q-1}^{Q_0-1} \left(\frac{a}{2} + x_0 - w + y(k)\right) f(k) - (a + x_0 - w) \sum_{k=Q_0}^{N-1} f(k).
\end{aligned}$$

Using the supposition that $Q \leq Q_0 - 1$, we obtain:

$$\begin{aligned}
\hat{U}_o - U_o &= (a + x_0 - w) \sum_{k=Q}^{Q_0-1} f(k) - \sum_{k=Q-1}^{Q_0-1} \left(\frac{a}{2} + x_0 - w + y(k)\right) f(k) \\
&= \sum_{k=Q}^{Q_0-1} \left[\frac{a}{2} - y(k)\right] f(k) - \left[\frac{a}{2} + x_0 - w + y(Q-1)\right] f(Q-1).
\end{aligned}$$

Note now that the first term is positive and strictly increasing in Q_0. Even though it might be possible to keep this inequality non-positive (as would have to be the case to make the deviation unprofitable), it would depend on the distribution of the exogenous parameters, and at any rate it would not hold for Q_0 sufficiently larger than Q. It is for this reason that we will focus on equilibria in which $Q_0 = Q$: these would be least susceptible to deviations. ■

[34] He cannot fully block the action because the opponent only has $1 - x_0$ available when the agent tries to implement the action.

Proof (Proposition 5). Consider first the continuation game after the vote. Whenever the agent invests toward the action, it will succeed because $x_0(Q)$ ensures that any groups of opponents at $q \geq Q$ does not have enough resources left to derail it (even though supporters consume privately). If $q < Q$, the agent reimburses the players. Should the supporters now attempt to impose the action, $q < \overline{Q}_{\mathrm{a}}$ ensures that the opponents will have enough resources to block it, and since they move after the supporters, they will do so. Thus, neither opponents nor supporters have an incentive to deviate after the vote.

We now examine the voting stage given that the continuation game after the vote will be played according to the equilibrium strategies. Consider a player who learns that he is an opponent. If he votes sincerely, the action will be implemented if there are $q \geq Q$ supporters among the remaining $N-1$ players. Using (9) and using (NBC) to obtain

$$x_0(Q) - w = \frac{(1-w)(N-Q)+\theta}{2N-Q} \equiv \hat{x}(Q),$$

we get

$$U_o = 1 - w - (a + \hat{x}(Q)) \sum_{k=Q}^{N-1} f(k). \tag{10}$$

If he deviates and votes in favor, the agent will attempt implementation if $q \geq Q-1$ of the remaining players also vote in favor. Since the actual amount invested in support will remain the same, nothing will change for any realizations with $q \geq Q$ supporters among the remaining players. The only possible difference comes from the situation with $q = Q-1$ where had the opponent voted sincerely the agent would not have attempted to invest in the action. Now that the agent would, the opponent has to block it. At best, he can expect $1 - x_0(Q) < 1 - w$, which is what he would have gotten had he just voted sincerely. The best deviation payoff would be:

$$\begin{aligned}\hat{U}_o &= \sum_{k=0}^{Q-2}(1-w)f(k) + (1-x_0(Q))f(Q-1) + \sum_{k=Q}^{N-1}(1-x_0(Q)-a)f(k) \\ &= 1 - w - \hat{x}(Q)f(Q-1) - (a+\hat{x}(Q))\sum_{k=Q}^{N-1} f(k) < U_o.\end{aligned}$$

Thus, such a deviation is unprofitable, and any opponent has a strict incentive to vote sincerely.

Consider now a player who learns that he is a supporter. If he votes sincerely, the action will be implemented if there are $q \geq Q-1$ supporters among the remaining players, and his payoff would be:

$$U_s = \sum_{k=0}^{Q-2}(1-w)f(k) + \sum_{k=Q-1}^{N-1}(1-x_0(Q)+a)f(k) = 1-w+(a-\hat{x}(Q))\sum_{k=Q-1}^{N-1} f(k). \tag{11}$$

If he deviates and votes against the action, then the agent will attempt implementation when there are $q \geq Q$ supporters among the remaining players. Since he will not even try to implement the action with fewer votes, there is no point in the supporter spending anything toward it. Since the action will succeed in all other cases, his payoff will simply be:

$$\hat{U}_s = \sum_{k=0}^{Q-1}(1-w)f(k) + \sum_{k=Q}^{N-1}(1-x_0(Q)+a)f(k) = 1-w+(a-\hat{x}(Q))\sum_{k=Q}^{N-1}f(k) < U_s,$$

making this deviation unprofitable. Thus, any supporter has strict incentives to vote sincerely as well. ■

Proof (Lemma 5). Delegating with Q means that every player contributes $x_0(Q)$, votes sincerely after observing his preference, and consumes privately. The agent commits the resources toward the action if there are $q \geq Q$ supporting votes and reimburses the players (net his fee) otherwise. Define $\hat{x}(Q) = x_0(Q) - w$ to be the portion of the contribution that can be used for implementation. For any agreed-upon Q, the *ex ante* expected payoff to player i is:

$$\begin{aligned}
U_{\text{a}} = {} & p\left[1 - w + (a - \hat{x}(Q))\sum_{k=Q-1}^{N-1} f(k)\right] \\
& + (1-p)\left[1 - w - (a + \hat{x}(Q))\sum_{k=Q}^{N-1} f(k)\right] \\
= {} & 1 - w + p(a - \hat{x}(Q))f(Q-1) + [(2p-1)a - \hat{x}(Q)]\sum_{k=Q}^{N-1} f(k), \qquad (12)
\end{aligned}$$

where we used (11) for the payoff in case he turns out to be a supporter (with probability p), (10) for the payoff in case he turns out to be an opponent (with probability $1-p$). To see how U_{a} changes with Q, note that:

$$\begin{aligned}
U_{\text{a}}(Q+1) - U_{\text{a}}(Q) = {} & (1-p)\,[a + \hat{x}(Q+1)]\, f(Q) \\
& - p\,[a - \hat{x}(Q)]\, f(Q-1) + [\hat{x}(Q) - \hat{x}(Q+1)]\sum_{k=Q}^{N-1} f(k)
\end{aligned}$$

or, with $\gamma(Q) = \hat{x}(Q) - \hat{x}(Q+1)$, and $\beta(Q) = a(N-2Q) + N\hat{x}(Q+1) + Q\gamma(Q)$,

$$\begin{aligned}
= {} & \beta(Q)\left[\frac{(N-1)!}{Q!(N-Q)!}\right] p^Q (1-p)^{N-Q} \\
& + \gamma(Q)\sum_{k=Q}^{N-1}\left[\frac{(N-1)!}{k!(N-1-k)!}\right] p^k (1-p)^{N-1-k},
\end{aligned}$$

where we note that

$$\gamma(Q) = \frac{(1-w)N - \theta}{(2N-Q)(2N-Q-1)} > 0.$$

Thus, $U_{\mathrm{a}}(Q+1) - U_{\mathrm{a}}(Q) \gtreqless 0$ if, and only if,

$$\beta(Q)\left[\frac{(N-1)!}{Q!(N-Q)!}\right]p^{Q}(1-p)^{N-Q} + \gamma(Q)\sum_{k=Q}^{N-1}\left[\frac{(N-1)!}{k!(N-1-k)!}\right]p^{k}(1-p)^{N-1-k} \gtreqless 0,$$

or, after dividing both sides by $(N-1)!\,p^{Q}(1-p)^{N-Q}$, if, and only if,

$$\frac{\beta(Q)}{Q!(N-Q)!} + \left[\frac{\gamma(Q)}{1-p}\right]\sum_{k=Q}^{N-1}\left[\frac{1}{k!(N-1-k)!}\right]\left(\frac{p}{1-p}\right)^{k-Q} \gtreqless 0.$$

We re-index the summation term and multiply both sides by $Q!(N-Q)!$ to obtain:

$$\beta(Q) + \left[\frac{\gamma(Q)}{1-p}\right]\sum_{i=0}^{N-1-Q}\left[\frac{Q!(N-Q)!}{(Q+i)!(N-1-Q-i)!}\right]\left(\frac{p}{1-p}\right)^{i} \gtreqless 0.$$

Using the definition of $\beta(Q)$, and dividing both sides by Q, we can rewrite this as:

$$\frac{aN + (N-Q)\hat{x}(Q+1)}{Q} + \hat{x}(Q) + \left[\frac{\gamma(Q)}{Q(1-p)}\right]\sum_{i=0}^{N-1-Q}\left[\frac{Q!(N-Q)!}{(Q+i)!(N-1-Q-i)!}\right]\left(\frac{p}{1-p}\right)^{i} \gtreqless 2a. \quad (13)$$

Observe now that all three terms on the left-hand side of this inequality are positive. Furthermore, at $Q = 1$ the left-hand side is strictly larger because it reduces to aN plus three non-negative terms and $N \geq 2$. Thus, at $Q = 1$, the difference is strictly positive, so the payoff function is increasing. We now prove that the function is concave. For this, we only need to show that the left-hand side of (13) (which is essentially the first derivative of U_{a}) is decreasing in Q. First, note that

$$\hat{x}(Q) = \frac{(1-w)(N-Q)+\theta}{2N-Q} \Rightarrow \frac{\mathrm{d}\hat{x}(Q)}{\mathrm{d}Q} = \frac{\theta-(1-w)N}{(2N-Q)^2} < 0,$$

where the inequality follows from $w < \overline{w}$. This means that the first two terms on the left-hand side of (13) are decreasing in Q. If $Q > N-1$, then the third term is zero, and the claim holds. Consider then $Q \leq N-1$. We now wish to show that the third term decreases as well. Letting

$$D(Q) = \left(\frac{1}{1-p}\right)\left[\frac{\gamma(Q)}{Q}\right]\sum_{i=0}^{N-1-Q}\left[\frac{Q!(N-Q)!}{(Q+i)!(N-1-Q-i)!}\right]\left(\frac{p}{1-p}\right)^{i},$$

we note that $D(Q+1) - D(Q) < 0$ if, and only if,

$$\sum_{i=0}^{N-1-(Q+1)} \left[\frac{\gamma(Q+1)(Q+1)!(N-(Q+1))!}{(Q+1)(Q+1+i)!(N-1-(Q+1)-i)!} \right.$$
$$\left. - \frac{\gamma(Q)Q!(N-Q)!}{Q(Q+i)!(N-1-Q-i)!} \right] \left(\frac{p}{1-p} \right)^i$$
$$- \left[\frac{\gamma(Q)Q!(N-Q)!}{Q(N-1)!} \right] \left(\frac{p}{1-p} \right)^{N-1-Q} < 0.$$

Since the second term is positive but is being subtracted, the inequality must hold whenever the summation is negative. Simplifying the summation, this requirement becomes:

$$\sum_{i=0}^{N-1-(Q+1)} \left(\frac{p}{1-p} \right)^i \left[\frac{Q!(N-(Q+1))!}{(Q+i)!(N-1-(Q+1)-i)!} \right]$$
$$\times \left[\frac{\gamma(Q+1)}{Q+1+i} - \frac{\gamma(Q)(N-Q)}{Q(N-1-Q-i)} \right] < 0,$$

and since the first two multiplicative terms in this summation are positive, the inequality will certainly hold if the third term is negative. But since the first term in that expression is decreasing in i while the second one is increasing, it is sufficient to show that the inequality holds at $i = 0$, or that

$$\frac{\gamma(Q+1)}{Q+1} - \frac{\gamma(Q)(N-Q)}{Q(N-1-Q)} < 0,$$

because if this is true, then the term above will be negative for any $i > 0$ as well. Rearranging terms gives us:

$$Q(N-1-Q)\left[\gamma(Q+1) - \gamma(Q)\right] < \gamma(Q)N.$$

Using the definition of $\gamma(Q)$, dividing both sides by $(1-w)N - \theta$, and multiplying them by $2N - Q - 1$ gives us:

$$Q(N-1-Q)\left(\frac{1}{2N-Q-2} - \frac{1}{2N-Q} \right) < \frac{N}{2N-Q},$$

which, after simplifying and multiplying both sides by $2N - Q$, yields:

$$2Q(N-1-Q) < N(2N-Q-2)$$

or, after adding and subtracting QN on the right-hand side and re-arranging terms,

$$2(N-Q) < 2(N-Q)^2 + QN,$$

which simply reduces to

$$0 < 2(N-Q)(N-Q-1) + QN,$$

which holds because we have been considering the case with $Q \leq N-1$. Thus, all three terms on the left-hand side of (13) are decreasing in Q. We conclude that the payoff function is concave, which implies that it has a unique maximizer, which we denote $Q_{\text{a}}^{*}(w, p)$. (It is the smallest integer for which the left-hand side of (13) is less than the right-hand side.) It then follows immediately that the optimal quota is either at this interior solution or at the point where (NIC) binds: $Q_{\text{a}}(w, p) = \min(Q_{\text{a}}^{*}(w, p), \overline{Q}_{\text{a}})$.

We finally show that $Q_{\text{a}}(w, p)$ is non-decreasing in p. Since only the interior solution depends on p, we only need to prove the claim for $Q_{\text{a}}^{*}(w, p)$. From the FOC given by (13), it is sufficient to show that the summation term (the only one involving p) is increasing in p. Taking the derivative of that term with respect to p produces

$$\left[\frac{\gamma(Q)}{Q}\right] \sum_{i=0}^{N-1-Q} \left[\frac{Q!(N-Q)!}{(Q+i)!(N-1-Q-i)!}\right] \left[\frac{p^{i}}{(1-p)^{2+i}}\right] \left(1+\frac{i}{p}\right) > 0,$$

so the claim holds. To see why this is so, fix some p and consider the optimum $Q_{\text{a}}^{*}(w, p)$, which is the smallest integer for which the left-hand side of (13) is less than the right-hand side (that is, increasing the quota would make the payoff worse). If increasing p causes the left-hand side to increase, it will eventually exceed the right-hand side for some $\hat{p} > p$. But then $Q_{\text{a}}^{*}(w, \hat{p})$ will no longer be the smallest integer that makes the left-hand side less than the right-hand side (i.e., it will no longer be optimal). Since the left-hand side is decreasing in Q, the requirement for optimality can be restored by increasing the quota to $Q_{\text{a}}^{*}(w, \hat{p}) = Q_{\text{a}}^{*}(w, p) + 1$, which will make the left-hand side less than the right-hand side again. Continuing in this manner, we see that increasing p will cause the quota to increase in step-wise fashion until it reaches the ceiling $\overline{Q}_{\text{a}}$. ∎

Proof (Lemma 6). Note now that $\lim_{p \to 1} U_{\text{a}} = 1 - w + a - \hat{x}(Q_{\text{a}})$. This is strictly preferable to private consumption whenever this is greater than 1, or, after rearranging terms, whenever $aN + (a-1)(N - Q_{\text{a}}) > wN + \theta$. Since $N \geq Q_{\text{a}}$ and $a - 1 > 0$, the second term on the left-hand side is non-negative at the optimum quota. It then follows that it is sufficient to establish that $aN > wN + \theta$ holds. Since the right-hand side is increasing in w, we only need to establish the claim at $\overline{w}$, where it reduces to $aN > \overline{w}N + \theta = N \Leftrightarrow a > 1$, which holds. ∎

Proof (Proposition 6). Since the strategies are unconditional, deviation does not affect future play, and the discount factor is irrelevant. The only possibly profitable deviation is therefore limited to the stage-game. However, since delegation with Q_{a} is preferable to private consumption and because the strategies from Proposition 5 specify an equilibrium in the stage-game, no such deviation exists. ∎