

Organizing for Collective Action: Olson Revisited*

Marco Battaglini

Thomas R. Palfrey

Cornell University

California Institute of Technology

July 26, 2024

Abstract

We characterize optimal honest and obedient HO mechanisms for the classic collective action problem with private information, where group success requires costly participation by some fraction of its members. For large n , a simple HO mechanism, the Volunteer Based Organization, is approximately optimal. Success is achieved in the limit with probability one or zero depending on the rate at which the required fraction declines with n . For finite n , optimal HO-mechanisms provide substantial gains over unorganized groups when the success probability converges to zero, because the optimal HO success probability converges slowly and is always positive, while finite-sized unorganized groups have exactly zero probability of success.

*Support from the National Science Foundation (SES-2343948) is gratefully acknowledged. The paper has benefited from comments and suggestions by the editor and participants of the 2021 ETH Workshop on Theoretical Political Economy, 2021 SITE Workshop on Political Economic Theory, 2023 POLECONUK, 2023 Conference on Economic Theory at the Cowles Foundation, ESSET 2023, and seminar audiences at Bocconi University, Caltech, Columbia University, CUNEF Universidad, Oxford University, Stanford University, TU Dortmund, UC3M, UC Berkeley, University of Bern, University of Essex, University of Padova, and University of Zurich. We are especially grateful to the referees for their detailed comments and suggestions. We also wish to thank Felix Bierbrauer, Wioletta Dziuda, Hans Gersbach, John Ledyard, David Levine, and Andrew Postlewaite for discussions and comments. The research assistance of Francesco Billari, Francesco Billotta and Julien Neves is gratefully acknowledged. We are responsible for any remaining shortcomings. Edited by Emir Kamenica.

1 Introduction

Collective action is one of the most basic and ubiquitous forms of strategic interaction in societies. Examples of collective action problems range from the case of private citizens banding together in public demonstrations; to dissatisfied workers participating in union activities; to voters bearing up against bad weather to cast their ballots; to community members donating their time to organize charity or cultural events. At a more macro level, the choices by countries contemplating to join an international environmental agreement also constitute a collective action problem. These are all instances of environments in which a common goal can be achieved by a community, but only if a sufficiently large number of its members are willing to make individual contributions, thus overcoming the incentives to free ride. There are many concrete examples testifying that societies are indeed able to partially solve collective action problems; theories of voluntary behavior and free riding, however, find hard to explain significant levels of individual participation, except assuming that citizens like it or feel morally obliged to it.

In his seminal work, Mancur Olson [1965] provided a taxonomy of the factors determining success of collective action, and highlighted the presence of an organization as a key factor. This observation is intuitive, but it opens up practical and theoretical questions that, as we will argue, have not yet been fully explored in the literature. A first set of questions is positive: what type of organizations should we plausibly expect in collective action problems, and how effective should we expect them to be? A second related set of questions is normative: how do empirically plausible organizations compare to the theoretically optimal organization? To what extent can the presence of an organization (plausible or even optimal) explain the observed effectiveness of collective action even with a large number of agents? Understanding these questions is important to make sense of the limits and opportunities of collective action and may provide normative insights to improve it.

In this paper we make progress on these issues by studying the effectiveness of organizations in a classic threshold contribution game, widely studied in economics, biology, political science and sociology. In the game, a group of n agents pursue a collective goal that, if achieved, generates a benefit v per agent. The goal is achieved if at least a m_n out of n

agents choose to make a personal contribution.¹ The cost of a personal contribution is private information to each agent: it is independently distributed across agents according to some commonly known distribution, $F(c)$. The agents may or may not have an organization to coordinate individual decisions, and the organization may be strong (allowing for transfers and/or some form of coercion) or, more plausibly, weak (no transfers and no coercion). We ask how the probability of success changes as n increases, depending on (1) the rate of increase in m_n , (2) whether or not there is an organization, and (3) whether the organization, if it exists, is strong or weak. We also ask under what conditions the group of agents will endogenously form an organization.

Our analysis produces four new theoretical insights. As a preliminary step, we first revisit the equilibrium analysis without an organization when the threshold m_n is a general increasing function of n . Our first finding is that, even without an organization and with a threshold m_n that grows to infinity, failure of collective action is not inevitable: the key factor is the rate of increase of m_n versus n . Perhaps surprisingly, we show that, regardless of the shape of $F(c)$, success is achieved with probability converging to one if m_n grows at a rate slower than $n^{2/3}$; success is instead impossible if n grows faster than $n^{2/3}$. When m_n grows faster than $n^{2/3}$, moreover, there is a critical group size, n_U , such that the probability of success falls precipitously from a strictly positive success probability to becomes exactly zero for $n \geq n_U$. Collective action, therefore, does not require an organization to be successful if m_n grows sufficiently slowly; but it can be really valuable otherwise.

The other three main findings address the questions of how and to what extent the performance of collective action can be improved by an organization. The key issue here is how to model an organization. The standard approach in mechanism design theory has focused on the study of optimal organizations with transfers that are Bayesian incentive compatible (IC) and interim individually rational (IR), what we refer to as *strong mechanisms*.² This

¹Classic contributions are Palfrey and Rosenthal [1984] in economics and Diekmann [1985] in sociology, who coined the term the “volunteer’s dilemma” for the special case in which $m_n = 1$. A survey of the work using these games in biology is presented by Archetti and Scheuring [2012]. Applications to environmental economics include, for example, Tavoni et al. [2011] and Barrett et al [2014]. Recent contributions in economics included Harrington [2011], Bergstrom [2017], Battaglini and Benabou [2003], Battaglini [2017], Nöldeke and Peña (2020), Dziuda et al. [2021], among others.

²An exception is Dixit and Olson [2000], as we will discuss below.

approach, through the (IC) constraint, captures the problem of honestly aggregating the dispersed private information regarding the agents' types; it also partially captures, through the IR constraint, a moral hazard problem at the interim stage by guaranteeing a minimal *expected* utility to all types. In most environments of interest, however, this approach bypasses the moral hazard problem faced by the group, since some types might choose to disobey a recommendation by the mechanism if carrying out the recommendation would not be optimal. In the standard Bayesian mechanism design problem, a direct mechanism maps each reported profile of types to an allocation and a payment to or from each agent, which is then imposed on all agents even if the allocation/payment makes some agents worse off at the reported type profile. In contrast, in a collective action problem, a mechanism lacks the power to simply impose the outcome on all agents, and can only suggest recommended (i.e., not imposed) actions, one for each agent ("go protest", "sign a petition", "volunteer", "do nothing", etc.). The final outcome ultimately depends on the individual willingness of each agent to voluntarily carry out these recommendations.

To clarify this point with an example, consider a community asking for volunteers to organize an event. The event requires at least 3 out of 10 agents to spend one afternoon at the community center and yields a value $v = 0.5$ per person if the quota is met. If the support of F is $[0, 1]$, then a simple (IC) and (IR) mechanism can achieve the goal with probability 1 using a simple random mechanism:³ just randomly and anonymously select 3 agents and ask them to volunteer. This is (IC), since the information on the types is not used; and it is (IR) since the interim expected cost, $0.3c$, is lower than the benefit, v , even for the highest type ($0.3 < 0.5$). The problem with this mechanism is that it violates the moral hazard (obedience) constraint: no type $c > 0.5$ would agree to volunteer if asked.⁴ In such a situation one must add an obedience constraint, requiring that the agents asked to volunteer find it optimal to carry out the mechanism's recommendation. We refer to mechanisms that also satisfy obedience and no transfers as *weak mechanisms*.

This distinction between (IC) and (IR) mechanisms and honest and obedient mechanisms

³This mechanism is also optimal under some weak conditions as we will see in Section 3. This fact is however irrelevant for the present discussion.

⁴This random mechanism is also *ex post* IR, since outcomes do not depend on the profile of types. All types are willing to participate, even after conditioning on the entire type profile.

was not especially important for limiting results with large groups in the early literature that assumed constant returns to scale; i.e., the cost of the common project grows linearly with the number of agents. In that case, the limit probability of success is zero even if we ignore the obedience constraint (Rob [1989], Mailath and Postlewaite [1990], Ledyard and Palfrey [1994, 1999]). If one generalizes the constant returns assumption, however, the distinction becomes important. As we are able to show, when m_n grows slower than n , even if at a speed arbitrarily close to n , then optimal (IC) and (IR) mechanisms achieve a probability one of success for a large enough n even if we adopt a simple random mechanism with no transfers as outlined before. Such mechanisms, however, violate the obedience constraint, as explained in the example above. It therefore becomes important to understand what can be achieved with an honest and obedient mechanism.

Our second theoretical contribution is to show that a simple class of honest and obedient mechanisms that we call *Volunteer Based Organizations* (VBO) is asymptotically optimal. The mechanism is a simple extension of the random mechanism described above. In a VBO, agents are asked to report whether they are willing to be activated (*volunteers*) or not (*free riders*). If the number of agents who state they are willing to be volunteers is lower than m_n , then no agent is asked to be active and the group fails, but wastes no cost of action by any agent. If the number of volunteers is greater than or equal to m_n , then the collective goal is achieved by randomly and anonymously selecting m_n volunteers. These volunteers are willing to follow the recommendation because they know that exactly $m_n - 1$ volunteers will also carry out similar recommendations. Free riders are never asked to be active.

In our third theoretical result we use the previous characterization to explore the limits of optimal honest and obedient organizations. This allows us to extend the negative limit results of the previous literature: i.e., that the limit probability of success in an optimal (IC) and (IR) with constant returns to scale is zero, both with an organized and unorganized group (Rob [1989], Mailath and Postlewaite [1990], Ledyard and Palfrey [1994, 1999, 2002]). We show that with honest and obedient mechanisms the limiting probability of success is the same with an organized and an unorganized group for *any* rate of growth of m_n : as with the Bayesian Nash equilibrium for unorganized groups, the optimal honest and obedient mechanism achieves a limiting success probability equal to 1 if m_n grows at a rate slower

than $n^{2/3}$; and it achieves a limit probability equal to 0 if it grows faster than $n^{2/3}$. An implication of this result is that the probability of success converges to zero even if the total group benefit is strictly greater than the maximum possible total cost, a case in which, as we will show, strong mechanisms are successful with probability one even with no transfers.

So is there any value in having an organization? The fourth lesson from our analysis is that organizations are indeed very useful, and that focusing only on limit results for infinite-sized groups misses an important part of the problem. We show that even when m_n grows faster than $n^{2/3}$, the limit probability of success in a honest and obedient organization converges to zero at a rate that is infinitely slower than without an organization (which indeed achieves exactly zero probability after a finite threshold, \bar{n}).

Taken together these results confirm and sharpen Olson's intuition for the importance of organizations for collective action, and also highlight important limitations to the power of organizations. Simple forms of cooperation such as a VBO, however, are approximately optimal for finite n and can provide an effective institution for group success for large groups, at least on the order of thousands of members, even in environments where the limiting probability of success with extremely large groups approaching infinity would be zero. This observation may help explain why numerous cases of successful collective action have been documented, even if collective action is not a panacea for all social problems.

Regarding the limitations to the power of weak organizations, our $n \rightarrow \infty$ results have implications about solving collective action problems in very large societies, for example on the scale of nation states with tens of millions of citizens. An extremely large society, even one that is ideally organized in a way that respects honesty and obedience constraints, will perform no better in providing public goods than an identical society operating under autarky, with no organization at all. Voluntary behavior, even if guided by a perfectly designed organization to coordinate activity, is not sufficient to provide a satisfactory solution to collective action problems for arbitrarily large societies. In these cases, what is required is the establishment of institutions (e.g., a government) that are enabled with the authority of ex post coercive powers to implement and enforce individual compliance with the outcomes imposed by a strong mechanism.

These implications are confirmed and strengthened when we endogenize the formation of an organization, as we do in Section 6. That analysis suggests that even when successful collective action is only possible with an organization, we should observe the formation of organizations only for values v larger than a threshold $v(n)$, increasing in n .

Related literature Olson [1965] was arguably the first to highlight the importance of an organization in solving collective action problems, providing a first informal description of the features of an organization useful to solve them.⁵ Formal analysis of this issue, however, had to wait for the development of the theory of optimal mechanisms in Bayesian environments. Our work follows this tradition, departing from it in two ways: first, because we assume no transfers; second, and most importantly, because we require the optimal mechanism to be honest and obedient as discussed above. Most previous theoretical research on optimal mechanisms for public good provision in Bayesian environments consider only strong organizations, which allow unlimited side payments and interim individual rationality constraints, ignoring the obedience constraint (Mailath and Postlewaite [1990], [1989], Ledyard and Palfrey [1994], [1999], [2002], Hellwig [2003]).⁶ As far as we know, the problem of optimal public goods mechanisms in Bayesian environments that satisfy obedience has never been studied. The first three groups of authors have presented negative results of strong organizations assuming constant returns to scale, showing that limit probabilities converge to zero with or without an optimal (IC) and interim (IR) mechanism. Hellwig [2003] has shown that with increasing returns, limit probabilities equal to 1 are feasible with an optimal (IC) and (IR) mechanism with unlimited transfers, indeed always achieved when the demand for the public good is bounded above. When we consider honest and obedient mechanism, results are very different, both with constant returns and without. Allowing for increasing returns, we extend the insight that organizations are not useful in the limit, since we show

⁵Other factors for the success of collective action that have been emphasized by Olson [1965] and the following literature include the cohesiveness of the preferences of the group's members, the elasticity of their cost function as a function of the contributions, and the degree of excludability of the common goal's benefits. Important works on these dimensions see Chamberlin [1974], and more recently Esteban and Ray [2001].

⁶Other research on Bayesian mechanism design with public goods analyzes "super-strong" organizations that require incentive compatibility, but allow for unlimited side payments and no participation constraints (d'Aspremont and Gerard-Varet [1979], Crémer and McLean [1985], d'Aspremont, Crémer, and Gerard-Varet [1990], Ledyard and Palfrey [1999], [2002]).

with HO mechanisms they can only obtain the same limit probabilities of success than unorganized groups as $n \rightarrow \infty$: with sufficiently increasing returns, however, this probability may be one both with and without an organization, a case that we precisely characterize. With constant returns, we also show that the failure of organizations in the limit is a more severe phenomenon than previously believed, since it extends to cases in which the total societal value of the collective action goal is strictly higher than the cost of achieving success in the worst case scenario, i.e. $vn > m_n \bar{c}$, where \bar{c} is the maximum possible cost. In contrast, success is guaranteed with strong organizations in this worst case.⁷

Following Olson [1965], a significant literature has also studied organizations for collective action from a positive perspective, providing empirical studies of the type of organizations that emerge in concrete examples, both using case studies (Ostrom [1990], for instance) and laboratory experiments (De Kragt et al [1983], Braver and Wilson [1986], Palfrey and Rosenthal [1991], Ostrom and Walker [1991], Ostrom et al. [1992], Palfrey et al. [2017] among others). Several of these experimental papers study public good games similar to ours, by allowing players to communicate before the contribution stage and ruling out coercion, and report the endogenous emergence of mechanisms similar to the VBO mechanism that we show to be asymptotically optimal.

2 The Collective Action Model

2.1 Setup

A group with n members, $I = \{1, 2, \dots, n\}$, desires an outcome generating a total value of W_n , with each member in the group receiving a personal, direct benefit of $v = W_n/n \in (0, 1)$

⁷The limits of organizations for collective action are also explored by Dixit and Olson [2000], who focus on the study of incentives to join organized groups. They take a cooperative perspective, assuming that organizations achieve the efficiency frontier through Coaseian bargaining; agents, however, have incentives to stay out, free riding on those who join the organization (for a similar approach in a dynamic setting, see also Battaglini and Harstad [2016]). Passarelli and Tabellini [2017] present a model of political unrest that incorporates psychological rewards for activism. Besides the contributions cited above, moreover, a recent significant literature has studied the limits of organizations in Bayesian mechanisms. See, for example, Healy [2010], Goldlucke and Troger [2020], and Bierbrauer and Hellwig [2016] and Bierbrauer and Winkelmann [2020].

which is independent of n .⁸ The policy is obtained if and only if at least $m_n > 1$ out of the $n \geq m_n$ members of the group are active. The fraction of agents that are required to be active for success is denoted by $\alpha_n = m_n/n \in (0, 1)$.

Different members have different activity costs, and we denote by c_i the cost of being active for member i . Member i 's payoff is given by:

$$\begin{aligned} u_i &= 0 \text{ if } i \text{ is not active and fewer than } m_n \text{ members are active} \\ &= v \text{ if } i \text{ is not active and at least } m_n \text{ members are active} \\ &= -c_i \text{ if } i \text{ is active and fewer than } m_n \text{ members are active} \\ &= v - c_i \text{ if } i \text{ is active and at least } m_n \text{ members are active} \end{aligned}$$

Costs are i.i.d. and distributed in $[0, \bar{c}]$ according to a distribution $F(c)$ with density $f(c)$. We normalize without loss of generality $\bar{c} = 1 > v$ and assume $0 < f(c) < \bar{f}$ for some bound $\bar{f} < \infty$ and all $c \geq 0$.⁹

We do not need to assume that m_n is monotonic in n , though typically we expect it to be non decreasing with $m_n \rightarrow \infty$ as $n \rightarrow \infty$.¹⁰ We refer to the case in which $m_n = \alpha n$ for some fixed constant $\alpha \in (0, 1)$ as the *constant returns to scale* case, since it represents a situation in which the fraction of active members required for success, m_n/n , is constant in n (or equivalently converges to a constant). We refer to the case in which $m_n/n \rightarrow 0$ as the *increasing returns to scale* case, since in this case the average cost of the common goal declines in n .

There are two basic forms of organization of the group. The first is no organization at all. In this case each member decides to be active or to free ride independently, given rational expectations about the other members activity decisions. This corresponds to a pure voluntary participation game with a threshold.

The second form is an organized group. We are interested in studying the benefits from

⁸It is straightforward to extend the analysis to the case in which we have a value v_n depending on n and $v_n \rightarrow v$ as $n \rightarrow \infty$.

⁹The analysis directly extends to more general environments. We discuss alternative assumptions about the cost distribution in Section 7.

¹⁰We give an example in the next section of a case where m_n is a constant for all $n \rightarrow \infty$.

organizing even when the organization has very limited tools at its disposal. To this end, we assume that the organization cannot directly observe the types of its members, it cannot exert any form of coercion on the members' actions and it cannot even commit to monetary transfers. We refer to such organizations as *weak organizations*. The organized group can only design an optimal communication mechanism. In such a mechanism, group members send messages to the mechanism, then the mechanism sends each member a recommended action, and then each member independently chooses an action. While the set of such mechanisms can be very large, Myerson [1982] has shown that the characterization of the set of all Bayesian Nash equilibria of all such communication mechanisms can be accomplished by considering only honest and obedient direct communication mechanisms.¹¹ In Section 2.2.1 below we provide a formal characterization of this class of mechanisms and its relationship to the class of incentive compatible and individually rational mechanisms.

In Section 6 we describe a stronger form of organization in which only incentive compatibility and interim individual rationality is required. This class of mechanism is a useful benchmark since the previous literature has focused on these mechanisms in the form presented here or in close variants (Rob [1989], Mailath and Postlewaite [1990], Ledyard and Palfrey [1994, 1999, 2002] and Hellwig [2003]). We refer to these as *strong organizations*.

2.2 Weak Organizations: Honest and Obedient Communication Mechanisms

In the absence of monetary transfers, a direct communication mechanism is fully characterized by a mapping from the set of possible type profiles into the set of probability distributions over the subsets of I , $\mu : [0, 1]^n \rightarrow \Delta(2^I)$, where we call μ either the *mechanism* or the *activity function*, $\Delta 2^I$ is the set of probability distributions over subsets of I , and we denote by $\mu_g(\mathbf{c})$ the probability the activity function selects subset $g \subseteq I$ of the group to be active at type profile \mathbf{c} . Members independently report their types to the mechanism; given the messages \mathbf{c} the mechanism selects a coalition g to activate according to $\mu_g(\mathbf{c})$ and sends the corresponding recommended action to each member; then each member observes their

¹¹This set is closely related to the set of correlated Bayesian equilibrium outcomes of the game.

own recommendation and decides whether to comply.

In the following it is sometimes useful to denote a coalition $g \subseteq I$ as an n -dimensional vector of zeros and ones, in which the i th component, g_i , is equal to 1 if $i \in g$ and equal to 0 if $i \notin g$. In this notation $(g_{-i}, 0)$ is a coalition with g_{-i} that excludes i ; and $(g_{-i}, 1)$ is the coalition of g_{-i} plus i . We denote $|g| = \sum_i g_i$.

Define $I_i = \{g \subseteq I \mid i \in g\}$ as the subsets of I containing i and define $I^{m_n} = \{g \subseteq I \mid |g| \geq m_n\}$ as the set of subsets containing at least m_n members. Given an activity function, μ , the probability i is active at type profile \mathbf{c} is given by $A_i(\mathbf{c}; \mu) = \sum_{g \in I_i} \mu_g(\mathbf{c})$, and the probability that enough members are active so the group is successful is given by $P(\mathbf{c}; \mu) = \sum_{g \in I^{m_n}} \mu_g(\mathbf{c})$. A mechanism is *balanced* if and only if, for all \mathbf{c} , $\mu_g(\mathbf{c}) > 0 \Leftrightarrow |g| = m_n$. A mechanism has *undercontribution at \mathbf{c}* if $\mu_g(\mathbf{c}) > 0$ for some $|g| < m_n$ and a mechanism has *overcontribution at \mathbf{c}* if $\mu_g(\mathbf{c}) > 0$ for some $|g| > m_n$. Thus a mechanism is balanced if and only if it never has overcontribution or undercontribution.

For any mechanism μ define its *reduced form mechanism* by the functions $p_i(c_i) = E_{\mathbf{c}_{-i}}[P((c_i, \mathbf{c}_{-i}); \mu)]$ and $a_i(c_i) = E_{\mathbf{c}_{-i}}[A_i(c_i, \mathbf{c}_{-i}); \mu]$, which are, respectively the expected probability of success and the expected probability i is active, condition on i 's cost. We assume, without loss of generality, that the mechanism is symmetric, i.e., for any $i, j \in I$, $c \in [0, 1]$, $p_i(c) = p_j(c)$ and $a_i(c) = a_j(c)$.¹² To simplify notation, we drop the member subscripts and simply write these reduced form functions as $p : [0, 1] \rightarrow [0, 1]$ and $a : [0, 1] \rightarrow [0, 1]$. We call a reduced form mechanism, (p, a) , *feasible* if and only if there exists an activity function μ that generates (p, a) . Given any activity function, μ , the interim expected utility for type c who reports to be a type c' is denoted by $U(c', c) = vp(c') - ca(c')$ with $U(c) \equiv U(c, c)$.

In Myerson [1982] a coordination mechanism is honest and obedient (HO) if it provides incentive to reveal the true type and to follow the recommendations of the mechanism. Define $\chi(g)$ as the success indicator function when a coalition $g \subseteq I$ is activated: so $\chi(g) = 1$ if

¹²To see that the restriction to symmetric mechanisms is without loss of generality, consider any honest and obedient asymmetric mechanism, μ . For any permutation, ρ , of the member indices, define the mechanism μ_ρ by $p_i(c; \mu_\rho) = p_{\rho(i)}(c; \mu)$ and $a_i(c; \mu_\rho) = a_{\rho(i)}(c; \mu)$. Now define the symmetric mechanism, $\bar{\mu}$, by uniformly randomizing among all possible such permutations. Linearity of the member utility function will guarantee that $\bar{\mu}$ is also honest and obedient, and it generates the same total surplus as μ .

$|g| \geq m_n$ and $\chi(g) = 0$ if $|g| < m_n$. Given this, the utility for agent i , with cost type c , when the activated coalition is g can be written as:

$$u_g^i(c) = \begin{cases} v\chi(g) - c & \text{if } g \in I_i \\ v\chi(g) & \text{if } g \notin I_i \end{cases}$$

Using this notation, condition (HO) requires:

$$U(c) = E_{\mathbf{c}_{-i}} \left[\sum_{g \subseteq I} \mu_g(c, \mathbf{c}_{-i}) u_g^i(c) \right] \geq E_{\mathbf{c}_{-i}} \left[\sum_{g \subseteq I} \mu_g(c', \mathbf{c}_{-i}) u_{g_{-i}, \delta_i(g_i)}^i(c) \right] \quad (\text{HO})$$

for any $i = 1, \dots, n$, $c, c' \in [0, 1]$, and any function $\delta_i(g_i)$ mapping g_i to $\{0, 1\}$. If we fix $\delta_i(g_i) = g_i$, (HO) implies the standard interim incentive compatibility condition:

$$U(c) \geq U(c', c) = E_{\mathbf{c}_{-i}} \left[\sum_{g \subseteq I} \mu_g(c', \mathbf{c}_{-i}) u_g^i(c) \right] \quad (\text{IC})$$

for any $c, c' \in [0, 1]$.

If we fix $c_i = c$, (HO) implies the following *interim moral hazard condition* (IMH):

$$E_{\mathbf{c}_{-i}} \left[\sum_{g \subseteq I} \mu_g(c, \mathbf{c}_{-i}) u_g^i(c) \right] \geq \max_{\delta_i} E_{\mathbf{c}_{-i}} \left[\sum_{g \subseteq I} \mu_g(c, \mathbf{c}_{-i}) u_{g_{-i}, \delta_i(g_i)}^i(c) \right] \quad (\text{IMH})$$

This inequality states that members find it optimal to follow the mechanism's recommendation on the equilibrium path in which types are truthfully revealed. Condition (HO) however also rules out joint deviations, in which a member misreports his/her type and then disobeys to the recommendation that follows the misreport.

Condition (IMH) has two implications. First, since the right hand side is non negative and the left hand side is $U(c) = E_{\mathbf{c}_{-i}} [U(c, \mathbf{c}_{-i})]$, IMH *implies* interim individual rationality (INTIR):

$$U(c) = E_{\mathbf{c}_{-i}} [U(c, \mathbf{c}_{-i})] \geq 0 \quad (\text{INTIR})$$

It follows that an (HO) mechanism is also an (IC) and (INTIR) mechanism. Second, (IMH)

implies:

$$c > v \Rightarrow a(c) = 0 \tag{1}$$

since $U((g_{-i}, 0), c) > U((g_{-i}, 1), c)$ for any g_{-i} if $c > v$. Condition (1) is not required in an (IC) and (INTIR) mechanism, so an (IC) and (INTIR) mechanism is not generally an (HO) mechanism (as we will see in Section 3.2).¹³

3 Two Benchmarks

Before characterizing optimal HO mechanisms, it is natural to consider two polar benchmarks. The first is completely unorganized groups, which provides a lower bound on the success of HO mechanisms. While HO mechanisms correspond to the set of all correlated Bayesian equilibria of the game, completely unorganized groups have no means of communication so the equilibrium outcomes correspond to the set of symmetric uncorrelated Bayesian equilibria. Thus, the gains from HO organizations can be measured in terms of the improvement over the best symmetric Bayesian equilibrium for unorganized groups. We show that this lower bound is essentially complete failure, unless the returns to scale are sufficiently high. With constant returns to scale or sufficiently small returns to scale, no member ever participates and the group never succeeds except for groups with very few members.

The second benchmark is strong organizations, which relaxes two of the constraints imposed by HO mechanisms, monetary transfers and obedience, with the latter constraint replaced with interim individual rationality (INTIR). It is a natural benchmark because it corresponds to the standard approach taken in the public good mechanism design literature (Mailath and Postlewaite [1990], Ledyard and Palfrey [1994, 1999], Hellwig [2003]). We show that under fairly weak conditions, the optimal mechanism succeeds 100% of the time.

¹³An alternative assumption for the participation constraint is *ex post individual rationality* (EXIR). It is interesting to note that (EXIR) implies neither (IMH) nor (HO). An example is presented in Section 3.2, where we show conditions under which the optimal (IC) and (INTIR) mechanism is (EXIR), but fails (IMH) and (HO).

3.1 Unorganized groups

For an unorganized group, the payoff function and distribution of costs described above define a Bayesian game where each member simultaneously choose to be active or not. We consider only symmetric equilibria of the game. The symmetry assumption reflects the idea that an asymmetric equilibrium implicitly requires some degree of organization or communication.

3.1.1 Equilibrium with unorganized groups

Denote by p the probability a member is active in the voluntary contribution game. Given any value of $p \in [0, 1]$, each member has a best reply that is characterized by a cutpoint, $\hat{c}_n^U(p)$, with the property that member i is active if and only if $c_i \leq \hat{c}_n^U(p)$. If success requires at least m_n of the n members to be active, then an equilibrium cutpoint must satisfy:

$$\hat{c}_n^U(p) = vB(m_n - 1, n - 1, p), \quad (2)$$

where $B(m_n - 1, n - 1, p) \equiv \binom{n-1}{m_n-1} (p)^{m_n-1} (1-p)^{n-m_n}$ is the probability of being pivotal. In equilibrium, it must be that p coincides with the probability a member has $c_i \leq \hat{c}_n^U(p)$, which is simply equal to $F(\hat{c}_n^U(p))$. Hence the following condition is necessary and sufficient for c_n^U to be an equilibrium cutpoint:

$$c_n^U = vB(m_n - 1, n - 1, F(c_n^U)). \quad (3)$$

An equilibrium exists, trivially, because $c_n^U = 0$ is always a solution to equation (3) when $m_n > 1$. It is possible there are also equilibria with $c_n^U \in (0, v)$. In all the analysis that follows, c_n^U always refers to the largest solution to equation 3: this is without loss of generality since we just intend to find an upper bound to the effectiveness of unorganized groups.

An unorganized group succeeds with positive probability only if there is a strictly positive solution c_n^U . Given a $c_n^U > 0$ and associated $p_n^U > 0$, the equilibrium probability an

unorganized group is successful is:

$$P_n^U(p_n^U, \alpha_n) = \sum_{k=m_n}^n B(k, n, p_n^U) \quad (4)$$

It is relatively straightforward to see that in the case with constant returns to scale, i.e. $m_n = \alpha n$ for some $\alpha \in (0, 1)$, large groups completely fail for sufficiently large n , in the sense that no member is ever active, including members with arbitrarily small costs: formally, there is a finite \bar{n}_U such that $p_n^U = c_n^U = 0$ for all $n > \bar{n}_U$. To see this point, suppose for simplicity that F is uniform in $[0, 1]$, consider any $c_n^U \in (0, v]$ and multiply both sides of equation (3) by $\frac{1-c_n^U}{vc_n^U}$ and substitute $\alpha n = m_n$ to obtain:

$$\frac{1-c_n^U}{v} = \left(\frac{(1-\alpha)n+1}{\alpha n-1} \right) \binom{n-1}{\alpha n-2} (c_n^U)^{\alpha n-2} (1-c_n^U)^{(1-\alpha)n+1} \quad (5)$$

The left hand side of equation (5) is greater than or equal to $\frac{1-v}{v}$ for all n and the right hand side converges to $\left(\frac{1-\alpha}{\alpha}\right) B(\alpha n-2, n-1, c_n^U)$ which converges to 0. Hence there exists \bar{n}_U such that for all $n > \bar{n}_U$ the only solution to equation (3) is $c_n^U = 0$.¹⁴

3.1.2 Unorganized Equilibrium in Large Groups: The effect of returns to scale

While it is natural to assume that m_n increases in n , it is also natural to expect that it grows slower than n . This opens the question of whether and to what extent an unorganized group can achieve success if m_n grows sufficiently slow. The following example shows that at least in the polar extreme case where m_n is constant in n , group success is achieved in the limit. This is a particularly extreme example of increasing returns to scale of activism, in which the ratio of required participation to population, m_n/n declines at the speed of $1/n$.

Example 1. The Volunteer's dilemma: Assume $F(c)$ is uniform in $[0, 1]$ and consider the so called “volunteer's dilemma,” in which only 1 volunteer is required, regardless of group size, so $m_n = 1$.¹⁵ Will the group be able to send 1 volunteer as $n \rightarrow \infty$? It is straightforward

¹⁴This is stated and proved more generally as Theorem 1 of the next section.

¹⁵The volunteer's dilemma is not consistent with our assumption that $m_n > 1$, which is assumed to hold throughout the rest of the paper, but it is an illustrative boundary case. The argument presented here generalizes to the case in which m_n is constant and equal to any integer $M > 1$.

to see that the answer is yes. From equation (3), a cutpoint equilibrium solves: $c_n^U = v(1 - c_n^U)^{n-1}$ which has a unique positive solution c_n^U for all n , with $\lim_{n \rightarrow \infty} c_n^U = 0$. The probability of success, from equation 4 is:

$$P_n^U = 1 - (1 - c_n^U)^n = 1 - \left(\frac{c_n^U}{v}\right)^{\frac{n}{n-1}}$$

so $\lim_{n \rightarrow \infty} P_n^U = 1 - \left(\frac{\lim_{n \rightarrow \infty} c_n^U}{v}\right) = 1$. \blacksquare

Can the logic of the volunteer's dilemma be generalized to the more realistic case in which m_n grows without bound? We say that m_n grows slower (resp., faster) than a sequence s_n , i.e. $m_n \prec s_n$ (resp., $m_n \succ s_n$) if $m_n/s_n \rightarrow 0$ (resp. $m_n/s_n \rightarrow \infty$). We say that m_n grows at the same speed as s_n , i.e. $m_n \simeq s_n$ if $m_n/s_n \rightarrow \rho$ for some finite $\rho > 0$.¹⁶ The following theorem establishes that, independently of the shape of F , there is an equilibrium in which unorganized groups are successful with probability 1 in the limit as $n \rightarrow \infty$ if $m_n \prec n^{2/3}$, and are completely unsuccessful in the limit in all equilibria if $m_n \succ n^{2/3}$.

Theorem 1. *With an unorganized group, for all $v \in (0, 1)$:*

1. *if $m_n \prec n^{2/3}$ then $\lim_{n \rightarrow \infty} P_n^U = 1$.*
2. *if $m_n \succ n^{2/3}$ then there exists \bar{n}_U such that the unique equilibrium is $c_n^U = 0$ for all $n > \bar{n}_U$, and hence $P_n^U = 0$ for all $n > \bar{n}_U$*

Proof: See Appendix.

Theorem 1 shows that we do not need constant returns for a group to fail in large finite groups; when the rate of growth of m_n is sufficiently high, i.e. $m_n \succ n^{2/3}$, the probability of success in the unorganized group collapses to *exactly* zero for all sufficiently large n . The result is stronger than a limiting result: failure always occurs with large finite unorganized groups. Perhaps more significantly, however, the theorem also shows that the negative results on collective action cannot generalize to all cases in which m_n grows slower than n . Even

¹⁶We will also use the notation $m_n \succeq s_n$ (resp. $m_n \preceq s_n$) for the case in which m_n does not grow slower (resp. faster) than a sequence s_n .

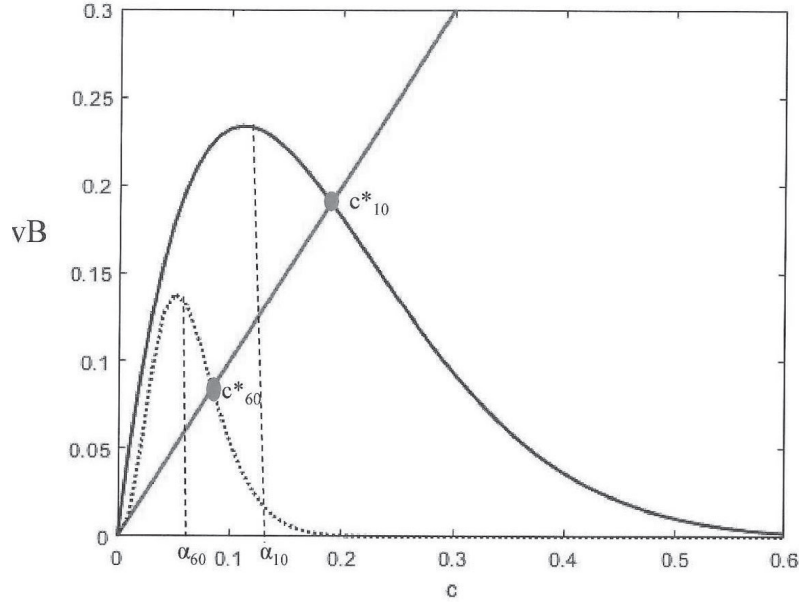


Figure 1: Intuition for equilibria with positive limit probability of success in unorganized groups: $v = .6$, $n = 10, 60$, $m_n = \lceil .5n^{.6} \rceil$.

with no organization the group can achieve a limit success probability of 1; but the no organization case can be seen as a trivial HO mechanism, so full success is also possible in the limit in an optimal HO mechanism when $m_n \prec n^{2/3}$.

Figure 1 illustrates part 1 of Theorem 1, the case where $m_n \prec n^{2/3}$. The diagonal in the figure is the left hand side of equation (3), and the two single-peaked curves show the right hand side of equation (3), for $v = .6$, $m_n = .5n^{.6}$, F uniform, and two different group sizes, $n = 10, 60$. An unorganized group equilibrium is any value of c where the single peaked curve intersects the diagonal. An increase in n has two effects on the right hand side of equation (3). First, it pushes the peak down, since the probability of exactly $m_n - 1$ active agents goes down: this makes it harder to have a positive intersection. Second, since $m_n \prec n$, it shifts the peak of the curve to the left since the share of required active agents $\alpha_n = m_n/n$ is also reduced. As n increases, the probability of success remains bounded above zero and eventually converges to 1 as long the sequence of intersections c_n remains sufficiently higher than the sequence of thresholds α_n (i.e., $F(c_n) > \alpha_n$). From Figure 1 we can see that a necessary condition for this is that for all n sufficiently large the right hand side of equation

(3), evaluated at α_n , is higher than the 45° degree line, so higher than $F^{-1}(\alpha_n)$,¹⁷ i.e.:

$$\frac{vB(m_n - 1, n - 1, \alpha_n)}{F^{-1}(\alpha_n)} \geq 1. \quad (6)$$

When this is the case, then the highest intersection point, c_n , remains on the right of the threshold α_n . The proof of Theorem 1 uses Stirling's approximation formula to show that, for any choice of F , a necessary and essentially sufficient condition for this to happen is that m_n increases at a rate slower than $n^{2/3}$. In that case the left hand side of (6) diverges to infinity; when m_n increases at a rate faster than $n^{2/3}$, the left hand side converges to zero so the probability of success also converges to zero, violating (6). Referring back to Figure 1, $m_n \prec n^{2/3}$ ensures that the second effect of increasing n (shifting the peak to the left) dominates the first effect (shifting the peak down).

The logic behind the “magic number” $2/3$ in Theorem 1 can be heuristically explained as follows. Let us first see why, when $m_n \succ n^{2/3}$, the expected share of volunteers in equilibrium $F(c_n^U)$ falls short of the threshold α_n for all n sufficiently large, thus leading to a limit probability of success equal to zero. As $n \rightarrow \infty$, $c_n^U \rightarrow 0$, so $F(c_n^U) \simeq f(0)c_n^U$. We therefore have $\alpha_n \leq F(c_n^U)$ only if:

$$\frac{m_n}{n} = \alpha_n \leq F(c_n^U) \simeq f(0)c_n^U = vf(0)B(m_n - 1, n - 1, F(c_n^U)) \quad (7)$$

where the last equality follows from the equilibrium condition for c_n^U . Since $B(m_n - 1, n - 1, F(c_n^U))$ is maximized at the value of c such that $F(c) = (\alpha_n - 1/n) / (1 - 1/n) \simeq \alpha_n$, in the limit $B(m_n - 1, n - 1, F(c_n^U)) \leq B(m_n - 1, n - 1, \alpha_n)$, so (7) is implied by $vf(0) \cdot B(\alpha_n n - 1, n - 1, \alpha_n) / \alpha_n \geq 1$. The binomial probability of $\alpha_n n - 1$ successes converges to zero at the slowest rate when the probability of success is α_n , and this rate is on the order of $1/\sqrt{\alpha_n n}$. This implies that:

$$B(\alpha_n n - 1, n - 1, F(c_n^U)) \leq B(\alpha_n n - 1, n - 1, \alpha_n) \simeq m_n^{-1/2},$$

¹⁷The function of c_n $F(vB(\alpha_n n - 1, n - 1, c_n))$ has a maximum at $c_n = (\alpha_n n - 1) / (n - 1) < \alpha_n$, and it is increasing (resp. decreasing) in c for $c < (\alpha_n n - 1) / (n - 1)$ (resp., $c > (\alpha_n n - 1) / (n - 1)$).

where \preceq here means that the right hand side converges to zero at the same or slower rate than the left hand side. A necessary condition for (7) therefore is that $\frac{m_n}{n}$ converges to zero faster than $m_n^{-1/2}$, but this condition cannot hold if $m_n \succ n^{2/3}$. When $m_n \prec n^{2/3}$, we have $\alpha_n < F(vB(m_n - 1, n - 1, \alpha_n))$ for all n sufficiently large. Since $F(vB(m_n - 1, n - 1, 1)) = 0 < 1$, continuity implies that there is a solution $c_n^U > \alpha_n$ for all n sufficiently large. Indeed, the proof of Theorem 1 establishes that this solution remains sufficiently larger than α_n , so that the probability of success converges to 1. When instead $m_n \succ n^{2/3}$, then $\alpha_n > F(vB(m_n - 1, n - 1, \alpha_n))$ for all n sufficiently large. Again, the proof of Theorem 1 establishes that the only solution of $c_n^U = vB(m_n - 1, n - 1, F(c_n^U))$ is actually zero in this case for n sufficiently large, so that the probability of success is 0 for n sufficiently large.

3.2 Strong Organizations

The standard Bayesian mechanism design approach to study collective action and public good provision is to characterize the optimal direct mechanism allowing for monetary transfers, requiring incentive compatibility (IC) and interim individually rationality (INTIR), see Rob [1989], Mailath and Postlewaite [1990] and Ledyard and Palfrey [1994, 1999]. We refer to a mechanism with transfers requiring (IC) and (INTIR) as a strong organization, since in both cases it needs to satisfy weaker constraints than in the weak organization defined in 2.2.1, and thus they can achieve more.

The best (IC) and (IR) mechanism can be characterized as the solution of the following maximization problem, where the interim expected monetary payment of type c is denoted by $t(c) = E_{\mathbf{c}_{-i}}[t^i(c, \mathbf{c}_{-i})]$:

$$\begin{aligned}
& \max_{p,a} \int_0^1 U(c) dF(c) & (8) \\
& \text{s.t. } vp(c) - ca(c) - t(c) \geq vp(c') - ca(c') - t(c') \quad \forall c, c' \in [0, 1] \\
& \quad vp(c) - ca(c) - t(c) \geq 0 \quad \forall c \in [0, 1] \\
& \quad p, a, t \text{ feasible}
\end{aligned}$$

where the first constraint is the (IC) constraint, the second is the (INTIR) constraint and the third is the feasibility constraint discussed in Section 2 with the additional feasibility condition that monetary transfers balance: $\sum_{i=1}^n t^i(\mathbf{c}) = 0$ for all \mathbf{c} . Following standard methods, the (IC) constraint is equivalent to requiring $U'(c) = -a(c)$ and $a(c)$ is non-increasing. Substituting (IC) into the objective function and simplifying gives:

$$\begin{aligned} & \max_{p(0), a(\cdot)} \left\{ vp(0) - \int_0^1 a(c) \frac{1 - F(c)}{f(c)} dc \right\} \\ & \text{s.t. } U'(c) = -a(c), a(c) \in [0, 1] \text{ and non-increasing,} \\ & U(c) \geq 0 \forall c \in [0, 1], \text{ and } p, a, t \text{ feasible} \end{aligned} \tag{9}$$

To solve (9), consider a relaxed version in which we ignore the (INTIR) constraint. In the appendix, we prove that when $F(c)$ satisfies the Monotone Hazard Rate Assumption (MHRA) the optimal way to solve this relaxed problem is to keep $a(c)$ flat. Intuitively, when $F(c)$ satisfies MHRA, then in the objective function $a(c)$ is weighted by an increasing function, $-\left[\frac{1-F(c)}{f(c)}\right]$. In this case, if $a(c)$ is strictly decreasing, then it is optimal to shift the probability of activation $a(c)$ from lower to higher values of c . Since $a(c)$ must be non-increasing, the best way to do it satisfying feasibility is to keep $a(c)$ and (by incentive compatibility) $p(c)$ constant: $a_n(c) = a_n$ and $p_n(c) = p_n$. In the absence of INTIR, it is efficient for the group to always be successful as long as $\frac{m_n}{n} \cdot E(c) < v$, we have $p_n = 1$ and a_n is chosen to be the smallest possible, so $a_n = \frac{m_n}{n}$ for all c . The solution of the relaxed problem is implemented (without any need to report types) by simply selecting exactly m_n agents at random to be active. We call this the *random mechanism*. This mechanism is also a solution to the full problem with INTIR (9) as long as $\frac{m_n}{n} \cdot 1 < v$ which guarantees that INTIR is not violated for the highest cost type, $c = 1$. With constant returns to scale this holds for all n if $\alpha < v$. With increasing returns, it holds as long as n is sufficiently large ($n \geq n^* = \min\{n | \frac{m_n}{n} < v\}$).

Theorem 2. For all $v \in (0, 1)$:

1. If $m_n \prec n$, then there exists a critical group size, n^* , such that for $n > n^*$ the random

mechanism satisfies (IC) and (INTIR) and achieves a probability of success equal to 1. If F satisfies (MHRA), the random mechanism is optimal.

2. *if $m_n = \alpha n$ and $\alpha < v$, then, for all n , the random mechanism satisfies (IC) and (INTIR) and achieves a probability of success equal to 1. If F satisfies (MHRA), the random mechanism is optimal.*

Proof: See Appendix.

Theorem 2 is relevant for two reasons. First, because it shows that the optimal (IC) and (IR) mechanism generally violates obedience. This can be seen from the fact that it requires all types to be active with positive probability, but this directly violates (1) since no type with $c > v$ would find it optimal to be active. The mechanism satisfies (INTIR) since the probability of being activated is small, so $v(1 - \alpha_n \frac{c}{v}) > 0$ even if $c > v$; but this only guarantees *interim* participation in the mechanism, not that a type $c > v$ will obey a recommendation to be active. Second, Theorem 2 is relevant because it highlights the need to study more realistic, honest and obedient mechanisms: by ignoring moral hazard, the optimal mechanism achieves complete success as $n \rightarrow \infty$. In order to understand why collective action can only be partially successful in more realistic environments, we need to integrate the obedience constraint into the analysis of the optimal mechanism.

Note that the monotone hazard rate assumption in Theorem 2 is used only for the characterization of the shape of the optimal mechanism, not for the substantive result that there is an (IC) and (INTIR) mechanism that achieves success with probability one for n large when $m_n \prec n$.

Note also that Theorem 2 is not in conflict with the main result in Mailath and Postlewaite [1990] where it was shown that the probability of success converges to zero in the best (IC) and (INTIR) mechanisms (even allowing for monetary transfers).¹⁸ That earlier result relied on an assumption that the total benefit of success, nv , is strictly lower than the cost of

¹⁸The random mechanism that succeeds with probability one does not use transfers. Hence, the gain from strong mechanisms is entirely due to the violation of obedience. The considerable group benefits from a strong organization are achieved via ex post coercion - by sacrificing the sovereignty of individual high-cost members of the group.

obtaining it in the worst case scenario where all members have cost equal to 1.¹⁹ In our setting, this assumption reduces to $v < m_n/n = \alpha_n$, which is not satisfied for large n when $m_n \prec n$, nor for the case of constant returns if $v \in (\alpha, 1)$. Requiring $v < \alpha_n$ for all n is essentially the same as assuming constant returns to scale with $\alpha = \liminf_{n \rightarrow \infty} \alpha_n$. Indeed, there are many situations in which it is natural to assume $v > \alpha_n$, such as situations where the “sacrifice” of a small share of population is all that is needed to guarantee group success. As we will see in the next two sections, with honest and obedient mechanisms, $v > \alpha_n$ does *not* imply success even though strong organizations with sufficiently many members will succeed with probability 1. The failure or success of collective action in HO mechanisms with large n depend only on the returns to scale and is the same for all $v \in (0, 1)$.

The observation that when we relax that assumption then the collective good can be financed in an (IC) and (INTIR) mechanism with monetary transfers is not completely new, as it was previously made by Hellwig [2003] in a more general environment in which the public good can be chosen as a continuous variable. Theorem 2 differs from Hellwig’s result in two ways: it shows that unlimited monetary transfers are not necessary, and it provides a full characterization of the optimal mechanism, even for large but finite n , as a simple random mechanism.

4 The Volunteer Based Organization (VBO)

As discussed in the previous section, strong mechanisms that only require (IC) and (INTIR) are not a good description of organizations for collective action because they ignore the obedience constraint, implicitly allowing the mechanism to coerce high cost members to be activated, or alternatively for all members to commit *ex ante* to obey any recommendation. An optimal strong organization, moreover, cannot explain collective action as an empirical phenomenon, since it either predicts complete success of any group when $m_n \prec n$ or if $m_n = \alpha n$ and $\alpha < v$, or complete failure otherwise. The result for the other benchmark, unorganized groups, is as negative as the results from strong organizations are positive.

¹⁹Formally, the key assumption in Mailath and Postlewaite [1990] is assumption (iv) in their Theorem 2. Translating it to our notation, it is equivalent to requiring that there exists $\epsilon > 0$ such that $nv + n\epsilon < n\alpha_n$.

Unorganized groups suffer *complete failure*, even in relatively small groups, except for the case where returns to scale are very high.

The key question then is: Where does the performance of optimal HO mechanisms fall in the very wide range between these two benchmarks? What type of HO organization is optimal and what type should one expect to see in practice? In this section we identify a simple and natural communication mechanism called a *Volunteer Based Organization* (VBO). In a VBO, which is obviously honest and obedient, members self-identify as either *volunteers* (low-cost) or *free riders* (high-cost) and the mechanism coordinate the activity of volunteers in order to activate a minimal coalition for success if there are enough self-reported volunteers. It turns out that the VBO is an approximately optimal HO mechanism for large n , as we show in this section.

4.1 Unique Honest and Obedient VBO

In a VBO each member reports his or her type: if the reported type is higher than some threshold $c_n^O > 0$, the agent is excused and not asked to be active, irrespective of what the other members report; if the type is less than or equal to c_n^O , then the agent is deemed a *volunteer* and is activated with positive probability, determined by the following rule. If the number of volunteers is greater than m_n , then a coalition of exactly m_n volunteers is randomly selected and activated, resulting in group success. If the number of volunteers is fewer than m_n , then no volunteer is activated and the group is unsuccessful. In case the group activates m_n volunteers, then all volunteers have the same probability of being included. Using the notation introduced in Section 2.2 a VBO is defined formally as follows:²⁰

Definition 1. For any $c \in [0, 1]$ and any profile of types, \mathbf{c} , let $k(\mathbf{c}; c) = |\{j \in I | c_j \leq c\}|$. For any given m_n and n , a simple VBO mechanism is defined by a volunteer cutoff $c_n^O \in (0, v)$ such that (1) $A_i(\mathbf{c}) = 0$ for all \mathbf{c} and for all i such that $c_i > c_n^O$; (2) $k(\mathbf{c}; c_n^O) < m_n \Rightarrow P(\mathbf{c}) = 0$ and $A_i(\mathbf{c}) = 0$ for all i ; (3) $k(\mathbf{c}; c_n^O) \geq m_n \Rightarrow P(\mathbf{c}) = 1$ and $A_i(\mathbf{c}) = \frac{m_n}{k(\mathbf{c}; c_n^O)}$ for all i such that $c_i \leq c_n^O$.

²⁰The definition requires $c_n^O > 0$. If c_n^O were exactly equal to 0, then IC implies no restriction on cost reports: all types are indifferent between all cost reports, including reporting $c' = 0$. Choosing c_n^O is obviously never optimal.

We first establish some properties of the VBO mechanism. The following result characterizes the unique incentive compatible *VBO*. Define the function:

$$Y_n(c) = \frac{vB(m_n - 1, n - 1, F(c))}{\sum_{k=m_n-1}^{n-1} \frac{m_n}{k+1} B(k, n - 1, F(c))} \quad (10)$$

Proposition 1. *For any v , m_n and n :*

1. *The function Y_n is strictly decreasing in c and has a unique fixed point $c_n^O \in (c_n^U, v)$.*
2. *A VBO is Honest and Obedient if and only if it has the volunteer cutoff c_n^O that satisfies $c_n^O = Y_n(c_n^O)$.*

Proof: See Appendix.

Condition (10) provides a simple way to compute the equilibrium threshold c_n^O and characterize its qualitative properties. Proposition 1 also makes clear why it is natural to refer to such a mechanism as “volunteer based”. The honest and obedient VBO can be implemented as a simple modification of the participation game with an unorganized group. In this implementation, each member is asked to choose to be either a “volunteer” or a “free rider”. The group is successful if the number of volunteers is greater than or equal to m_n , in which case exactly m_n of them are randomly selected to be active. If the number of volunteers is less than m_n , then no member is activated.

Volunteers are always willing to follow a recommendation to be active, since they have types $c_i \leq c_n^O < v$ and they know that, conditional on receiving such a recommendation, the mechanism has also activated exactly $m_n - 1$ other volunteers so they are pivotal. At the interim stage, however, an agent might still have an incentive to misreport, since it would prefer some other agent to be called in case of $k \geq m_n + 1$. The cost of misreporting as a free rider is that indeed that there are exactly $m_n - 1$ volunteers in the rest of the group, so the misreport would be pivotal in inducing the group’s failure. The expected value of this cost is $vB(m_n - 1, n - 1, F(c_n^O))$, pretty much the same as in the unorganized group in (3). The critical difference for the organized group is that volunteers are not called to action indiscriminately, but only if they are needed, and never in excess. These qualifications are

reflected in the denominator of (10), which is, instead, simply equal to 1 for unorganized groups. Hence $c_n^O > c_n^U$, and the probability that a member volunteers is $F(c_n^O) > F(c_n^U)$. Therefore, there are *three* improvements of the VBO relative to an unorganized group: (1) nobody contributes if the threshold is not reached; (2) no more than m_n members contribute; and (3) the threshold is reached more often because $F(c_n^O) > F(c_n^U)$.

4.2 The performance of VBO in large groups: The effect of returns to scale

Condition (10) establishes that for any finite n , the VBO outperforms an unorganized group by a combination of eliminating waste and increasing the equilibrium probability of group success. We might expect that the presence of an organization that allows for coordination, increases the probability of success, and eliminates wasteful participation makes it possible to achieve higher limit probabilities of success in large group, at least for some parametrizations. However, as we show below this conjecture is incorrect: the limiting performance of the VBO *is the same* as the limiting performance of unorganized groups.

To evaluate how a group performs using the honest and obedient VBO when n is large requires evaluating how the solution to condition (10), c_n^O , converges as $n \rightarrow \infty$. The next result has important implications for the probability of success of the organized group and the welfare of its members.

Proposition 2. *With an organized group using the honest and obedient VBO, then for all $v < 1$ we have that:*

- *If $m_n = \alpha n$ then $c_\infty^O \equiv \lim_{n \rightarrow \infty} c_n^O > 0$.*
- *If $m_n \prec n$ then $c_\infty^O = 0$ and $\lim_{n \rightarrow \infty} F(c_n^O) / \alpha_n \geq 1$.*

Proof: See Online Appendix.

The first bullet point establishes that the share of volunteers is always strictly positive in a VBO with constant returns: this implies that the probability of success in a VBO, P_n^O , is positive with constant returns for any finite n . In contrast, $p^U = F(0) = 0$ for

large enough finite unorganized groups, so the relative success of the VBO compared to unorganized groups is *infinite* for large enough finite groups. The intuition for the fact that $\lim_{n \rightarrow \infty} c_n^O > 0$ is that the numerator and the denominator of the ratio on the right hand side of (10) both converge to zero at the same rate and indeed the ratio is strictly positive in the limit. The following result highlights this property, but qualifies it showing that this is not enough to guarantee positive probability of success for the VBO in the limit:

Corollary 1. *With an organized group using the honest and obedient VBO and constant returns (i.e. $m_n = \alpha n$), then there exists $n_U(\alpha, v)$ such that for all $n > n_U(\alpha, v)$, $P_n^U / P_n^O = 0$. Despite this, we have that $P_\infty^O \equiv \lim_{n \rightarrow \infty} P_n^O = 0$.*

Proof: See Online Appendix.

The second point of Proposition 2 is that, when $m_n \prec n$, then, while the equilibrium participation rate converges to zero as $n \rightarrow \infty$, it does so at the same (or slower) rate as the threshold fraction α_n , since $\lim_{n \rightarrow \infty} (F(c_n^O) / \alpha_n) \geq 1$. If $\lim_{n \rightarrow \infty} (F(c_n^O) / \alpha_n) > 1$, then obviously $P_\infty^O = 1$, but if $\lim_{n \rightarrow \infty} (F(c_n^O) / \alpha_n) = 1$, then it depends on exactly how $F(c_n^O) / \alpha_n$ converges to 1. If $F(c_n^O) / \alpha_n$ converges from below and convergence is slow, then the probability that the number of volunteers passes the threshold converges to zero; if convergence is fast or $F(c_n^O) / \alpha_n$ converges from above, then the probability of success of the group will be strictly positive even for an arbitrary large number of activists. Does the fact that c_n^O remains bounded or converges to zero at the same speed of α_n imply that if $m_n \prec n$ we can get strictly positive probability of success even with large or arbitrarily large groups, and/or we can achieve a higher limit probability of success than without an organization? The answer is no: the conditions for the limiting probability of success for a VBO in large groups are exactly the same as the result for unorganized groups.

Theorem 3. *With an organized group using the honest and obedient VBO, then for all $v \in (0, 1)$:*

- *If $m_n \prec n^{2/3}$, then $\lim_{n \rightarrow \infty} P_n^O = 1$, so $\lim_{n \rightarrow \infty} \frac{P_n^U}{P_n^O} = 1$.*
- *If $m_n \succ n^{2/3}$, then $\lim_{n \rightarrow \infty} P_n^O = 0$, but there is a n_U such that for all $n > n_U$, $\frac{P_n^U}{P_n^O} = 0$.*

Proof: See Appendix.

As we will show in the next section, Theorem 3 holds generally for *all* HO mechanisms. Surprisingly, the limit probability of success is the same with a VBO or in an unorganized group. When $m_n \prec n^{2/3}$, a limit probability of success is 1, but this was also true for an unorganized group; when $m_n \succ n^{2/3}$, the limit probability of success is zero with a VBO, once again, just as in an unorganized group. With a VBO, however, the probability of success is positive for any n even when $m_n \succ n^{2/3}$, a feature that is not shared by the equilibrium in an unorganized group (where the probability is exactly zero after a finite n), so for sufficiently high n , as highlighted by the second bullet, $\frac{P_n^U}{P_n^O} = 0$. This can imply a significant benefit of adopting a simple VBO compared to having an unorganized group. As we will show in the next section where we quantify by numerical methods the success probability in a VBO, for reasonable parameter values groups with VBO can achieve high probabilities of success even for large groups, even if $m_n \succ n^{2/3}$.

It is useful to go over the intuition of Theorem 3 since, suitably generalized, it will also help with understanding the extension to general HO mechanisms in the next section. The expected share of volunteers does not fall short of the threshold for success only if $\alpha_n \leq F(c_n^O)$. Since $c_n^O \rightarrow 0$, we have $F(c_n^O) \simeq f(0)c_n^O$, implying:

$$\frac{\alpha_n}{f(0)} \lesssim c_n^O = Y(c_n^O) \leq Y(F^{-1}(\alpha_n)) = \frac{vB(\alpha_n n - 1, n - 1, \alpha_n)}{a_n(\alpha_n)}$$

where $a_n(\alpha_n) = \sum_{k=\alpha_n n - 1}^{n-1} \frac{\alpha_n n}{k+1} B(k, n - 1, \alpha_n)$ and \lesssim denotes that the left-hand side converges to a value less than or equal to the right-hand side. The second inequality, $Y(c_n^O) \leq Y(F^{-1}(\alpha_n))$ follows from the fact that, $Y(\cdot)$ is a decreasing function, by Proposition 1. Hence:

$$a_n(\alpha_n) \lesssim v f(0) \cdot \frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} \quad (11)$$

If $m_n \succ n^{2/3}$, we know from the proof of Theorem 1 that the right hand side of (11) converges to zero. Condition (11), therefore, implies that $a_n^O(\alpha_n)$ must also converge to zero. It is however intuitive to see that this is impossible. Note that $a_n(\alpha_n)$ is the probability that a volunteer is activated when the threshold used by the other members for volunteering is

α_n . But when this is the case, for large n there will be a share of volunteers roughly equal to α_n . In this case, the probability that a volunteer is activated cannot be arbitrarily small since, even conditioning on having at least a share α_n of volunteers, the share of volunteers will almost surely be only marginally greater than α_n , the minimal requirement for success.

4.3 The value of a VBO with finite n

While the limit probability of success is the same for a VBO and an unorganized group, the VBO mechanism has *three* improvements of the VBO relative to an unorganized group: (1) nobody contributes if the threshold is not reached; (2) no more than m_n members contribute; and (3) the threshold is reached more often because $c_n^O > c_n^U$. These properties have an immediate positive impact on the willingness to volunteer, which leads to a higher probability of success of the group. On the other hand, $c_n^U = 0$ for finite sized unorganized groups. As a result a VBO performs infinitely better than an unorganized group, as shown earlier.

The following result confirms these results, and shows that limit results understate the potential value of a VBO, because when $m_n \succ n^{2/3}$ the the limit probability with an organized group converges to zero slowly. The next result bounds below this rate of convergence. We say that P_n^* converges at a strictly slower rate than exponential if $P_n^*/e^{-\gamma n} \rightarrow \infty$ for any $\gamma > 0$. We have:

Proposition 3. *For any $m_n \succ n^{2/3}$, P_n^O converges to zero at a rate that is strictly slower than exponential.*

Proof: See Appendix.

The example illustrates how this slow rate of convergence implies that a VBO can lead to group success with high probability, even for medium and large sized groups, while unorganized groups are unsuccessful.

Example 2: Comparison of VBO and Unorganized groups $v = 0.8$, $m_n = 0.2n^\beta$, $F \sim \mathcal{U}[0, 1]$.

In this example, we compare the equilibrium in the VBO mechanism and the equilibrium for the unorganized group for different group sizes and different rates of increase in m_n . From Proposition 1, the organized group's optimal threshold satisfies: $c_n^O = Y_n(c_n^O, \alpha_n, v)$ where we make explicit the dependence of Y_n on α_n, v for convenience:

$$Y_n(c_n^O, \alpha_n, v) = v \frac{B(m_n - 1, n - 1, c_n^O)}{\sum_{k=m_n-1}^{n-1} \frac{m_n}{k+1} B(k, n - 1, c_n^O)}. \quad (12)$$

With no organization the equilibrium condition is: $c_n^U = Z_n(c_n^U, \alpha_n, v)$, where:

$$Z_n(c_n^U, \alpha_n, v) = v B(m_n - 1, n - 1, c_n^U) \quad (13)$$

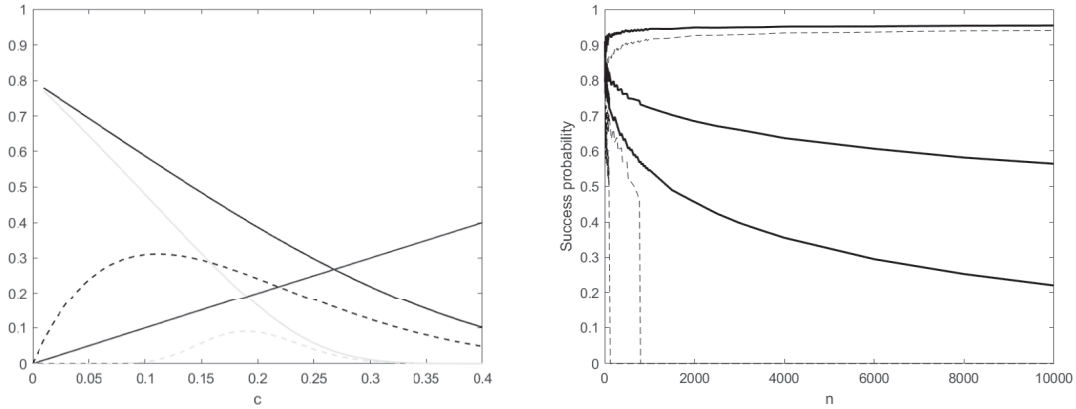


Figure 2: Comparison of VBO mechanism (solid curves) and unorganized group equilibrium (dashed curves). $v = 0.8$, $m_n = .2n^\beta$, $F(c)$ Uniform. Left Panel: Equilibrium cutpoints. $\beta = 1$. Equilibrium cutpoints are located at the intersection of each curve ($n = 10$ (black) and $n = 80$ (gray)) with the diagonal. Right Panel: Probability of success. $n = 10, \dots, 10000$, $\beta = .65$ (top curve), $.80$ (middle curve), $.85$ (bottom curve).

The equilibrium probabilities of success for the unorganized and organized groups are, respectively:

$$P_n^U(c_n^U, \alpha_n) = \sum_{k=m_n}^n B(k, n, c_n^U) \text{ and } P_n^O(c_n^O, \alpha_n) = \sum_{k=m_n}^n B(k, n, c_n^O)$$

The left hand panel of Figure 2 illustrates the $c_n^O = Y_n(c_n^O, \alpha_n, v)$ and $c_n^U = Z_n(c_n^U, \alpha_n, v)$ equilibrium conditions for groups with 10 and 80 members, with $v = 0.8$, $m_n = 0.2n$, and

$F \sim \mathcal{U}[0, 1]$. The downward sloping solid black curve is $Y_{10}(c, 0.2, 0.8)$ and the downward sloping solid gray curve is $Y_{80}(c, 0.2, 0.8)$, corresponding to the right-hand side of equation 12 for the VBO. The respective equilibrium cutpoints, c_{10}^O and c_{80}^O , are given by the intersection of these two $Y(\cdot)$ curves with the diagonal (black). The black dashed single-peaked curve is $Z_{10}(c, 0.2, 0.8)$, and c_{10}^U is given by the highest intersection of this curve with the diagonal. The gray dashed single-peaked curve is $Z_{80}(c, 0.2, 0.8)$, which does not intersect the diagonal at any positive value, so the unique equilibrium for the unorganized group is $c_{80}^U = 0$, with zero participation.²¹

The right hand panel of Figure 2 illustrates the equilibrium probability of success for organized (solid curves) and unorganized groups (dashed curves) for group sizes up to 10,000, with $v = 0.8$, $F \sim \mathcal{U}[0, 1]$, and $m_n = .2n^\beta$ for and three values of $\beta < 1$. One can see that for $\beta = .65 < 2/3$, both unorganized and organized group can achieve success with probability that is high even for small groups and converges to 1. For $\beta > 2/3$, instead the probability of success converges to zero, and groups without an organization fail completely after reaching a threshold size that is decreasing in β . In contrast, organized groups obtain much higher probability of success and this success declines slowly in group size. For $\beta = 0.85$, while the probability of success in a VBO eventually converges toward zero as $n \rightarrow \infty$, the probability of success is more than 20% even for groups with more than 10,000 members. In contrast, for unorganized groups and $\beta = 0.85$, the probability of success is *exactly zero* for all $n \geq 200$. This illustrates the importance of considering finite n even when the limit probability of success is 0.

5 Asymptotic optimality of the VBO

The previous section illustrated a group of large finite size can achieve significant (albeit imperfect) levels of success even when n is large by organizing with a very simple honest and obedient mechanism. Theorem 3 established that the advantage of the VBO is not in the limit probability of success that can be achieved with infinitely large groups, which is equal

²¹Since $m > 1$, there is always a solution at $c^{U*} = 0$. When n is sufficiently small, there are can also be positive equilibrium cutpoints for the unorganized group.

to the limit probability obtainable without an organization. However, this leaves open the possibility that there is an even better HO mechanism that achieves higher probability of success than in an unorganized group, even in the limit, and a positive limiting probability even when $m_n \succ n^{2/3}$.

In this section we prove two key results that demonstrate that the limiting result in Theorem 3 applies not only to VBO but to all HO mechanisms. First, we formalize the optimization problem for HO mechanisms. Second, we show that, exactly like the VBO, the optimal HO has a limiting probability of success equal to 0 if $m_n \succ n^{2/3}$, and equal to 1 if $m_n \prec n^{2/3}$. Thus Theorem 3 holds for *any HO mechanism* (including the unorganized group, the VBO, and the optimal HO mechanism).

5.1 Optimal HO Mechanisms

Formally, the optimal HO mechanism is defined as the solution $a^*(c)$, $p^*(c)$ of the following problem:²²

$$\begin{aligned} \max_{a(c), p(c)} \int_0^1 U(c) dc & \tag{14} \\ \text{s.t. (HO) and } a(c), p(c) \text{ is feasible.} \end{aligned}$$

where (HO) is the honest and obedient constraint specified in Section 2.2. As discussed in Section 2, the (HO) constraint implies (IC), (INTIR) and Condition (1). In the following we will first study the solution of a relaxed problem in which only (IC) and (1) are considered, we will then prove that this solution satisfies the omitted constraints and thus solves (14).

²²Whenever it does not create confusion, as here and when we take n as given, we omit the subscript n in the equilibrium variables $a_n^*(c)$, $p_n^*(c)$, c_n^* .

By standard methods one can rewrite the problem as:

$$\begin{aligned}
& \max_{p(0) \in [0,1], a(c) \in [0,1]} \left\{ vp(0) - \int_0^1 a(c) \frac{1 - F(c)}{f(c)} dc \right\} \\
& \text{s.t. } U'(c) = -a(c) \text{ with } a(c) \text{ non-increasing} \\
& a(c) = 0 \text{ for } c > c^*, \text{ where } c^* = \min \{c \leq v \mid vp(c) - ca(c) \leq vp_2\} \\
& \text{and } p, a \text{ feasible}
\end{aligned} \tag{15}$$

where p_2 is the (constant) expected probability of success for all types $c > c^*$. In (15) we derived the objective function using the (IC) constraint in a similar way as in (9). The constraints in the second line of (15) is a monotonicity constraint implied by IC, also present in (9); the constraint in the third line follows from (1) and incentive compatibility, and it is new to (15).²³ Note that in the problem above we have no (IR) constraint; IR, however, follows from the monotonicity of $U(c)$ and the definition of c^* .²⁴

The HO optimization problem in (15) appears to be very similar to the problem in (9) for strong organizations. In fact, the objective functions are identical, the only difference being the new c^* constraint, which replaces INTIR. We showed earlier that the optimal strong mechanism for any $m_n \prec n$ can be easily characterized when n is large by observing that one can improve the objective function in (9) without violating the constraints by “flattening” the mechanism, i.e. by making the mechanism less sensitive to an agent’s type c . If n is sufficiently large, then the optimal mechanism flattens the mechanism completely, the group is always successful, and all types are asked to be activated with positive probability and randomly selected with the same probability: $p(c) = 1$ and $a(c) = \alpha_n$. IC is trivially satisfied, and INTIR is satisfied for n large since the probability of being activated, α_n , converges to zero so is eventually smaller than any $v > 0$.

The same logic cannot be applied to (15). Here too, the objective function (which is the same in (9) and (15)) improves if we “flatten” the mechanism; however, now flattening the mechanism may affect the obedience constraint. To see this note that a mechanism in which all types are activated with positive probability is certainly impossible, since no type with

²³See the proof of Proposition 5 for the details on how it follows from (1) and incentive compatibility.

²⁴Indeed, $U(c) \geq U(c^*) \geq vp_2 \geq 0$ for all $c \in [0, c^*]$.

$c > v$ will ever accept to be activated even if it is interim individually rational (INTIR) to commit to participate in the mechanism. In the optimal HO mechanism there will necessarily be a maximal type $c_n^* < v < 1$, who is indifferent between volunteering and free riding. By flattening the mechanism, we now necessarily require higher expected participation from this type c_n^* which would break that indifference. Having a flatter mechanism therefore involves a trade off: on the one hand, for a given c_n^* , it improves the objective function since it relaxes the (IC) constraint; on the other hand, however, it may imply lower participation, in the form of a lower c_n^* . Hence the shape of the mechanism could depend on the trade-off between the benefit of keeping the volunteer cutoff high (a higher c_n^*), which produces a larger pool of volunteers, and keeping the mechanism flat, which relaxes incentive compatibility.

5.2 Optimal HO Mechanisms for large groups

In this section, we show that the performance of an optimal HO mechanism is no better than the performance of the VBO in the limit as $n \rightarrow \infty$. Thus, *in the limit*, with infinitely-sized groups, voluntary organizations for collective action accomplish nothing relative to unorganized groups.

The result is established in three steps. First, define a mechanism as *binary*, if it allows the agents to send at most two messages (volunteer or not volunteer); and define a mechanism as *binary HO*, if it is honest and obedient (HO) and binary. We show that the optimal *binary HO mechanism* is a straightforward generalization of the VBO. Second, we show that the optimal binary HO mechanism has the same limiting performance as the optimal HO mechanism, as $n \rightarrow \infty$. Third, we show that the optimal binary HO mechanism has a limiting probability of success equal to 0 if $m_n \succ n^{2/3}$, and equal to 1 if $m_n \prec n^{2/3}$.

5.2.1 Optimal *Binary* HO Mechanisms: The Generalized VBO

For the first step, we introduce a class of mechanisms that generalizes the VBO. A *generalized VBO* is defined by a threshold $k_n^G \geq m_n$ and a volunteer cutoff c_n^G . A k_n^G -generalized VBO with a threshold k_n^G greater than or equal to m_n works as follows. If there are more than k_n^G volunteers, then the mechanism selects and activates exactly m_n volunteers, each with

equal probability, thus guaranteeing that the group succeeds. If there are fewer than k_n^G volunteers, then the mechanism selects 0 volunteers and the group fails. If there are exactly k_n^G volunteers, then the mechanism selects and activates exactly m_n volunteers with some probability $q_{k_n^G} \in (0, 1]$ and selects exactly 0 volunteers with probability $1 - q_{k_n^G}$. Thus, the simple VBO analyzed above corresponds to the special case $k_n^G = m_n$ and $q_{k_n^G} = 1$. The random mechanism that in Section 3.2 we showed solves problem (8) for the best IC and IR mechanism when n is large can also be seen as a VBO with $k_n^G = m_n$, $q_{k_n^G} = 1$ and $c_n^G = 1$.

Definition 2. For any $c \in [0, 1]$ and any profile of types, \mathbf{c} , let $k(\mathbf{c}; c) = |\{j \in I | c_j \leq c\}|$. For any given m_n and n , a generalized VBO mechanism is defined by a volunteer cutoff $c_n^G \in (0, v)$ and a critical mass threshold $k_n^G \geq m_n$, such that (1) $A_i(\mathbf{c}) = 0$ for all \mathbf{c} and for all i such that $c_i > c_n^G$; (2) $k(\mathbf{c}; c_n^G) < k_n^G \Rightarrow P(\mathbf{c}) = 0$ and $A_i(\mathbf{c}) = 0$ for all i ; (3) $k(\mathbf{c}; c_n^G) = k_n^G \Rightarrow P(\mathbf{c}) = 1$ and $A_i(\mathbf{c}) = \frac{m_n}{k(\mathbf{c}; c_n^G)}$ for all i such that $c_i \leq c_n^G$; (4) $k(c; c_n^G) = k_n^G \Rightarrow P(c) = q_{k_n^G} \in (0, 1]$ and $A_i(\mathbf{c}) = q_{k_n^G} \frac{m_n}{k(\mathbf{c}; c_n^G)}$ for all i such that $c_i \leq c_n^G$.

The following proposition establishes that the optimal binary HO mechanism is a generalized VBO.

Proposition 4. For any $v \in (0, 1)$, m_n and n , there exists $c_n^b \in (0, 1)$, $k_n^b \geq m_n$, and q_n^b such that a k_n^b -generalized VBO mechanism with volunteer cutoff c_n^b and critical mass threshold k_n^b is an optimal honest and obedient binary mechanism.

Proof: The proof is carried out in two steps. In Step 1, we establish that the optimal binary mechanism is non-wasteful, meaning that it only activates 0 or m_n agents. In Step 2, we show that if the optimal binary mechanism activates m_n agents with positive probability with k volunteers, then it must activate m_n agents with probability one when there are more than k volunteers. This implies that the optimal binary mechanism is characterized by a threshold k_n^b as specified in Definition 1. See the appendix for details. ■

One might conjecture that the unique VBO characterized by equation 10 would be the optimal honest and obedient binary mechanism; i.e., the group succeeds if and only if there are enough volunteers (at least m_n). Using a higher threshold than m_n seems wasteful and is ex post suboptimal since it implies that there are events in which the group fails even

thought the number of volunteers is known (by the mechanism) to exceed the minimum number required for success. However, in principle, it could be ex ante optimal for the mechanism to commit to failure in some such events (e.g., $k_n^G = m_n + 1$) in order to create better incentives for the agents to self identify as volunteers and more generally relax the (HO) constraint. This could happen if increasing k_n^G above m_n leads to a higher volunteer cutoff, c_n^G .

5.2.2 Asymptotic Optimality of Generalized VBO mechanisms

In an optimal binary mechanism, the mechanism does not use detailed information regarding the type of the agent, just whether the agent is willing to be a volunteer. The optimal binary mechanism is therefore asymptotically optimal if the best HO mechanism also does not use detailed information regarding the types c and instead treats all low cost member (volunteers) approximately the same and treats all high cost members (free riders) essentially the same. The following lemma establishes this property:

Lemma 1. *Let $a_n^*(c), p_n^*(c)$ be an optimal honest and obedient mechanism. For any two types c' and c'' with $c'' > c' > 0$, $p_n^*(c') - p_n^*(c'') \rightarrow 0$ and $a_n^*(c') - a_n^*(c'') \rightarrow 0$ as $n \rightarrow \infty$.*

Proof: See Online Appendix.

To see the intuition of this result, note that by IC, $p_n^*(c)$ is non-increasing in c , so $p_n^*(c') \leq p_n^*(c \leq c')$ and $p_n^*(c'') \geq p_n^*(c \geq c'')$, where $p_n^*(c \leq c')$ and $p_n^*(c \geq c'')$ are the interim probabilities of success conditioning on, respectively, $c \leq c'$ and $c \geq c''$. Then we have,

$$p_n^*(c \leq c') = \tau_{0,n-1}P_B^n + (1 - \tau_{0,n-1})P_0^n$$

where $\tau_{0,n-1}$ is the probability that, out of the remaining $n - 1$ agents, there is at least one type $\tilde{c} \geq c''$, P_B^n is the expected probability of success conditioning on the presence of at least a type $\tilde{c} \geq c''$ and a type $\tilde{c} \leq c'$, and P_0^n is the expected probability of success conditioning on the presence of at least a type $\tilde{c} \leq c'$ and the absence of a type $\tilde{c} \geq c''$. Similarly, we have:

$$p_n^*(c \geq c'') = \tau_{1,n-1}P_B^n + (1 - \tau_{1,n-1})P_1^n$$

where $\tau_{1,n-1}$ is the probability that, out of the remaining $n - 1$ agents, there is at least one type $\tilde{c} \leq c'$, and P_1^n is the expected probability of success conditioning on the presence of at least a type $\tilde{c} \geq c''$ and the absence one type $\tilde{c} \leq c'$. But then we have:

$$0 \leq p_n^*(c') - p_n^*(c'') \leq (\tau_{0,n-1} - \tau_{1,n-1}) P_B^n + (1 - \tau_{0,n-1}) P_0^n - (1 - \tau_{1,n-1}) P_1^n$$

As $n \rightarrow \infty$, both $\tau_{0,n-1}$ and $\tau_{1,n-1}$ converge to 1. Since P_0^n , P_1^n and P_B^n are all bounded, we have that for any $\varepsilon > 0$, there is a n_ε such that $p_n^*(c') - p_n^*(c'') < \varepsilon$ for all $n > n_\varepsilon$.

An implication of Lemma 1 is that when n is large the optimal mechanism is characterized by a c_n^* such that for $c > c_n^*$, the required participation $a_n^*(c)$ is zero; and for $c \leq c_n^*$, participation is a non-increasing function which is approximately flat, even when the probability of success converges to a positive value. The next result shows that the utility obtained in such a mechanism converges to the utility that can be obtained in a binary mechanism. This fact combined with Proposition 4 implies that the optimal VBO is asymptotically optimal. Let V_n^G and V_n^* be the expected welfare generated in, respectively, the optimal binary HO mechanism (generalized VBO) and in the optimal HO mechanism when the number of agents is n . Putting this all together, we have:

Proposition 5. $\lim_{n \rightarrow \infty} V_n^G = \lim_{n \rightarrow \infty} V_n^*$.

Proof: See Appendix.

This proposition allows us to rule out situations in which the limit probability of success is positive in the optimal HO mechanism, but is zero in the optimal VBO. An implication of this is that whenever the limit probability of the optimal VBO converges to zero, then it converges to zero in *every* honest and obedient mechanism. We will use this fact in the next section to show that the limit probability in the best HO mechanism is the same as in the unorganized case.

5.2.3 The irrelevance of organizations in the limit as $n \rightarrow \infty$

In our earlier analysis of the simple VBO mechanism (Theorem 3), we proved that in the limit large groups succeed with probability 1 if $m_n \prec n^{2/3}$ and large groups fail with probability 1

if $m_n \succ n^{2/3}$. However that left open the question of whether the optimal honest and obedient mechanism might succeed with positive probability for some values of $m_n \succ n^{2/3}$. Since the simple VBO is not necessarily optimal, it could be the case that the optimal mechanism does much better than a simple VBO in large groups. In this section, we prove that the limiting properties of the simple VBO are shared by the optimal honest and obedient mechanism.

Let P_n^* denote the probability of success in the optimal honest and obedient mechanism as a function of n , for a given threshold m_n and value v . We have:

Theorem 4. *For any $v \in (0, 1)$:*

- *If $m_n \prec n^{2/3}$, then $\lim_{n \rightarrow \infty} P_n^* = \lim_{n \rightarrow \infty} P_n^G = \lim_{n \rightarrow \infty} P_n^O = \lim_{n \rightarrow \infty} P_n^U = 1$.*
- *If $m_n \succ n^{2/3}$, then $\lim_{n \rightarrow \infty} P_n^* = \lim_{n \rightarrow \infty} P_n^G = \lim_{n \rightarrow \infty} P_n^O = \lim_{n \rightarrow \infty} P_n^U = 0$.*

Proof: The first bullet point follows immediately, since the VBO achieves success with probability 1 in the limit and is approximately optimal. The second bullet point is proved in two steps. The first step establishes that if $m_n \succ n^{2/3}$, then the probability of success in the generalized VBO converges to zero, again using Stirling's approximation of the binomial distribution as in the proofs of Theorems 1 and 3. The second step applies Proposition 5 to show that the probability of success in the optimal HO mechanism also converges to zero. See the appendix for details. ■

We conclude this section discussing three implications of Theorem 4. The first implication, which we view as positive, is that a very simple and natural HO mechanism - the VBO - is approximately optimal. Furthermore, as illustrated in Section 4.3, the VBO produces large gains over unorganized groups in finite groups, even when n is large. Thus, the free rider problem can be significantly (though not entirely) mitigated by voluntary organizations that do not require coercion or taxes. The second implication is that the moral hazard problem is much worse than one might have thought for extremely large groups, i.e., in the limit. If mechanisms must satisfy honesty and obedience and operate voluntarily without coercion, then in the limit case of arbitrarily large groups such organizations are no more successful than unorganized groups. When $m_n \succ n^{2/3}$, this occurs even if the total benefit is strictly

higher than the total cost in the worst scenario in which $c_i = 1$ for all players. In contrast, strong organizations that allow for coercion and taxes/transfers (or even only coercion) will always achieve success as long as the total benefit for society $V_n = nv$ is larger than the cost in the worst case scenario in which $c_i = 1$ for all $i \in I$, regardless of the returns to scale. In our environment this situation arises when $\alpha_n < v$, so it holds for large groups if there are any positive returns to scale and in *all groups* if there are constant returns to scale and $\alpha < v$; in other environments, such a situation would arise, for example, when total demand for a public good is bounded above (Hellwig [2003]). Hence, a third implication is that for extremely large groups (call them *societies*), overcoming the free rider problem requires some form of coercion or taxation, except in those cases where the free rider problem disappears so fast ($m_n \prec n^{2/3}$) that no organization whatsoever is needed for success to be achieved.

6 Endogenous organizations: when do groups choose to organize?

The key insight in Olson [1965] is that we should expect successful collective action only when the free rider problem is not too severe: that is, for a given value v of the public good, when the number of interested agents n is not too large; or for a given n , when the value of the public good v is sufficiently large.²⁵ These observations motivate his claim that small groups with strong individual incentives will be much more effective than large groups in which individuals have weak incentives.

The results obtained above have implications that relate to these conjectures and also some additional insights about when one might expect successful groups to arise endogenously. On the one hand, the marginal impact of v or n on the probability of success depends on whether the group is organized or not and the extent to which the underlying technology displays increasing returns to scale: for example, without an organization, the marginal impact of v and n is exactly zero when n is large; it is positive only with an organization. So we cannot understand the true impact of individual preferences and the size of

²⁵See the discussion in chapters 1-2 in Olson [1965].

the population without first specifying their impact of the presence and quality of a group's organization.

On the other hand, whether the group might become organized or not depends on the underlying fundamentals of the economy, thus on v and n as well. We can evaluate the importance of these variables for the success of a group *only* when we include in the analysis their impact on the presence and effectiveness of an organization. To explore this idea, in this section we capitalize on the previous analysis to endogenize the presence of an organization and study how its endogeneity affects the impact of v and n on the ultimate success of a group.

We model the process of formation of an organization in a stylized, yet general way. Suppose the agents composing a group evaluate the opportunity of establishing an organization *ex ante*, before they know their individual costs of activism c . A VBO is formed if the increase in expected utility with the organization is larger than a given organizational fixed cost, κ : $\Delta V_n^* = V_n^O - V_n^U \geq \kappa(n)$, where V_n^O and V_n^U are, respectively, the expected utilities with and without an organization; and $\kappa(n)$ is the per person cost of forming and operating an organization with these n agents. We assume $\kappa(2) > 0$, and $\lim_{n \rightarrow \infty} \kappa(n) = \kappa > 0$.²⁶

The expected utility in a VBO with cut off c_n^O can be written as:

$$V_n^O = v \left[F(c_n^O) p_{1,n}(c_n^O) + (1 - F(c_n^O)) p_{2,n}(c_n^O) \right] - F(c_n^O) E(c; c_n^O) a_n^*, \quad (16)$$

where $p_{1,n}(c_n^O)$ and $p_{2,n}(c_n^O)$ are the probability of success for an agent, conditioning on being a volunteer and on not being a volunteer; and $E(c; c')$ is the expected c , conditioning on not being larger than c' . In a VBO, we must have $c_n^O a_n^* = v(p_{1,n}(c_n^O) - p_{2,n}(c_n^O))$, the expected

²⁶This simple model is intended capture a variety of environments. Consider these two polar examples. First, assume perfect substitutability and that there is an elite of $l \leq n$ agents each of whom can form the organization paying a fixed cost $\hat{\kappa}$. The organization is created if at least one member of the elite pays the cost; if the organization is formed, the members of the elite capture a share $v \leq 1$ of the total surplus $n\Delta V^*$. In this case, there is an equilibrium in which each member of the elite pays the cost with probability $\phi < 1$ and the condition for the establishment is $\Delta V^* \geq \kappa$ with $\kappa = \hat{\kappa} / [nB(0, l-1, \phi)v]$. The elite members internalize only a share of the benefit because they themselves may face a free rider problem. Second, assume perfect complementarity in the technology for the formation of the organization, so that each member of the elite needs to pay a cost $\hat{\kappa}$. In this case the organization forms if and only if $\Delta V^* \geq \kappa$ is satisfied with $\kappa = \hat{\kappa}/nv$ as described in the main text.

utility with an optimal organization can therefore be written as:

$$V_n^O = v \left[F(c_n^O) \left(1 - \frac{E(c; c_n^O)}{c_n^O} \right) B(m_n - 1, n - 1, F(c_n^O)) + \sum_{j=m_n}^{n-1} B(j, n - 1, F(c_n^O)) \right] \quad (17)$$

Following similar steps, the expected utility without an organization can be written as:

$$V_n^U = v \left[F(c_n^U) \left(1 - \frac{E(c; c_n^U)}{c_n^U} \right) B(m_n - 1, n - 1, F(c_n^U)) + \sum_{j=m_n}^{n-1} B(j, n - 1, F(c_n^U)) \right] \quad (18)$$

The effect of n on ΔV_n^* is complicated by the fact that n indirectly affects the thresholds for equilibrium participation c_n^O and c_n^U . Still, from the continuity of the functions in the square parentheses with respect to c_n^O and c_n^U , and the fact that we know for n large enough $c_n^U = 0$ and $c_n^O \rightarrow 0^+$, we can deduce the no organization will ever be formed for arbitrarily large groups:²⁷

Proposition 6. *There is a $n_\kappa > 0$ such that a VBO is formed only if $n \leq n_\kappa$.*

A similar discontinuity as highlighted above is generated by a change in v if we keep n constant. Again, signing the comparative statics in full generality is difficult because it involves evaluating how the mechanism cutoff for volunteers, c_n^O , changes relative to c_n^U as we change v . However, the effect can easily be signed when n is sufficiently large. We have:

Proposition 7. *There is a $n^* > 0$ such that for any $n > n^*$, ΔV_n^* is strictly increasing in v , so an organization is formed only if v is larger than a threshold v_n^* .*

Proof: See Appendix.

Propositions 6 and 7 are interesting because they suggests why the two factors highlighted by Olson (size and individual incentives) matter for a group's effectiveness. It is not just that as n increases or v decreases, we have a more severe free rider problem that depresses the probability of success. If it were only for this, the probability of success would change very little. A more important point is that the group organizes only for $n \leq n_\kappa$ and this

²⁷Note that the fact that c_n^U and c_n^O converge to zero does not imply the terms in parenthesis converge to zero; indeed, as we know from Theorem 3, they both converge to zero if $m_n \succ n^{2/3}$ and to one if $m_n \prec n^{2/3}$.

has important implications for effectiveness.²⁸ As n passes n_κ , effectiveness collapses to almost zero, since without an organization the probability of success is extremely small and insensitive of v and n . Proposition 6 also explains why we should expect a dichotomy of organizations: the small, organized, and effective groups on the one hand; and large, unorganized, and ineffective group on the other hand. What creates the dichotomy is the decision to organize that transform a continuous effect in a discrete drop in effectiveness.

7 Variations and discussions

7.1 On High-value environments ($v \geq 1$)

An assumption that we have maintained throughout the analysis is that $v < 1$, where 1 is the highest possible cost \bar{c} . This assumption is standard and, for the constant returns to scale case, implied by the stronger assumption requiring that the total benefit of success vn is less than the marginal cost αn in the worst case scenario in which all types have a cost of 1, so that $v < \alpha$ (see, for example, Mailath and Postlewaite [1990]). The case with $v \geq 1$, however, has an interesting peculiarity that is worth discussing. If $v \geq 1$, then, for weak organizations we obtain a result similar to Theorem 2, because randomly selecting a group of size m_n , regardless of individual costs, does not violate the obedience constraint for any type. Specifically:

Proposition 8. *If $v \geq 1$, MHRA is satisfied, and either $m_n \prec n$ or $m_n = \alpha n$ for some fixed $\alpha < 1$, then for all n the optimal direct mechanism satisfying (IC) and (IMH) is a random mechanism in which each g such that $|g| = \alpha n$ is activated with probability $1/\binom{n}{m_n}$ and each g such that $|g| \neq m_n$ is activated with probability 0. The probability of success equals 1.*

When $v \geq 1$, however, the limit probability of success in the symmetric equilibrium of an unorganized group remains 0 when $m \succ n^{2/3}$. An implication of Proposition 8, therefore, is that the limit equivalence of the probabilities of success in organized and unorganized groups is not valid anymore when $v \geq 1$. In this case, moral hazard is not a problem; the

²⁸The size threshold, n_κ , as well as the value threshold, v_n^* , both vary with the returns to scale. In principle, n_κ could be quite large.

only strategic problem faced by the members is coordination. Coordination can be easily solved by a honest and obedient mechanism, but is unsolvable in a symmetric equilibrium without an organization.²⁹

7.2 On the divisibility of tasks

In the previous analysis we have assumed that the decision to contribute is dichotomous: agent i either contributes at a cost c_i or not. For example, an agent participates in a rally or not; an agent signs a petition or not; joins a union or a committee or not. There are however cases in which the contribution can be split up. For example, suppose that an agent has up to one day to donate to a cause, say the organization of a charity. However, if the agent cannot donate one day, perhaps the agent can donate less, say one hour. It is easy to see that the analysis can be easily extended to this case, though the results are interesting only when we assume some economies of scale, if the task cannot be “atomized” too much relative to the cost of providing the effort: in this case the obedience constraint becomes moot (and the optimal mechanism becomes too powerful to generate plausible predictions).

To see this, assume that a contribution now can be divided in λ parts: when $\lambda = 1$ the contribution is, say, one day; when $\lambda = 24$, it is one hour, etc. Now the mechanism can ask each agent to contribute any discrete amount $x \in \{0, 1/\lambda, 2/\lambda, \dots, 1\}$, say from 0 hours to 24 hours. Assume that we need a total of m_n contribution units to achieve the collective goal and recall that the costs are distributed in $[0, \bar{c}]$ and \bar{c} can be possibly larger than 1: now we can require m_n agents providing one unit, or up to λm_n agents providing one hour. If we can choose λ so large that $\bar{c}/\lambda < v$ and $\lambda m_n \leq n$, then we can achieve the common goal with probability one with a mechanism equivalent to the optimal (IC) and (INTIR) mechanism of Theorem 2. In this case, we just ask λm_n agents at random to contribute $1/\lambda$ each. If $m_n \prec n$, then for any λ such that $\bar{c}/\lambda < v$, it will be true that $\lambda m_n \leq n$ for large n , so success with probability 1 is feasible for n large enough. In many plausible environments, however, economies of scale make it unrealistic to assume that divisibility is fine enough to

²⁹Of course, we can design an asymmetric equilibrium that achieves success with probability 1 in an unorganized group, but such an equilibrium would implicitly assume a solution of the coordination problem by ex ante selecting the “volunteers”.

guarantee that *all* types including the most extreme would be willing to obey it if asked. If we assume that there is a $\bar{\lambda}$ satisfying $\bar{c}/\bar{\lambda} > v$, then the obedience constraint will always be binding as in the analysis presented above. For instance, this is always true for any $\bar{\lambda}$, for \bar{c} large enough.

The analysis in previous sections carries over to this case where contributions can be discretely divisible rather than just dichotomous. In the dichotomous case a (reduced form) mechanism specifies a probability of success $p(c)$ and a probability of contributing $a(c)$, where $a(c) \in [0, 1]$ and non-increasing in c . As before, a mechanism specifies an interim probability of success $p(c)$ and an interim *expected contribution* $a(c)$, where again $a(c) \in [0, 1]$ and non-increasing in c . The analysis is completely analogous. Indeed, the same logic as in Section 5 suggests that, for finite n , $a(c)$ will be non increasing and positive up to a threshold $c^* = \min \{c \leq vp(c^*) - c^*a(c^*) \leq vp(\bar{c})\}$ and then $a(c) = 0$ for $c > c^*$, just as above. Moreover, $a(c)$ will become flat as $n \rightarrow \infty$, so a VBO with $a(c) > a$ for “volunteers” and zero for free riders with a higher c will be asymptotically optimal, just as in the previous analysis.

7.3 Alternative cost distributions

While we allow for general distributions of the players’ costs, three assumptions are maintained throughout the analysis. The first assumption is that there is a positive density of zero cost types, i.e., $f(0) > 0$. This assumption is standard in the literature and has been adopted in classic pivotal agent models to study voter turnout and candidate competition (e.g., Ledyard (1984), Myerson (2000)). Relaxing this assumption has no impact on the result for $m_n \succ n^{2/3}$, as well as for all the results concerning strong organizations. When $m_n \prec n^{2/3}$, the existence of a sequence of equilibria with the probability of success converging to one may depend on the rate at which $f(c)$ converges to zero as $c \rightarrow 0$.

A second assumption is that $F(0) = 0$, i.e., there are no types that actually like to contribute. With negative cost types, i.e. $F(0) > 0$, it is trivial to see that the probability of success equals one for any increasing returns, i.e., if $m_n \prec n$, with or without an organization, since $\alpha_n < F(0)$ for sufficiently large n , except if the distribution of types also depends on

n and becomes sufficiently small as n increases, i.e. $F_n(0) < m_n/n$ for n sufficiently large. The results for constant returns to scale with HO mechanisms are essentially unchanged as long as $F(0) < \alpha$.

A third assumption is that there is a continuum of types. The analysis can be extended to allow for an arbitrary finite number of types. More can be said in the extreme case where there are only two types, c_L and c_H , as studied in Ledyard and Palfrey (1994). For this case Battaglini and Palfrey (2023) have shown that a simple VBO is exactly optimal for finite n if $\frac{c_L}{v} < \frac{1-\phi}{1-\phi} \frac{\alpha n}{\alpha n}$, where $\phi \in (0, 1)$ is the probability of the low type.

8 Conclusions

We have developed a model of collective action in which a group can organize by constructing communication mechanisms to elicit private information and coordinate the actions of its members. We have stressed the importance of requiring the mechanism to be obedient, besides the more familiar requirements of incentive compatibility and individual rationality. Mechanisms that are only incentive compatible and individually rational make sure that members are willing to join a group and reveal their types, but they require members to commit to carry out the mechanism's recommendations, thus assuming away a key aspect of the moral hazard problem. Obedience is not generally included in classic mechanism design problems, since in these applications mechanisms map vectors of type profiles to allocations; in collective action problems, on the contrary, mechanisms only map vectors of type profiles to recommendations: allocations are the decentralized results of the members' individual actions.

Strong mechanisms that satisfy interim incentive compatibility and individually rationality, but omit the obedience constraint, can fully solve a group's problem, achieving success with probability one even with no side payments. Moreover, regardless of how fast m_n grows, success is always guaranteed when the total benefit of the group is larger than the cost in the worst case scenario (in which all members have maximal costs). The predicted success of optimal honest and obedient mechanisms is typically much more modest, which highlights

the importance of realistic moral hazard or obedience constraints, and at the same time suggests that interim individual rationality is too weak a condition. We showed that when m_n grows slower than $n^{2/3}$, success is achievable with certainty, with or without an organization. When m_n grows faster than $n^{2/3}$, however, success is impossible in the limit even if total benefit is always larger than total cost. Nonetheless organizing a group with a simple honest and obedient institution gives a group a key advantage even in large groups. The real benefit of an organization is that even when the probability of success of the optimal HO mechanism converges to zero, it does so at a much slower rate than without an organization (which is exactly zero after a finite n). The rate of convergence of the probability of success when m_n grows faster than $n^{2/3}$ is always strictly slower than exponential.

There are numerous ways the theory might usefully be extended. In our analysis we have relied on a very simple base model of collective action, a classic threshold public good game. It may be possible to explore the themes described above in much more general economic environments in which the size of the common goal that can be chosen by the collectivity is a continuous variable: as, for example, when the group does not only choose to build a bridge, but also its quality and its capacity. In addition, we have studied a completely static model. Many collective action problems are dynamic. The ideas presented here could be embedded in dynamic environments to extend previous work that has studied contribution games in dynamic environments with no organizations (see, for instance, Matthews [2013] and Battaglini et al. [2014]).

Another direction that we have begun to explore is the study of how multiple groups, each facing their own potentially competing collective action projects, strategically interact with each other. Groups may strategically interact because their respective goals are substitutes, as when there is a budget constraint that allows only a subset of projects to be realized. Or they can interact in environments with complementarities, which leads to “a collective action problem in a collective action problem”: that is, the groups need to solve a collective action problem between themselves in the face of common goals, but each group also needs to solve its own internal collective action design problem in order for the group to elicit contributions.

The theory presented here also provides inspiration for new empirical questions that can be studied with laboratory experiments and possibly field work. We mentioned a significant literature in experimental economics that has studied contribution games with structured and unstructured preplay communication. Most of this literature has focused on environments with complete information, or with only a few players. We leave for future research an empirical investigation of the effectiveness of the VBOs characterized in this paper and the comparison of their performance with unorganized groups.

Appendix

Proof of Theorem 1

In Step 1, we show that $m_n \prec n^{2/3}$ implies that $F(c_n^U)$ is sufficiently larger than α_n for all n sufficiently large and in the limit. In Step 2, we show that this guarantees that $\lim_{n \rightarrow \infty} P_n^U = 1$. In Step 3, we show that if $m_n \succ n^{2/3}$, then $P_n^U = 0$ for n sufficiently large.

Step 1. We first show that $m_n \prec n^{2/3}$ implies that $F(c_n^U) > \alpha_n$ for any n sufficiently large. Recall that the threshold, c_n^U , is the largest solution for $c \in [0, 1]$ to the following equation: $c = vB(\alpha_n n - 1, n - 1, F(c))$. Therefore, $F(c_n^U) > \alpha_n$ for any n large if, for sufficiently large n , we have:

$$\alpha_n < F(vB(\alpha_n n - 1, n - 1, \alpha_n))$$

This condition guarantees that there is an intersection on the right of α_n . See Figure 1. The following lemma will prove useful in the argument.

Lemma A1. *If $m_n \prec n^{2/3}$ then $B(\alpha_n n - 1, n - 1, \alpha_n)/\alpha_n \rightarrow \infty$; if $m_n \succ n^{2/3}$ then $B(\alpha_n n - 1, n - 1, \alpha_n)/\alpha_n \rightarrow 0$.*

Proof. We can approximate the binomial combinatorial term for large n using Stirling's formula: $\binom{n}{k} = \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}}$. First note that:

$$\binom{n-1}{\alpha_n n - 1} = \frac{(n-1)!}{(\alpha_n n - 1)!(n - \alpha_n n)!} = \frac{\alpha_n n}{n} \frac{n!}{\alpha_n n \cdot (\alpha_n n - 1)!(n - \alpha_n n)!} = \alpha_n \binom{n}{\alpha_n n}$$

Applying Stirling's formula yields:

$$\binom{n-1}{\alpha_n n - 1} \simeq \alpha_n \sqrt{\frac{1}{2\pi \alpha_n (1 - \alpha_n) n}} \left[\frac{1}{(\alpha_n)^{\alpha_n} (1 - \alpha_n)^{(1 - \alpha_n)}} \right]^n$$

where \simeq means that the two sequences converge to zero at the same speed. So an approximation of $B(\alpha_n n - 1, n - 1, \alpha_n)$ is given by:

$$B(\alpha_n n - 1, n - 1, \alpha_n) = \binom{n-1}{\alpha_n n - 1} \frac{[(\alpha_n)^{\alpha_n} (1 - \alpha_n)^{1 - \alpha_n}]^n}{\alpha_n} \simeq \sqrt{\frac{1}{2\pi \alpha_n (1 - \alpha_n) n}}$$

We therefore have:

$$\frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} \simeq \frac{1}{\alpha_n} \sqrt{\frac{1}{2\pi(\alpha_n)(1 - \alpha_n)n}}$$

We have $\frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} \rightarrow \infty$ if $\left(\frac{m_n}{n}\right)^3 \left(1 - \frac{m_n}{n}\right)n \rightarrow 0$ as $n \rightarrow \infty$, a condition satisfied if $m_n/n^{2/3} \rightarrow 0$, or $m_n \prec n^{2/3}$; we have $\frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} \rightarrow 0$ if $\left(\frac{m_n}{n}\right)^3 \left(1 - \frac{m_n}{n}\right)n \rightarrow \infty$ as $n \rightarrow \infty$, a condition satisfied if $m_n/n^{2/3} \rightarrow \infty$, or $m_n \succ n^{2/3}$. ■

We can now prove that $m_n \prec n^{2/3}$ implies that $F(c_n^U) > \alpha_n$ for all n sufficiently large. To this goal, note that as $n \rightarrow \infty$, $vB(\alpha_n n - 1, n - 1, \alpha_n) \rightarrow 0$, so we can write

$$F[vB(\alpha_n n - 1, n - 1, \alpha_n)] = vf(0) \cdot B(\alpha_n n - 1, n - 1, \alpha_n) + o(B(\alpha_n n - 1, n - 1, \alpha_n))$$

where $o(B(\alpha_n n - 1, n - 1, \alpha_n))/\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. It follows that:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{F[vB(\alpha_n n - 1, n - 1, \alpha_n)]}{\alpha_n} &= \lim_{n \rightarrow \infty} \left[\frac{vf(0) \cdot \frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} + \frac{o(B(\alpha_n n - 1, n - 1, \alpha_n))}{\alpha_n}}{\frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n}} \right] \\ &= \lim_{n \rightarrow \infty} \left[\frac{vf(0)}{\frac{o(B(\alpha_n n - 1, n - 1, \alpha_n))}{B(\alpha_n n - 1, n - 1, \alpha_n)}} \right] \frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} \\ &= vf(0) \lim_{n \rightarrow \infty} \frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n} \end{aligned}$$

We conclude that whenever $\frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n}$ converges to 0 or diverges at ∞ , so does $\frac{F[vB(\alpha_n n - 1, n - 1, \alpha_n)]}{\alpha_n}$.

This implies that when $m_n \prec n^{2/3}$, then $\frac{F[vB(\alpha_n n - 1, n - 1, \alpha_n)]}{\alpha_n} \rightarrow \infty$, implying $\alpha_n < F[vB(\alpha_n n - 1, n - 1, \alpha_n)]$, so $F(c_n^U) > \alpha_n$.

Step 2. We now prove that if $m_n \prec n^{2/3}$ the probability of success for the group converges to 1. We proceed in two sub-steps.

Step 2.1. Assume first that $\frac{F(c_n^U)}{\alpha_n} \rightarrow 1$. We have:

$$\begin{aligned} B(\alpha_n n - 1, n - 1, F(c_n^U)) &= \binom{n-1}{\alpha_n n - 1} \frac{[(F(c_n^U))^{\alpha_n} (1 - F(c_n^U))^{1-\alpha_n}]^n}{F(c_n^U)} \\ &= \binom{n-1}{\alpha_n n - 1} \frac{[(\alpha_n)^{\alpha_n} (1 - \alpha_n)^{1-\alpha_n}]^n}{\alpha_n} \frac{[(F(c_n^U))^{\alpha_n} (1 - F(c_n^U))^{1-\alpha_n}]^n}{[(\alpha_n)^{\alpha_n} (1 - \alpha_n)^{1-\alpha_n}]^n} \\ &\simeq \binom{n-1}{\alpha_n n - 1} \frac{[(\alpha_n)^{\alpha_n} (1 - \alpha_n)^{1-\alpha_n}]^n}{\alpha_n} \end{aligned}$$

But note that by the definition of c_n^U , $\alpha_n \simeq v f(0) c_n^U$, and the fact that $m_n \prec n^{2/3}$ we must have:

$$\begin{aligned} 1 &= v \frac{B(\alpha_n n - 1, n - 1, F(c_n^U))}{c_n^U} \simeq v f(0) \frac{B(\alpha_n n - 1, n - 1, F(c_n^U))}{\alpha_n} \\ &\simeq \frac{v f(0)}{\alpha_n} \sqrt{\frac{1}{2\pi (\alpha_n) (1 - \alpha_n) n}} \rightarrow \infty, \end{aligned}$$

a contradiction. We must therefore have that in equilibrium: $\frac{F(c_n^U)}{\alpha_n} \rightarrow \zeta > 1$, with ζ possibly arbitrarily large.

Step 2.2. It follows from steps 2.1 that we can assume that $F(c_n^U)/\alpha_n \rightarrow \zeta > 1$. Note that the probability of failure is equal to the probability that fewer than $\alpha_n n$ agents volunteer and thus it can be bounded above by:

$$\begin{aligned} \Pr(k \leq \alpha_n n) &= \Pr\left(\frac{k}{n} \leq \alpha_n\right) = \Pr\left(\frac{k}{n} \leq F(c_n^U) - F(c_n^U) \left(1 - \frac{\alpha_n}{F(c_n^U)}\right)\right) \\ &\leq \Pr\left(\frac{k}{n} \leq F(c_n^U) - \vartheta F(c_n^U)\right) \leq \Pr\left[\left|\frac{k}{n} - F(c_n^U)\right| \geq \vartheta F(c_n^U)\right] \\ &= \Pr\left[\left|\frac{k}{n} - F(c_n^U)\right| \geq \frac{\sqrt{F(c_n^U)(1 - F(c_n^U))}}{\sqrt{n}} \frac{\sqrt{n\vartheta F(c_n^U)}}{\sqrt{F(c_n^U)(1 - F(c_n^U))}}\right] \\ &= \Pr\left[\left|\frac{k}{n} - F(c_n^U)\right| \geq \sigma_{c_n^U}\left(\frac{k}{n}\right) \cdot \frac{\vartheta \sqrt{n F(c_n^U)}}{\sqrt{(1 - F(c_n^U))}}\right] \leq \left(\frac{\sqrt{(1 - F(c_n^U))}}{\vartheta \sqrt{n F(c_n^U)}}\right)^2 \rightarrow 0 \end{aligned}$$

where in the second line we used $\frac{\alpha_n}{F(c_n^U)} < 1$, so $\left(1 - \frac{\alpha_n}{F(c_n^U)}\right) \geq \vartheta$ for some $\vartheta > 0$; in the fourth

line we define $\sigma_{c_n^U}(\frac{k}{n}) = \frac{\sqrt{F(c_n^U)(1-F(c_n^U))}}{\sqrt{n}}$ and used Chebyshev's inequality; and in the last step of the fourth line (“ \rightarrow ”), we used the fact that $n\alpha_n = m_n \rightarrow \infty$.

Step 3. We now show that if $m_n \succ n^{2/3}$, then $\lim_{n \rightarrow \infty} P_n^U = 0$. We first establish that $\lim_{n \rightarrow \infty} \frac{p_n^U}{\alpha_n} = 0$. Assume not, so $\lim_{n \rightarrow \infty} \frac{p_n^U}{\alpha_n} = \zeta$ for some $\zeta > 0$. From the equilibrium condition we must have:

$$p_n^U = F(vB(\alpha_n n - 1, n - 1, p_n^U))$$

Note however that:

$$\begin{aligned} B(\alpha_n n - 1, n - 1, p_n^U) &= \binom{n-1}{\alpha_n n - 1} \frac{[(p_n^U)^{\alpha_n} (1 - p_n^U)^{1-\alpha_n}]^n}{p_n^U} \\ &= \frac{\alpha_n}{p_n^U} \binom{n}{\alpha_n n} [(p_n^U)^{\alpha_n} (1 - p_n^U)^{1-\alpha_n}]^n \simeq \frac{1}{\zeta} \cdot \sqrt{\frac{1}{2\pi\alpha_n(1-\alpha_n)n}} [\xi_n]^n \end{aligned}$$

with $\xi_n = \frac{(p_n^U)^{\alpha_n} (1 - p_n^U)^{1-\alpha_n}}{(\alpha_n)^{\alpha_n} (1 - \alpha_n)^{1-\alpha_n}} \leq 1$ for any n (since for any p_n^U , $(p_n^U)^{\alpha_n} (1 - p_n^U)^{1-\alpha_n} \leq (\alpha_n)^{\alpha_n} (1 - \alpha_n)^{1-\alpha_n}$). So, by Lemma A1, we have that for large n :

$$1 = \frac{F(vB(\alpha_n n - 1, n - 1, p_n^U))}{p_n^U} \simeq \frac{vf(0)(\xi_n)^n}{\alpha_n} \sqrt{\frac{1}{2\pi\alpha_n(1-\alpha_n)n}} \rightarrow 0$$

where in the second step (“ \simeq ”) we used the fact that $p_n^U = \zeta\alpha_n$, and in the last step (“ \rightarrow ”), we used the fact that $\frac{1}{\alpha_n} \sqrt{\frac{1}{2\pi\alpha_n(1-\alpha_n)n}} \rightarrow 0$ since $m_n \succ n^{2/3}$ and $(\xi_n)^n \leq 1$. This is a contradiction, implying that $\lim_{n \rightarrow \infty} \frac{p_n^U}{\alpha_n} = 0$. We next use this to show that $p_n^U = 0$ for large n . By definition we have:

$$p_n^U = F(vB(\alpha_n n - 1, n - 1, p_n^U)) = \Psi(\alpha_n, n, p_n^U) \quad (19)$$

Note that since we don't have an equilibrium $p_n^U > 0$ on the right of α_n , we must have an equilibrium $\tilde{p}_n > 0$ on the left of α_n with $\Psi'(\alpha_n, n, p_n^U) < 1$, where $\Psi'(\alpha_n, n, p_n^U)$ denotes the derivative of $\Psi(\alpha_n, n, p_n^U)$ with respect to p_n^U for a given α_n and n . Note that for any constant $\epsilon > 0$ arbitrarily small, there is a n_ϵ such that for $n > n_\epsilon$, $\Psi'(\alpha_n, n, p_n^U) > vf(0)(1-\epsilon)B'(\alpha_n n - 1, n - 1, p_n^U)$, where $B'(\alpha_n n - 1, n - 1, p_n^U)$ denotes the derivative of $B(\alpha_n n - 1, n - 1, p_n^U)$

with respect to p_n^U for a given α_n and n . We can write:

$$\begin{aligned}
B'(\alpha_n n - 1, n - 1, p_n^U) &= B(\alpha_n n - 1, n - 1, p_n^U) \left[\frac{\alpha_n n - 1}{p_n^U} - \frac{n - \alpha_n n}{1 - p_n^U} \right] \\
&\rightarrow \frac{p_n^U}{f(0)v} \left[\frac{\alpha_n n - 1}{p_n^U} - \frac{n - \alpha_n n}{1 - p_n^U} \right] = \frac{(\alpha_n - p_n^U) n - 1 + p_n^U}{f(0)v (1 - p_n^U)} \\
&= \frac{\left(1 - \frac{p_n^U}{\alpha_n}\right) \alpha_n - \frac{1 - p_n^U}{n}}{f(0)v (1 - p_n^U)} n = \frac{\left[\left(1 - \frac{p_n^U}{\alpha_n}\right) - \frac{1 - p_n^U}{\alpha_n n}\right]}{f(0)v (1 - p_n^U)} \alpha_n n \rightarrow \frac{m_n}{f(0)v} \rightarrow \infty
\end{aligned}$$

where the equilibrium condition (19) and the fact that

$$F(vB(\alpha_n n - 1, n - 1, p_n^U)) \simeq f(0)vB(\alpha_n n - 1, n - 1, p_n^U)$$

for n large is used in the second line; and the last line follows from the earlier result that $\frac{p_n^U}{\alpha_n} \rightarrow 0$ when $m_n \succ n^{2/3}$ and $m_n \rightarrow \infty$. This leads to a contradiction, since it implies that for n large at any positive intersection $\Psi'(\alpha_n, n, p_n^U)$ is arbitrarily large. We conclude that the only equilibrium is $p_n^U = 0$. ■

Proof of Theorem 2

The paragraph immediately before the statement of Theorem 2 already proved that the random mechanism satisfies (IC) and (INTIR) under the conditions of parts 1 and 2 of the theorem, and it obviously achieves a probability of success equal to 1. Here we show that if, in addition, F satisfies MHRA then it is also the optimal mechanism.

For a given n , consider the relaxed problem:

$$\begin{aligned}
&\max_{p(\cdot), a(\cdot)} \left\{ vp(0) - E \left[a(c) \cdot \frac{1 - F(c)}{f(c)} \right] \right\} \\
&\text{s.t. } a(c) \text{ is non-increasing with } a(c) \in [0, 1] \\
&\text{and } p, a \text{ feasible}
\end{aligned} \tag{20}$$

derived from (9) by eliminating the (INTIR) constraint, and let $\mu_n(\mathbf{c})$ with associated reduced form mechanism $a_n(c)$, $p_n(c)$ be its solution. We proceed in three steps. In Step 1, starting

from $\mu_n(\mathbf{c})$ we present a perturbed mechanism $\mu_n^\gamma(\mathbf{c})$ and show it is incentive compatible and feasible. In Step 2 we show that such a perturbation strictly improves the relaxed problem (20) if $a_n(c)$ is strictly decreasing. In Step 3, we show that the solution of the relaxed problem is $p_n(c) = 1$, $a(c) = \alpha_n$. Moreover, when $m_n \preceq n$ and $v > \alpha$, or when $m_n \prec n$ and n is large, then this solution is a solution of the full problem (9).

Step 1. Since the argument is true for any n , we omit here the subscript n for simplicity. Let μ be any feasible and incentive compatible mechanism. Consider the following “flattening” perturbation of the mechanism. After a profile of reports \mathbf{c} , the perturbed mechanism is defined by a new activity function that uses $\mu(\mathbf{c})$ with probability $1 - \gamma$, and $\mu(\tilde{\mathbf{c}})$ and with probability γ , where $\tilde{\mathbf{c}}$ is a vector in which all components $c_i > 0$ are replaced with i.i.d realizations in $(0, 1]$ from $F(x)$ (and components $c_i = 0$ are left unchanged). Let $\bar{a} = \int_0^1 a(x)f(x)dx$ and $\bar{p} = \int_0^1 p(x)f(x)dx$. This new allocation generates a reduced form mechanism:

$$p^\gamma(c_i) = \gamma\bar{p} + (1 - \gamma)p(c_i), \text{ and } a^\gamma(c_i) = \gamma\bar{a} + (1 - \gamma)a(c_i)$$

for $c_i \in (0, 1]$ and $p^\gamma(0) = p(0)$, $a^\gamma(0) = a(0)$. Note that since $a(0) \geq a(c_i)$ and $p(0) \geq p(c_i)$ for all $c_i \in [0, 1]$, we must have that $a(0) \geq \bar{a} = \int_0^1 a(x)f(x)dx$ and similarly $p(0) \geq \bar{p}$.

The new reduced form allocation is clearly feasible since we have shown the feasible activity function that generates it. It also does not change $p(0)$. Note that after the change incentive compatibility is satisfied since $a^\gamma(c_i)$ is non increasing in $[0, 1]$ and, after the change, we have:

$$\begin{aligned} U^\gamma(x) &= \gamma[v\bar{p} - \bar{a}x] + (1 - \gamma)[vp(x) - a(x)x] \\ &= vp^\gamma(x) - a^\gamma(x)x \end{aligned}$$

For $c > 0$ and $c' \geq 0$, we have:

$$\begin{aligned} vp^\gamma(c) - a^\gamma(c)c &= \gamma[v\bar{p} - \bar{a}c] + (1 - \gamma)[vp(c) - a(c)c] \\ &\geq \gamma[v\bar{p} - \bar{a}c] + (1 - \gamma)[vp(c') - a(c')c] \\ &= vp^\gamma(c') - a^\gamma(c')c \end{aligned}$$

Moreover a type 0 does not want to imitate a type $c > 0$:

$$vp^\gamma(0) \geq \gamma [v\bar{p}] + (1 - \gamma) [vp(0)] \geq \gamma [v\bar{p}] + (1 - \gamma) [vp(c')] = vp^\gamma(c')$$

where the first inequality follows from the fact that $p(0) \geq \bar{p}$. Given IC, by the usual argument, we have:

$$[U^\gamma]'(x) = -\gamma\bar{a} - (1 - \gamma)a'(x) = -a^\gamma(c_i)$$

So the change is feasible in the relaxed problem.

Step 2. To see that it increases the objective function, we need to show that $-\int_0^1 a^\gamma(x)(1 - F(x))dx$ increases in γ , since $vp(0)$ is unchanged by the change. We can write it as:

$$\bar{a} \cdot \int_0^1 G(x) \frac{a^\gamma(x)}{\bar{a}} f(x) dx$$

where $G(x) = -\frac{1-F(x)}{f(x)}$ and \bar{a} is $\int_0^1 a(x)f(x)dx$. Note that $\frac{a^\gamma(x)f(x)}{\bar{a}}$ is a density since $\frac{a^\gamma(x)f(x)}{\bar{a}} \geq 0$ and $\int_0^1 \frac{a^\gamma(x)f(x)}{\bar{a}} dx = 1$. By MHRA, $G(x)$ is monotone non-decreasing in x , so the result is proven if we prove that an increase in γ implies a first order stochastic dominance improvement in $\frac{a^\gamma(x)}{\bar{a}} f(x)$. Define: $\Gamma^\gamma(t) = \int_0^t \frac{a^\gamma(x)}{\bar{a}} f(x) dx$. We prove the result if $\partial\Gamma^\gamma(t)/\partial\gamma < 1$ for all $t < 1$. We have:

$$\begin{aligned} \frac{\partial}{\partial\gamma}\Gamma^\gamma(t) &= \frac{\partial}{\partial\gamma} \left[\frac{1}{\bar{a}} \int_0^t [\gamma\bar{a} + (1 - \gamma)a(x)] f(x) dx \right] \\ &= \int_0^t \left[1 - \frac{a(x)}{\bar{a}} \right] f(x) dx = F(t) \left[1 - \frac{E[a(x); x \leq t]}{E[a(x)]} \right] \leq 0 \end{aligned}$$

where the last inequality follows from the fact that $a(x)$ non-increasing in x . It follows that increasing γ improves the relaxed problem, which is maximized at $\gamma = 1$. When $\gamma = 1$, feasibility and the (IC) are satisfied: so $\gamma = 1$ is optimal for the original problem as well.

Step 3. From Step 2 we know that the optimal mechanism solving the relaxed problem is independent of c : a_n^S, p_n^S . It is easy to see that this mechanism will always activate a coalition of size m_n . Assume not. Three cases are possible. First, the mechanism activates a coalition of size larger than m_n ; second the mechanism selects a non empty coalition of

size smaller than m_n . In the first case, just modify the mechanism by imposing that all activated coalitions are reduced to a size m_n by randomly selecting agents to drop; in the second, modify the mechanism by imposing that coalitions that are smaller than m_n are not selected and a coalition of size m_n is selected instead, with equal probability on all coalitions of size m_n . This leaves p_n^S unchanged and reduces a_n^S . No constraint is violated and the objective function is increased, a contradiction. The third case is that the mechanism always select a coalition of size m_n , but with probability $p_n^S < 1$. It is easy to see that this is not optimal since by increasing p_n^S we obtain a marginal improvement in utility equal to $1 - \alpha_n \frac{c}{v}$. We conclude that the optimal solution of the relaxed problem is $p_n^S = 1$ and $a_n^S = \alpha_n$.

Proof of Proposition 1

We first prove that for any α_n, n , $Y_n(c)$ has a unique fixed point c_n^O . We then prove that a simple VBO is incentive compatible if and only if the volunteer cutoff is c_n^O , and $c_n^O \in (c_n^U, v)$. Finally, we establish the VBO is honest and obedient. The following lemma will prove useful.

Lemma A2. $\frac{B(\alpha_n n - 1 + j, n - 1, p)}{B(\alpha_n n - 1, n - 1, p)} = \prod_{k=1}^j \frac{(n - \alpha_n n - 1 - k)}{\alpha_n n - 1 + k} \cdot \left(\frac{p}{1-p}\right)^j$.

Proof: See online appendix ■

We now proceed in three steps.

Step 1. For any α_n, n , $Y_n(c)$ is defined as:

$$Y_n(c) = \frac{v B(\alpha_n n - 1, n - 1, F(c))}{\sum_{j=\alpha_n n - 1}^{n-1} \frac{\alpha_n n}{j+1} B(j, n - 1, (F(c)))}.$$

We can rewrite it as:

$$Y_n(c) = \frac{v}{1 + \sum_{j=\alpha_n n}^{n-1} \frac{\alpha_n n}{j+1} \frac{B(j, n-1, F(c))}{B(\alpha_n n - 1, n-1, F(c))}} = F \left(\frac{v}{1 + \sum_{j=1}^{n-\alpha_n n} \frac{\alpha_n n}{j+\alpha_n n} \frac{B(\alpha_n n - 1 + j, n-1, F(c))}{B(\alpha_n n - 1, n-1, F(c))}} \right)$$

We now show that:

$$\sum_{j=1}^{n-\alpha_n n} \frac{\alpha_n n}{j + \alpha_n n} \frac{B(\alpha_n n - 1 + j, n - 1, F(c))}{B(\alpha_n n - 1, n - 1, F(c))}$$

is strictly increasing in c , so $Y_n(c)$ is continuous and strictly decreasing in c . By Lemma A2, we have:

$$\frac{B(\alpha_n n - 1 + j, n - 1, F(c))}{B(\alpha_n n - 1, n - 1, F(c))} = \prod_{i=1}^j \frac{n - \alpha_n n + 1 - i}{\alpha_n n - 1 + i} \cdot \left(\frac{F(c)}{1 - F(c)} \right)^j$$

It follows that:

$$\sum_{j=1}^{n-\alpha_n n} \frac{\alpha_n n}{j + \alpha_n n} \frac{B(\alpha_n n - 1 + j, n - 1, F(c))}{B(\alpha_n n - 1, n - 1, F(c))} = \sum_{j=1}^{n-\alpha_n n} \frac{\alpha_n n}{j + \alpha_n n} \cdot \prod_{i=1}^j \frac{n - \alpha_n n + 1 - i}{\alpha_n n - 1 + i} \cdot \left(\frac{F(c)}{1 - F(c)} \right)^j$$

which is increasing in c . Moreover, it is easy to see that $Y_n(0) = F(v) > 0$ and $Y_n(v) = 0 < v$. Hence $Y_n(c)$ has a unique fixed point c_n^O in $(0, v)$.

Step 2. Incentive compatibility requires that $U(c) = vp_n^O(c) - ca_n^O(c) \geq vp_n^O(c') - ca_n^O(c')$ for all $c, c' \in [0, 1]$, where $p_n^O(c), a_n^O(c)$ is the reduced form direct mechanism described by the VBO. We now show that the mechanism is IC when the threshold is c_n^O such that $c_n^O = Y_n(c_n^O)$. Given c_n^O , we let $p_{1,n}^O$ denote the (constant) interim probability of success for all types $c \leq c_n^O$, let $p_{2,n}^O$ denote the (constant) interim probability of success for all types $c > c_n^O$, and let a_n^O denote the (constant) interim probability of being activated for all types $c \leq c_n^O$. From the definition of a VBO, $p_{1,n}^O, p_{2,n}^O, a_n^O$ are given by the following formulas:

$$a_n^O = \sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{k+1} B(k, n-1, F(c_n^O)) \quad (21)$$

$$p_{1,n}^O = \sum_{k=\alpha_n n-1}^{n-1} B(k, n-1, F(c_n^O)) \quad (22)$$

$$p_{2,n}^O = \sum_{k=\alpha_n n}^{n-1} B(k, n-1, F(c_n^O)) \quad (23)$$

The equilibrium condition for IC is:

$$ac_n^O = v(p_{1,n}^O - p_{2,n}^O) \quad (24)$$

By substituting equations (21), (22), and (23) into equation (24) for any n , we can obtain an expression for $c_n^O(v)$, the VBO volunteer threshold cost as a function of group size and

the value of success:

$$c_n^O(v) = v \frac{B(\alpha_n n - 1, n - 1, F(c_n^O))}{\sum_{k=\alpha_n n - 1}^{n-1} \frac{\alpha_n n}{k+1} B(k, n - 1, F(c_n^O))} \quad (25)$$

where the numerator on the right hand side is $p_{1,n}^O - p_{2,n}^O$ and the denominator is a_n^O , the probability a volunteer is activated. It is easy to see that (25) implies the statement in the Proposition. Moreover, we get $c_n^O > c_n^U$ because $Y_n(c) > vB(m_n - 1, n - 1, F(c))$ for all c, v, m_n, n . It follows that $c_n^O \in (c_n^U, v)$, as stated.

Step 3. To establish that the VBO is honest and obedient, first observe that all group members whose recommended action is to free ride will obey the recommendation because, if all other members are obedient, then either 0 or exactly m_n other members will volunteer. Hence, their participation will not affect success or failure so they are better off free riding. Second, all members with type c_i whose recommended action is to activate will obey the recommendation if $c_i \leq v$ because, if all other members are obedient, then exactly $m_n - 1$ other members have been recommended to activate and will do so. Hence their payoff is 0 if they disobey the recommendation to activate and $v - c_i > 0$ if they obey. It follows that all types $c_i \leq v$ find it optimal to obey no matter what type they have previously reported. Types $c_i > v$, instead, always find it optimal to free ride regardless of the recommendation. Therefore they cannot strictly improve their payoff by reporting $c'_i \neq c_i$. Formally: $E_{\mathbf{c}_{-i}} \left[\sum_{g \in I} \mu_g(\tilde{c}_i, \mathbf{c}_{-i}) u_g^i(c) \right] \geq E_{\mathbf{c}_{-i}} \left[\sum_{g \in I} \mu_g(\tilde{c}_i, \mathbf{c}_{-i}) u_{\xi_i(g)}^i(c) \right]$. This condition and (IC) imply:

$$E_{\mathbf{c}_{-i}} \left[\sum_{g \in I} \mu_g(c_i, \mathbf{c}_{-i}) u_g^i(c) \right] \geq E_{\mathbf{c}_{-i}} \left[\sum_{g \in I} \mu_g(\tilde{c}_i, \mathbf{c}_{-i}) u_g^i(c) \right] \quad (26)$$

$$\geq E_{\mathbf{c}_{-i}} \left[\sum_{g \in I} \mu_g(\tilde{c}_i, \mathbf{c}_{-i}) u_{\xi_i(g)}^i(c) \right] \quad (27)$$

for any $i = 1, \dots, n$, $c_i, \tilde{c}_i \in [0, 1]$ and any function $\xi_i(g)$ mapping g to either $\{g, g \setminus \{i\}\}$ if $g \in I_i$, or $\{g, g \cup \{i\}\}$ if $g \notin I_i$. ■

Proof of Theorem 3

The first part of Theorem 3 follows from Proposition 1 and part (1) of Theorem 1: for any n , the probability of success of an organized group using a VBO mechanism is greater than or equal to the probability of success of an unorganized group. Since the probability of success of an unorganized group converges to one when $m_n \prec n^{2/3}$, the same must be true for an organized group using a VBO mechanism.

For the second part, we now proceed in three Steps.

Step 1. We first prove that there is a constant $a_\infty > 0$ such that

$$\lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) = a_\infty.$$

The details of this step of the proof are in the online appendix.

Step 2. Given that $\lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) = a_\infty > 0$, we can now prove that there is a constant $\vartheta < 1$ such that $\lim_{n \rightarrow \infty} \frac{F(c_n^O)}{\alpha_n} < \vartheta$.

The details of this step of the proof are in the online appendix.

Step 3. Note that by Step 2, $\alpha_n - F(c_n^O) \geq (1 - \vartheta)\alpha_n$ for some $\vartheta < 1$. Following standard steps (see for instance the proof of Theorem 1), we have:

$$\begin{aligned} \Pr(k \geq \alpha_n n) &\leq \Pr\left[\left|\frac{k}{n} - F(c_n^O)\right| \geq (1 - \vartheta)\alpha_n\right] \\ &\leq \Pr\left[\left|\frac{k}{n} - F(c_n^O)\right| \geq \sigma_{c_n^O}\left(\frac{k}{n}\right) \cdot \frac{\sqrt{n\alpha_n}(1 - \vartheta)}{\sqrt{\vartheta(1 - F(c_n^O))}}\right] \leq \left(\frac{\sqrt{\vartheta(1 - F(c_n^O))}}{\sqrt{n\alpha_n}(1 - \vartheta)}\right)^2 \rightarrow 0 \end{aligned}$$

where $\sigma_{c_n^O}\left(\frac{k}{n}\right) = \frac{\sqrt{F(c_n^O)(1 - F(c_n^O))}}{\sqrt{n}}$. This proves the result. \blacksquare

Proof of Proposition 3

We can bound the probability of success in an optimal honest and obedient mechanism as follows. Let $D\left(\frac{k}{n} \| p\right) = \frac{k}{n} \log \frac{k/n}{p} + (1 - \frac{k}{n}) \log \frac{1 - k/n}{1 - p}$ be the Kullback–Leibler divergence, or

relative entropy, from p to $\frac{k}{n}$. We can write:

$$\begin{aligned} \lim_{n \rightarrow \infty} P_n^* &= \lim_{n \rightarrow \infty} \sum_{k=m_n}^n B(k, n, p_n^*) \geq \lim_{n \rightarrow \infty} P_n^O = \sum_{k=m_n}^n B(k, n, p_n^O) \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{\sqrt{8m_n(1 - \frac{m_n}{n})}} \exp\left(-nD\left(\frac{m_n}{n} \parallel p_n^O\right)\right) \simeq \lim_{n \rightarrow \infty} \frac{1}{\sqrt{8m_n}} \exp\left(-nD\left(\frac{m_n}{n} \parallel p_n^O\right)\right) \end{aligned}$$

where in the last line we used the lower bound on the tail of a Binomial distribution (Lemma 4.7.2 in Ash [1990]). By Proposition 2, $\lim_{n \rightarrow \infty} \frac{p_n^O}{m_n/n} \geq 1$, so for any $\epsilon > 0$, there exists n_ϵ such that for $n > n_\epsilon$, we have $D\left(\frac{m_n}{n} \parallel p_n^*\right) \leq \epsilon/2$. So

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{P_n^*}{e^{-\epsilon n}} &\geq \lim_{n \rightarrow \infty} \frac{\frac{1}{\sqrt{8m_n}} \exp\left(-nD\left(\frac{m_n}{n} \parallel p_n^*\right)\right)}{e^{-\epsilon n}} \\ &= \lim_{n \rightarrow \infty} \frac{e^{[\epsilon - \exp(-nD(\frac{m_n}{n} \parallel p_n^*))]n}}{\sqrt{8m_n}} \geq \lim_{n \rightarrow \infty} \frac{e^{\frac{\epsilon}{2}n}}{\sqrt{8m_n}} = \infty \end{aligned}$$

So for any $\epsilon > 0$, P_n^* converges strictly faster than $e^{-\epsilon n}$. We conclude that P_n^* converges to zero at a rate that is strictly slower than exponential. \blacksquare

Proof of Proposition 4

We will omit n here as a subscript for simplicity whenever it does not create confusion. Let μ^b be an optimal binary mechanism and let c^b be the volunteer cut-point associated with μ^b , and q_k be the corresponding probability of success when there are k volunteers. We will prove the result that if $q_k > 0$ for some $k \geq m$, then $q_{k+j} = 1$ for $j > 0$. This implies that there is a k^b such that $q_j = 0$ for $j < k^b$ and $q_j = 1$ for $j > k^b$ and at most at one k we have $q_{k^b} \in (0, 1)$. So the optimal binary mechanism is a k^b -VBO except at most for an event with probability that converges to zero as $n \rightarrow 0$, i.e. when there are exactly k^b volunteers.

We proceed in two steps. In Step 1, we establish that the optimal binary mechanism is non-wasteful, meaning that it does not ever activate more agents than necessary; Step 2 shows that it is characterized by a threshold k^b .

Step 1. We first show that the optimal honest and obedient binary mechanism must be non-wasteful in the sense that whenever a group is activated, there are exactly m members

in the activated group. We prove this by contradiction by supposing that μ^b is wasteful at a positive measure set of profiles and then showing that it can be improved. First, define a new mechanism, μ' , that is exactly the same as μ for all coalitions of size m , but eliminates all waste by reducing all activated successful coalitions to a size m by randomly selecting agents to drop out, and by not activating unsuccessful coalitions that are smaller than m . This leaves c^b , p_1^μ and p_2^μ unchanged and reduces a^μ to $a^{\mu'} < a^\mu$. This implies that $a^{\mu'}c^b < v(p_1^\mu - p_2^\mu)$, so some cost types in a neighborhood above c^b are strictly better off volunteering, which violates incentive compatibility. Now, for any $\tilde{c} > c^b$, consider a modified version of μ' , denoted by $\tilde{\mu}'$ that has the same success probabilities, $\{q_k\}_{k=m}^n$ as μ' except that all members with $c < \tilde{c}$ are volunteers, so there is a bigger pool of volunteers. This increases p_1^μ and p_2^μ to $p_1^{\tilde{\mu}'} > p_1^\mu$ and $p_2^{\tilde{\mu}'} > p_2^\mu$ and changes $a^{\mu'}$ to $\tilde{a}^{\mu'}$. Denote by $\tilde{c}^b > c^b$ the first such value of $\tilde{c} > c^b$ such that $\tilde{a}^{\mu'}\tilde{c}^b = v(p_1^{\tilde{\mu}'} - p_2^{\tilde{\mu}'})$. (Such a point exists by the intermediate value theorem.) Denote by $u(c; c^b)$ the interim expected utility of a member with cost c under μ , and denote by $u(c; \tilde{c}^b)$ the interim expected utility of a member with cost c under the modified mechanism, $\tilde{\mu}'$ with volunteer cutoff $\tilde{c}^b > c^b$. Because $p_2^{\tilde{\mu}'} > p_2^\mu$, we know that $u(c; \tilde{c}^b) = u(\tilde{c}^b; \tilde{c}^b) > u(c; c^b)$ for all $c \geq \tilde{c}^b$, so these members are strictly better off. For $c \in (c^b, \tilde{c}^b)$ we have $u(c; \tilde{c}^b) \geq u(\tilde{c}^b; \tilde{c}^b) > u(c; c^b)$, so these members are also better off. Finally, for all $c \in [0, c^b)$ (the volunteers under μ) are better off because for each $k \geq m$ for which $q_k > 0$ there are a positive measure of additional profiles \mathbf{c} with exactly k volunteers, and for each such additional profile, the c -type member in $[0, \tilde{c}^b)$ gets a conditional expected utility of $(v - \frac{m}{k}c)q_k > 0$. (Such members receive the same conditional expected utility for all other profiles.) Hence, $u(c; \tilde{c}^b) > u(c; c^b)$ for all $c \leq c^b$. Hence all agents are better off under $\tilde{\mu}'$ than under μ . All constraints are satisfied and the objective function is increased, a contradiction. Hence the optimal mechanism is non-wasteful

Step 2. If $q_k > 0$ for $k \geq m$ and $q_{k+j} < 1$ for $j > 1$, then there must be a k' such that $q_{k'} > 0$ for $k' \geq m$ and $q_{k'+1} < 1$, so we only need to prove the result for the case of $j = 1$.

Assume by contradiction that $q_k > 0$ for some $k \geq m$ and $q_{k+1} < 1$. Let c^b be the minimum cost above which an agent is activated with probability zero. Then incentive

compatibility is binding at c^b if $ac^b = v(p_1^b - p_2^b)$, where:

$$p_1^b - p_2^b = B(n-1, n-1, F(c^b))q_n + \sum_{k=m}^{n-1} [B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b))]q_k$$

and

$$a = \sum_{k=m-1}^{n-1} \frac{m}{1+k} B(k, n-1, F(c^b))q_{k+1}.$$

We can marginally reduce q_k by $-dq_k < 0$ and marginally increase q_{k+1} by $dq_{k+1} > 0$ so that the (IC) constraint is unchanged, thus keeping c^b constant. This requires:

$$\begin{aligned} c^b \left[-\frac{m}{k} B(k-1, n-1, F(c^b)) + \frac{m}{k+1} B(k, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \right] dq_k \\ = \left[\begin{array}{c} - (B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b))) \\ + B(k, n-1, F(c^b)) - B(k+1, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \end{array} \right] dq_k \end{aligned} \quad (28)$$

Note we can write:

$$\begin{aligned} p_1^b - p_2^b &= B(n-1, n-1, F(c^b))q_n \\ &+ \sum_{k=m}^{n-1} [B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b))]q_k = \sum_{k=m}^n \Theta_k q_k \end{aligned} \quad (29)$$

where we denote:

$$\Theta_n = B(n-1, n-1, F(c^b))$$

$$\Theta_k = B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b)) \text{ for } k = n-1, \dots, m$$

We can rewrite the previous expression as:

$$\Theta_k = B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b)) = B(k, n-1, F(c^b)) \left[\frac{k}{n-k} \frac{1-F(c^b)}{F(c^b)} - 1 \right]$$

Similarly, we have:

$$\begin{aligned}
\Theta_{k+1} &= B(k, n-1, F(c^b)) - B(k+1, n-1, F(c^b)) \\
&= \frac{(n-1)!}{(k)!(n-k-1)!} (F(c^b))^k (1-F(c^b))^{n-k-1} \\
&\quad - \frac{(n-1)!}{(k+1)!(n-k-2)!} (F(c^b))^{k+1} (1-F(c^b))^{n-k-2} \\
&= B(k+1, n-1, F(c^b)) \left[\frac{k+1}{n-k-1} \frac{1-F(c^b)}{F(c^b)} - 1 \right]
\end{aligned}$$

Substituting into (28) gives:

$$\begin{aligned}
&c^b \left[-\frac{m}{k} B(k-1, n-1, F(c^b)) + \frac{m}{k+1} B(k, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \right] dq_k \\
&= \left[\begin{aligned} &-(B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b))) \\ &+ B(k, n-1, F(c^b)) - B(k+1, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \end{aligned} \right] dq_k
\end{aligned}$$

or

$$\begin{aligned}
&c^b \left[\begin{aligned} &-\frac{m}{k} \frac{k}{n-k} \frac{1-F(c^b)}{F(c^b)} B(k, n-1, F(c^b)) \\ &+ \frac{m}{k+1} \frac{k+1}{n-k-1} \frac{1-F(c^b)}{F(c^b)} B(k+1, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \end{aligned} \right] dq_k \\
&= \left[\begin{aligned} &-B(k, n-1, F(c^b)) \left[\frac{k}{n-k} \frac{1-F(c^b)}{F(c^b)} - 1 \right] \\ &+ B(k+1, n-1, F(c^b)) \left[\frac{k+1}{n-k-1} \frac{1-F(c^b)}{F(c^b)} - 1 \right] \frac{dq_{k+1}}{dq_k} \end{aligned} \right] dq_k
\end{aligned}$$

It follows that:

$$\begin{aligned}
&\frac{dq_{k+1}}{dq_k} \left[\left(c^b \frac{m}{k+1} - 1 \right) \frac{k+1}{n-k-1} \frac{1-F(c^b)}{F(c^b)} + 1 \right] B(k+1, n-1, F(c^b)) \\
&= \left[\left(c^b \frac{m}{k} - 1 \right) \frac{k}{n-k} \frac{1-F(c^b)}{F(c^b)} + 1 \right] B(k, n-1, F(c^b)) \\
&\Leftrightarrow \frac{dq_{k+1}}{dq_k} = \frac{1 - \left(1 - c^b \frac{m}{k} \right) \frac{k}{n-k} \frac{1-F(c^b)}{F(c^b)}}{1 - \left(1 - c^b \frac{m}{k+1} \right) \frac{k+1}{n-k-1} \frac{1-F(c^b)}{F(c^b)}} \frac{B(k, n-1, F(c^b))}{B(k+1, n-1, F(c^b))} = T_{n,k} \frac{B(k, n-1, F(c^b))}{B(k+1, n-1, F(c^b))}
\end{aligned}$$

where:

$$T_{n,k} = \frac{1 - \left(1 - c^b \frac{m}{k}\right) \frac{k}{n-k} \frac{1-F(c^b)}{F(c^b)}}{1 - \left(1 - c^b \frac{m}{k+1}\right) \frac{k+1}{n-k-1} \frac{1-F(c^b)}{F(c^b)}} > 1 \Leftrightarrow \frac{k - c^b m}{n - k} < \frac{k - c^b m + 1}{n - k - 1} \Leftrightarrow n > c^b m$$

After this change, the probability of success increases, indeed we have:

$$\begin{aligned} dP_n &= \left[-B(k, n-1, F(c^b)) + B(k+1, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \right] dq_k \\ &= \left[-B(k, n-1, F(c^b)) + B(k+1, n-1, F(c^b)) \cdot T_{n,k} \frac{B(k, n-1, F(c^b))}{B(k+1, n-1, F(c^b))} \right] dq_k \\ &> (T_{n,k} - 1) B(k, n-1, F(c^b)) \cdot dq_k > 0 \end{aligned}$$

Since the probability of success increases but the average probability of participation remains constant (since c^b is unchanged), we must increase welfare, ceteris paribus. This implies that the original mechanism was not optimal, a contradiction. ■

Proof of Proposition 5

As discussed in Section 5.1, we derive the objective function in the relaxed problem (15) using the (IC) constraint in a similar way as in (9). The constraint in the second line of (15) is just the implication of IC, also present in (9). The constraint in the third line of (15) follows from the following lemma. For simplicity, we omit the subscript n in the expressions of Lemma A4.

Lemma A4: *If (IC) and (1) hold, then $a(c) = 0$ for $c > c^*$, where*

$$c^* = \min \{c \leq v \mid vp(c) - ca(c) \leq vp_2\}.$$

Proof: By (1), $c > v$ implies $a(c) = 0$. Consider any $c \in [c^*, v]$. By the definition of c^* , $U(c) \leq U(v)$, and (IC) implies $U(c) \geq U(v)$, so $U(c) - U(v) = 0$ for $c \in [c^*, v]$. This implies: $\int_{c^*}^v a(x)dx = \int_v^{c^*} U'(x)dx = U(c^*) - U(v) = 0$. Since $a(c)$ is nonnegative, we get $a(c) = 0$ for $c > c^*$. ■

Let $V_n^{**}(c)$ denote the expected value for a type c in an optimal mechanism that solves the relaxed problem (15) and $V_n^{**} = E_c\{V_n^{**}(c)\}$ be the value of the objective function in (15). Note that $V_n^{**} \geq V_n^*$, where V_n^* is the optimal mechanism in an honest and obedient mechanism, since (15) is a relaxed version of (14).

Define an ε -bounded mechanism, $\tilde{\mu}_n^\varepsilon(\mathbf{c})$, and associated reduced form mechanism $\tilde{a}_n^\varepsilon(c)$, $\tilde{p}_n^\varepsilon(c)$ as follows. It solves the problem for the optimal mechanism (15), but with an additional condition:

$$\tilde{a}_n(c) > 0 \Rightarrow p_n(0) - \tilde{p}_n(c) < \varepsilon$$

The value for a type c and the expected values of this mechanism are $\tilde{V}_n^\varepsilon(c)$ and \tilde{V}_n^ε , respectively. When $\varepsilon = 1$ (or larger), the additional constraint is slack, so $\tilde{V}_n^\varepsilon = V_n^{**}$. When $\varepsilon = 0$, $\tilde{\mu}_n^\varepsilon(\mathbf{c})$ is a binary mechanism, that is there is a \tilde{c}_n^ε such that $\tilde{p}_n^\varepsilon(c) = \tilde{p}_n^\varepsilon(0)$ for $c \leq \tilde{c}_n^\varepsilon$ and $\tilde{p}_n^\varepsilon(c) = \tilde{p}_n^\varepsilon(1)$ for $c > \tilde{c}_n^\varepsilon$. Moreover, incentive compatibility implies that $\tilde{a}_n^\varepsilon(c) = \tilde{a}_n^\varepsilon(0)$ for $c \leq \tilde{c}_n^\varepsilon$ and $\tilde{a}_n^\varepsilon(c) = 0$ for $c > \tilde{c}_n^\varepsilon$. We denote a binary mechanism as follows: $\mu_n^b(\mathbf{c})$ and associated $a_n^b(c)$, $p_n^b(c)$ with values $V_n^b(c)$ and V_n^b .

We proceed in two steps:

Step 1: For any η , there exists n_η such that $n > n_\eta \implies V_n^b \geq V_n^{**} - \eta$, and hence $V_n^b \geq V_n^* - \eta$.

Consider $D_n = V_n^{**} - \tilde{V}_n^0 = V_n^{**} - V_n^b$. There are two possibilities:

(1) $\lim_{n \rightarrow \infty} D_n = 0$. In this case:

$$\lim_{n \rightarrow \infty} D_n = \lim_{n \rightarrow \infty} (V_n^{**} - V_n^b) = \lim_{n \rightarrow \infty} (V_n^{**} - V_n^G) = 0$$

where V_n^G is the value in the optimal generalized VBO and we are done.

(2) $\lim_{n \rightarrow \infty} D_n = D > 0$. In this case let $\eta \in (0, D)$, and for any such η and any n , define $\varepsilon(n, \eta)$ as follows:

$$V_n^b = \tilde{V}_n^{\varepsilon(n, \eta)} - \eta/2$$

Note that $\varepsilon(n, \eta) \in (0, 1)$ for any n . It follows that $\lim_{n \rightarrow \infty} \varepsilon(n, \eta)$ exists and $\lim_{n \rightarrow \infty} \varepsilon(n, \eta) = \varepsilon(\eta) \in [0, 1]$.

Suppose that $\varepsilon(\eta) > 0$. Then for any $\varepsilon' \leq \varepsilon(\eta)$, there is a n_η^1 such that for $n > n_\eta^1$ we

have $V_n^b \geq \tilde{V}_n^{\varepsilon'} - \eta/2$. From Lemma 1, we know that there is a n_η^2 such that for $n > n_\eta^2$ we have $\tilde{V}_n^\varepsilon = V_n^{**}$, since the additional constraint in the ε -bounded mechanism becomes slack. We conclude that for $n > \max\{n_\eta^1, n_\eta^2\}$, $V_n^b \geq V_n^{**} - \eta$, a contradiction with the assumption that $\lim_{n \rightarrow \infty} D_n = D > 0$. We conclude that we must have $\varepsilon(\eta) = 0$.

The rest of the proof of Step 1 relies on the following lemma:

Lemma A5. *If $\lim_{n \rightarrow \infty} \varepsilon(n, \eta) = 0$, then for any arbitrarily small $\epsilon \in (0, \eta/2)$, there is an n_ϵ such that for $n > n_\epsilon$, we have $\tilde{V}_n^{\varepsilon(n, \eta)} \leq \tilde{V}_n^0 + \epsilon$.*

Proof. See online appendix. ■

Take $\epsilon < \eta/2$. From Lemma A5 there is a n_ϵ such that for $n > n_\epsilon$, $V_n^b = \tilde{V}_n^{\varepsilon(n, \eta)} - \eta \leq \tilde{V}_n^0 + \epsilon - \eta/2 < V_n^b$, a contradiction. Form the fact that we obtain a contradiction for any $\varepsilon(\eta) \geq 0$, we conclude that $\lim_{n \rightarrow \infty} (V_n^{**} - V_n^G) = 0$. ■

Step 2. We can now put together Step 1 and Proposition 4 to argue that a VBO is approximately optimal for n large. Since $V_n^b \leq V_n^*$, Step 1 implies that $|V_n^b - V_n^*| \rightarrow 0$ as $n \rightarrow \infty$. Proposition 4, moreover, shows that the optimal binary mechanism is a generalized VBO with threshold k_n^* . Let $VBO(k_n^*)$ be a VBO with threshold $k = k_n^*$ and no mixing for $k = k_n^*$. The $VBO(k_n^*)$ generates utility that converges to the utility of the generalized VBO with threshold k_n^* , since the probability of exactly $k = k_n^*$ volunteers converges to zero. Since the generalized VBO is equivalent to an optimal binary mechanism that generates utility V_n^b , we have $|V_n^b - V_n^{VBO(k_n^*)}| \rightarrow 0$. Hence, we have that for any η there is a n_η such that for $n > n_\eta$ $V_n^{VBO(k_n^*)} \geq V_n^* - \eta$ for some threshold k_n^* , which implies the result. ■

Proof of Theorem 4

We focus here on the second bullet point of the proposition. We prove the result in two steps:

Step 1. We first prove that $\lim_{n \rightarrow \infty} \frac{p_n^{\theta_n}}{\theta_n} < 1$, where $p_n^{\theta_n} = F(c_n^{\theta_n})$ and $c_n^{\theta_n}$ is the cutoff for

participation with θ_n . In equilibrium we must have:

$$p_n^{\theta_n} = F \left(\frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})} \right)$$

Consider the right hand side. By the Mean Value Theorem we can write:

$$F \left(\frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})} \right) = F(0) + v f(\xi) \frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})}$$

where $\xi \in \left[0, \frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})} \right]$ and the last term uses the residual in the Lagrange form. Thus, if we define $\bar{f} = \max_{c \in [0,1]} f(c)$, we have:

$$F \left(\frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})} \right) \leq (v\bar{f}) \cdot \frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})} \quad (30)$$

since $F(0) = 0$. We now prove that there is a constant $\vartheta < 1$ such that $\lim_{n \rightarrow \infty} \frac{p_n^{\theta_n}}{\theta_n} < \vartheta$. Assume not, so $\lim_{n \rightarrow \infty} \frac{p_n^{\theta_n}}{\theta_n} = \zeta \geq 1$. Following a very similar argument as in Step 1 of Theorem 3, we can prove that there is a constant $a_\infty > 0$ such that $\sum_{j=\theta_n n-1}^{n-1} \frac{\theta_n}{j+1} B(j, n - 1, p_n^{\theta_n}) \geq a_\infty$. We therefore have:

$$\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n}) = \frac{\alpha_n}{\theta_n} \sum_{j=\theta_n n-1}^{n-1} \frac{\theta_n n}{j+1} B(j, n - 1, p_n^{\theta_n}) \simeq \frac{\alpha_n}{\theta_n}$$

Following a very similar argument as in Step 2 of Theorem 3, we can prove:

$$\frac{B(\theta_n n - 1, n - 1, p_n^{\theta_n})}{p_n^{\theta_n}} \leq \frac{B(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\zeta \theta_n} \simeq \frac{1}{\theta_n} \sqrt{\frac{1}{2\pi \theta_n (1 - \theta_n) n}} \simeq \sqrt{\frac{1}{\theta_n^3 n}}$$

Using these facts, we now note that in equilibrium we must have

$$\begin{aligned} 1 &= \frac{1}{p_n^{\theta_n}} F \left(\frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1} \frac{m_n}{j+1} B(j, n - 1, p_n^{\theta_n})} \right) \leq \frac{v\bar{f}}{a_\infty} \frac{\theta_n}{\alpha_n} \frac{B(\theta_n n - 1, n - 1, p_n^{\theta_n})}{p_n^{\theta_n}} \\ &\simeq \frac{v\bar{f}}{a_\infty} \sqrt{\frac{1}{(\alpha_n)^2 \theta_n n}} \rightarrow 0, \end{aligned}$$

where the last step follows since $(\alpha_n)^2 \theta_n n = (\alpha_n)^3 \frac{\theta_n}{\alpha_n} n > (\alpha_n)^3 n$, which converges to infinity if $m_n \succ n^{2/3}$ as assumed. We therefore have a contradiction. We conclude that there is a constant $\vartheta < 1$ such that $\lim_{n \rightarrow \infty} \frac{p_n}{\theta_n} < \vartheta$. Using this fact, the result follows from the same argument as in Step 3 of Theorem 3.

Step 2. We now show that if the probability of success in the optimal binary HO mechanism (which is a general VBO) converges to zero, then the probability of success in a fully optimal HO mechanism converges to zero as well. This implies that the expected welfare in the two mechanisms converge to the same value. For this we use Proposition 5, that the generalize VBO is an approximately optimal HO mechanism.

Suppose by contradiction that the probability of success in the optimal mechanism P_n^* (not necessarily binary or VBO) converges to some positive value $P^* > 0$, but the probability of success in the optimal generalized VBO mechanism P_n^G converges to zero. Let W_n^* and W_n^G be the expected per capita welfare in the optimal mechanism and in the optimal VBO. Note that for any ε , there is a $n_{1,\varepsilon}$ such that for $n > n_{1,\varepsilon}$:

$$W_n^G = vP_n^G \left(1 - E \left(\frac{a^G(c)}{P_n^G} \cdot \frac{c}{v} \right) \right) \leq vP_n^G \leq \varepsilon/2$$

since by assumption $P_n^G \rightarrow 0$. Moreover for any ε , there is a $n_{2,\varepsilon}$ such that for $n > n_{2,\varepsilon}$

$$W_n^* = vP_n^* \left(1 - E \left(\frac{a_n^*(c)}{P_n^*} \cdot \frac{c}{v} \right) \right) \geq vP^* - \varepsilon/2 > 0$$

since: (a) for all $c \leq v$, $a_n^*(c) \frac{c}{v} \leq p_n^*(c) - p_n^*(v) \rightarrow 0$, as proved earlier; and (b) $P_n^* \rightarrow P^* > 0$. It follows that for any ε , there is a $n_\varepsilon = \max\{n_{1,\varepsilon}, n_{2,\varepsilon}\}$ such that for $n > n_\varepsilon$, $W_n^* - W_n^G > vP^* - \varepsilon$.

By Proposition 5, for any arbitrarily small $\eta > 0$, there is a n_η such that for $n > n_\eta$, $|W_n^* - W_n^G| < \eta$, where W_n^* and W_n^G are the expected per capita welfare in the optimal mechanism and in the optimal VBO. It follows that for n large, $\eta + \varepsilon > |W_n^* - W_n^G| \geq vP^*$, which is a contradiction since η and ε are both arbitrarily small, and vP^* is bounded away from zero. ■

Proof of Proposition 7

If $m_n \succ n^{2/3}$, $c_n^O > 0$ for all n and $c_n^U = 0$ for n sufficiently large, so ΔV_n^* is proportional to v , and thus clearly increasing in v . If instead $m_n \prec n^{2/3}$, then both $F(c_n^O) > m_n/n$ and $F(c_n^U) > m_n/n$, so the first terms in the square parenthesis of (17) and (18) converge to zero faster than the second terms, hence it can be ignored for large n . For large enough n , we therefore have:

$$EU_O(c_n^O) - EU_U(c_n^U) \simeq v \left[\sum_{j=\alpha_n n}^{n-1} B(j, n-1, c_n^O) - \sum_{j=\alpha_n n}^{n-1} B(j, n-1, c_n^U) \right]$$

which is strictly increasing in v since $\sum_{j=\alpha_n n}^{n-1} B(j, n-1, c)$ is strictly increasing in c and $c_n^O > c_n^U$ by Proposition 1. ■

References

- [1] Archetti, M., and Scheuring, I. (2012). “Game theory of public goods in one-shot social dilemmas without assortment,” *Journal of Theoretical Biology*, 299, 9–20.
- [2] Ash, R. (1990). *Information Theory*. Dover Publications, New York.
- [3] d’Aspremont, Claude and L.-A. Gérard-Varet (1979), “Incentives and Incomplete Information,” *Journal of Public Economics*, 11, 25-45.
- [4] d’Aspremont C., J. Crémer, and L. A. Gérard-Varet (1990), “Incentives and the existence of Pareto-optimal revelation mechanisms,” *Journal of Economic Theory*, 51(2): 233-254.
- [5] Barrett, S. and Dannenberg, A. (2014). “Sensitivity of Collective Action to Uncertainty about Climate Tipping Points,” *Nature Climate Change*, 4(1), 36–39.
- [6] Battaglini M. and R. Benabou (2003): “Trust, Coordination, and The Industrial Organization of Political Activism,” *Journal of the European Economic Association*, 1(4), 2003.
- [7] Battaglini M. (2017), “Public Protests and Policy Making,” *Quarterly Journal of Economics*, 2017, 132(1): 485-549.
- [8] Battaglini M. and B. Harstad (2016), “Participation and Duration of Environmental Agreements”, *Journal of Political Economy*, 2016, 124(1): 160-204.
- [9] Battaglini M., S. Nunnari and T. Palfrey (2014), “Dynamic Free Riding with Irreversible Investments,” *American Economic Review*, 104(9): 2858-71.
- [10] Battaglini M. and T. Palfrey (2023), “Organizing Collective Action: Olson Revisited,” NBER working paper n. 30991.
- [11] Bierbrauer, F. and M. Hellwig (2016), “Robustly Coalition-Proof Incentive Mechanisms for Public Good Provision are Voting Mechanisms and Vice Versa,” *Review of Economic Studies*, 83: 1440–1464.

- [12] Bierbrauer, F. and J. Winkelmann (2020), “All or nothing: State capacity and optimal public goods provision,” *Journal of Economic Theory*, 185: 1-15.
- [13] Braver, S. L., and L. A. Wilson (1986), “Choices in Social Dilemmas: Effects of Communication within Subgroups,” *The Journal of Conflict Resolution*, 30(1): 51–62.
- [14] Chamberlin, J. R. (1974), “Provision of Collective Goods as a Function of Group Size,” *American Political Science Review*, 68: 707.
- [15] Cremer, J. and McLean, R.P. (1988), “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions,” *Econometrica*, 56, 1247-1257.
- [16] Diekmann, A. (1985), “Volunteer’s dilemma”, *Journal of Conflict Resolution*, 29: 605-610.
- [17] Dixit, A. and M. Olson (2000), “Does voluntary participation undermine the Coase Theorem?” *Journal of Public Economics*, 76 (3): 309-35.
- [18] Dziuda W. A. Gitmez and M. Shadmeher (2021), “The Difficulty of Easy Projects,” *American Economic Review: Insights*, 3(3): 285–302.
- [19] Esteban J. and D. Ray (2001), “Collective action and the group paradox,” *American Political Science Review*, 95(3):663–672.
- [20] Goldlucke, S. and T. Troger (2020), “The Multiple-Volunteers Dilemma,” *Manuscript*, University of Mannheim.
- [21] Healy, P. J. 2010. ”Equilibrium participation in public goods allocations.” *Review of Economic Design*. 14: 27–50.
- [22] Hellwig, M. F (2003), “Public Good Provision with Many Participants,” *The Review of Economic Studies*, 70: 589-614.
- [23] van de Kragt A., J. Orbell and R. Dawes (1983), “Organizing groups for collective action,” *American Political Science Review*, 80(4): 1171-1185.

- [24] Ledyard, J. O. (1984), "The pure theory of large two-candidate elections," *Public Choice*, 44:, 7-41.
- [25] Ledyard, J. O. and T. R. Palfrey (1994), "Voting and Lottery Drafts as Efficient Public Goods Mechanisms," *The Review of Economic Studies*, 61(2): 327-355.
- [26] Ledyard, John O. and Thomas R. Palfrey (1999), "A Characterization of Interim Efficiency with Public Goods," *Econometrica*, 67(2):435-48.
- [27] Ledyard, John O. and Thomas R. Palfrey (2002), "The Approximation of Efficient Public Good Mechanisms by Simple Voting Schemes," *Journal of Public Economics*, 83(2): 153-72.
- [28] Mailath, G. J. and A. Postlewaite (1990), "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, 57:351-67.
- [29] Matthews, S. (2013), "Achievable Outcomes of Dynamic Contribution Games," *Theoretical Economics*, 8(2): 365-403.
- [30] Myerson, R. B. (1982), "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics* 10: 67-81.
- [31] Myerson, Roger B. (2000), "Large Poisson Games," *Journal of Economic Theory*, 94: 7-45.
- [32] Nöldeke G. and J. Peña (2020), "Group size and collective action in a binary contribution game," *Journal of Mathematical Economics*, 88: 42-51.
- [33] Olson, M. (1965), *The Logic of Collective Action*. Harvard University Press: Cambridge.
- [34] Ostrom E. (1990), *Governing the Commons*, Cambridge University Press.
- [35] Ostrom E. and J. Walker (1990), "Communication in a Commons: Cooperation Without External Enforcement," *Contemporary Laboratory Research in Political Economy*, ed. Tom Palfrey, p. 287-322, University Of Michigan Press.

- [36] Ostrom E., J. Walker and R. Gardner (1992), “Covenants With and Without a Sword: Self-Governance is Possible,” *American Political Science Review*, 86(2): 404-417.
- [37] Palfrey, T.R. and H. Rosenthal (1984), “Participation and the Provision of Public Goods: a Strategic Analysis,” *Journal of Public Economics*, 24:171-193.
- [38] Palfrey, T. R. and H. Rosenthal (1991), “Testing for effects of cheap talk in a public goods game with private information,” *Games and Economic Behavior*, 3:183-220.
- [39] Palfrey, T. R., H. Rosenthal, and N. Roy (2017), “How Cheap Talk Enhances Efficiency in Threshold Public Goods Games,” *Games and Economic Behavior*, 101: 234-59.
- [40] Passarelli F. and G. Tabellini (2017), “Emotions and Political Unrest,” *Journal of Political Economy*, 125(3):903-946.
- [41] Rob, R. (1989), “Pollution Claim Settlements with Private Information”, *Journal of Economic Theory*, 47, 307- 33.
- [42] Tavoni A., A. Dannenbergb, G. Kallisc, and A. Löschelb (2011), “Inequality, communication, and the avoidance of disastrous climate change in a public goods game,” *Proceedings of the National Academy of Science*, 108(29): 11825-11-829.

Online Appendix for:

Organizing for Collective Action: Olson Revisited

Marco Battaglini

Thomas R. Palfrey

Cornell University, NBER

California Institute of Technology, NBER

January 2024

Supplementary Online Appendix for:

Organizing for Collective Action: Olson Revisited

Proof of Lemma A2

We prove this by induction. The formula is correct for $j = 1$ since $\frac{B(m_n+1, n-1, p)}{B(m_n, n-1, p)} = \frac{(n-m_n-1)}{m_n+1} \cdot \frac{p}{1-p}$, and for $j = 2$ since:

$$\begin{aligned} \frac{B(m_n+2, n-1, p)}{B(m_n, n-1, p)} &= \frac{B(m_n+2, n-1, p)}{B(m_n+1, n-1, p)} \cdot \frac{B(m_n+1, n-1, p)}{B(m_n, n-1, p)} \\ &= \frac{(n-m_n-2)}{m_n+2} \frac{(n-m_n-1)}{m_n+1} \cdot \left(\frac{p}{1-p}\right)^2 \end{aligned}$$

The induction hypothesis is that the formula is correct for $j-1$:

$$\frac{B(m_n+j-1, n-1, p)}{B(m_n, n-1, p)} = \prod_{i=1}^{j-1} \frac{(n-m_n-i)}{m_n+i} \cdot \left(\frac{p}{1-p}\right)^{j-1}$$

This implies:

$$\begin{aligned} \frac{B(m_n+j, n-1, p)}{B(m_n, n-1, p)} &= \frac{B(m_n+j, n-1, p)}{B(m_n+j-1, n-1, p)} \cdot \frac{B(m_n+j-1, n-1, p)}{B(m_n, n-1, p)} \\ &= \frac{(n-m_n-j)}{m_n+j} \left(\frac{p}{1-p}\right) \prod_{l=1}^{j-1} \frac{(n-m_n-l)}{m_n+l} \cdot \left(\frac{p}{1-p}\right)^{j-1} \\ &= \prod_{i=1}^j \frac{(n-m_n-i)}{m_n+i} \cdot \left(\frac{p}{1-p}\right)^j \end{aligned}$$

so the formula is correct for j , which proves the claim. ■

Proof of Proposition 2

We proceed in two steps.

Step 1. To prove the first bullet point, let $m_n = \alpha n$ for some $\alpha \in (0, 1)$. From equation

(10), p_n^O is the unique solution to:

$$p_n^O = F \left[v \frac{1}{1 + \sum_{j=1}^{n-\alpha n} \frac{\alpha n}{j+\alpha n} \frac{B(\alpha n-1+j, n-1, p_n^O)}{B(\alpha n-1, n-1, p_n^O)}} \right] \quad (1)$$

and from Lemma A2, we have:

$$\begin{aligned} \sum_{j=1}^{n-\alpha n} \frac{\alpha n}{j+\alpha n} \frac{B(\alpha n-1+j, n-1, p_n^O)}{B(\alpha n-1, n-1, p_n^O)} &= \sum_{j=1}^{n-\alpha n} \frac{\alpha n}{j+\alpha n} \cdot \prod_{i=1}^j \frac{n-\alpha n+1-i}{\alpha n-1+i} \left(\frac{p_n^O}{1-p_n^O} \right)^j \\ &\leq \sum_{j=1}^{n-\alpha n} \frac{\alpha n}{j+\alpha n} \left(\frac{n-\alpha n}{\alpha n} \frac{p_n^O}{1-p_n^O} \right)^j \\ &\leq \sum_{j=1}^{n-\alpha n} \left(\frac{1-\alpha}{\alpha} \frac{p_n^O}{1-p_n^O} \right)^j \end{aligned} \quad (2)$$

Assume by contradiction that $\lim_{n \rightarrow \infty} p_n^O = 0$, then:

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{n-\alpha n} \left(\frac{1-\alpha}{\alpha} \frac{p_n^O}{1-p_n^O} \right)^j \leq \lim_{n \rightarrow \infty} \sum_{j=1}^n \left(\frac{1-\alpha}{\alpha} \frac{p_n^O}{1-p_n^O} \right)^j < \infty$$

since $\frac{1-\alpha}{\alpha} \frac{p_n^O}{1-p_n^O} \rightarrow 0$. Hence

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{n-\alpha n} \frac{\alpha n}{j+\alpha n} \prod_{i=1}^j \frac{n-\alpha n+1-i}{\alpha n-1+i} \left(\frac{p_n^O}{1-p_n^O} \right)^j < \infty \Rightarrow \lim_{n \rightarrow \infty} p_n^O > 0$$

a contradiction.

Step 2. To prove the second bullet point, let $m_n \prec n$. From Proposition 1, p_n^O is again the unique positive solution to (1) and from Lemma A2, we have (2). Assume by contradiction that $\lim_{n \rightarrow \infty} \frac{p_n^O}{\alpha n} < 1$, so $\lim_{n \rightarrow \infty} \frac{1-\alpha_n}{\alpha_n} \frac{p_n^O}{1-p_n^O} = \theta < 1$. In this case:

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{n-\alpha_n n} \left(\frac{1-\alpha_n}{\alpha_n} \frac{p_n^O}{1-p_n^O} \right)^j \rightarrow \sum_{j=1}^{\infty} (\theta)^j = \frac{\theta}{1-\theta} < \infty$$

Hence

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{n-\alpha_n n} \frac{\alpha_n n}{j+\alpha_n n} \prod_{i=1}^j \frac{n-\alpha_n n+1-i}{\alpha_n n-1+i} \left(\frac{p_n^O}{1-p_n^O} \right)^j < \infty$$

$$\Rightarrow \lim_{n \rightarrow \infty} p_n^O > 0 = \lim_{n \rightarrow \infty} \alpha_n \Rightarrow \lim_{n \rightarrow \infty} \frac{p_n^O}{\alpha_n} = \infty > 1$$

where $\lim_{n \rightarrow \infty} \alpha_n = 0$ follows from $m_n \prec n$. We have a contradiction and conclude that $\lim_{n \rightarrow \infty} \frac{p_n^O}{\alpha_n} \geq 1$. ■

Proof of Corollary 1

The fact that the probability of success in a VBO is infinitely higher than in an unorganized group for large n follows from part (2) of Theorem 1, that $P_n^U = 0$ for all $n > \bar{n}_U$, and part (1) of Proposition 2, that $c_n^O > 0$ for all n , so $P_n^O > 0$ for all n . Hence, $P_n^U / P_n^O = 0$ for all $n > \bar{n}_U$. Consider the claim that the probability of success is zero in the limit in the VBO (with constant returns). From equations (22) and (23) we have:

$$p_1(c_n^O) - p_2(c_n^O) = B(\alpha n - 1, n - 1, F(c_n^O)) \rightarrow_n 0$$

since all single terms of the binomial expansion converge to 0 as $n \rightarrow \infty$. Hence, equation (25) implies that $\lim_{n \rightarrow \infty} c_n^O a(c_n^O) = 0$, which in turn implies $\lim_{n \rightarrow \infty} a(c_n^O) = 0$ since $\lim_{n \rightarrow \infty} c_n^O = c^O > 0$. This implies $\lim_{n \rightarrow \infty} P_n^O = 0$. ■

Proof details for Theorem 3

Proof of Step 1:

To see this, first note that, for any $\eta > 0$ arbitrarily small:¹

$$\begin{aligned} \sum_{k=\alpha_n n + \eta\sqrt{n}}^{n-1} B(k, n-1, \alpha_n) &= \Pr(k \geq \alpha_n n + \eta\sqrt{n}) = \Pr\left(\frac{k}{n} - \alpha_n \geq \frac{\eta}{\sqrt{n}}\right) \\ &\leq \Pr\left(\left|\frac{k}{n} - \alpha_n\right| \geq \sigma_{\alpha_n}(k/n) \frac{\eta}{\sqrt{\alpha_n(1-\alpha_n)}}\right) \\ &\leq \lim \frac{\alpha_n(1-\alpha_n)}{\eta^2} \simeq \frac{\delta}{\eta^2} \cdot \frac{m_n}{n} \rightarrow 0 \end{aligned}$$

¹The initial term of the summation below should be written as $\lceil \alpha_n n + \eta\sqrt{n} \rceil$, since $\alpha_n n + \eta\sqrt{n}$ may not be an integer. To keep the notation simple, and without loss of generality since irrelevant for the argument, in the following we ignore this issue.

where $\sigma_{\alpha_n}(k/n) = \sqrt{\frac{\alpha_n(1-\alpha_n)}{n}}$ and in the last step we use Chebyshev's inequality. So we have:

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) &= \lim_{n \rightarrow \infty} \left[\sum_{k=\alpha_n n-1}^{(\alpha_n + \frac{\eta}{\sqrt{n}})^{n-1}} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) \right. \\ &\quad \left. + \sum_{k=\alpha_n n+\eta\sqrt{n}}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) \right] \\ &= \lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{(\alpha_n + \frac{\eta}{\sqrt{n}})^{n-1}} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) \end{aligned}$$

and $\lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{(\alpha_n + \frac{\eta}{\sqrt{n}})^{n-1}} B(k, n-1, \alpha_n) \geq \frac{1}{2}$. We can also write:

$$\begin{aligned} &\lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{(\alpha_n + \frac{\eta}{\sqrt{n}})^{n-1}} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n) \\ &\geq \left(\lim_{n \rightarrow \infty} \frac{1}{1 + \eta \cdot \frac{n^{1/2}}{m_N}} \right) \cdot \lim_{n \rightarrow \infty} \sum_{k=\alpha_n n-1}^{(\alpha_n + \frac{\eta}{\sqrt{n}})^{n-1}} B(k, n-1, \alpha_n) \geq \frac{1}{2} \end{aligned}$$

where in the last line we use the fact that $\alpha_n n = m_n \succ n^{1/2}$, so $n^{1/2}/m_n \rightarrow 0$.

Proof of Step 2:

Define $p_n^O = F(c_n^O)$ and assume by contradiction that $\lim_{n \rightarrow \infty} \frac{p_n^O}{\alpha_n} = \zeta$ for some $\zeta \geq 1$.

From the equilibrium condition we must have:

$$p_n^O = F \left[\frac{v B(\alpha_n n - 1, n - 1, p_n^O)}{\sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, p_n^O)} \right]$$

Note however that:

$$\begin{aligned} B(\alpha_n n - 1, n - 1, p_n^O) &= \binom{n-1}{\alpha_n n - 1} \frac{[(p_n^O)^{\alpha_n} (1 - p_n^O)^{1-\alpha_n}]^n}{p_n^O} \\ &= \frac{\alpha_n}{p_n^O} \binom{n}{\alpha_n n} [(p_n^O)^{\alpha_n} (1 - p_n^O)^{1-\alpha_n}]^n = \frac{1}{\zeta} \cdot \sqrt{\frac{1}{2\pi\alpha_n(1-\alpha_n)n}} [\xi_n]^n \end{aligned}$$

with $\xi_n = \frac{(p_n^O)^{\alpha_n} (1-p_n^O)^{1-\alpha_n}}{(\alpha_n)^{\alpha_n} (1-\alpha_n)^{1-\alpha_n}} \leq 1$ for any n (since for any p_n^O , $(p_n^O)^{\alpha_n} (1-p_n^O)^{1-\alpha_n} \leq (\alpha_n)^{\alpha_n} (1-\alpha_n)^{1-\alpha_n}$). Note, moreover, that

$$\sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, p_n^O) \geq \sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n-1, \alpha_n),$$

since $p_n^O > \alpha_n$ for n large enough. It follows that, using Lemma A1 and Step 1 above, we have that for large n :

$$\begin{aligned} 1 &= \frac{1}{p_n^O} F \left[\frac{vB(\alpha_n n - 1, n - 1, p_n^O)}{\sum_{k=\alpha_n n - 1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n - 1, p_n^O)} \right] \simeq \frac{vf(0)}{p_n^O} \frac{B(\alpha_n n - 1, n - 1, p_n^O)}{\sum_{k=\alpha_n n - 1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n - 1, p_n^O)} \\ &\leq \frac{vf(0)}{p_n^O} \frac{B(\alpha_n n - 1, n - 1, p_n^O)}{\sum_{k=\alpha_n n - 1}^{n-1} \frac{\alpha_n n}{1+k} B(k, n - 1, \alpha_n)} \simeq \frac{vf(0)(\xi_n)^n}{\zeta \cdot a_\infty} \cdot \frac{1}{\alpha_n} \sqrt{\frac{1}{2\pi\alpha_n(1-\alpha_n)n}} \rightarrow 0 \end{aligned}$$

where in the second step (“ \simeq ”) we used the fact that $p_n^O = \zeta\alpha_n$, and in the last step (“ \rightarrow ”), we used the fact that $\frac{1}{\alpha_n} \sqrt{\frac{1}{2\pi\alpha_n(1-\alpha_n)n}} \rightarrow 0$ since $m_n \succ n^{2/3}$ and $(\xi_n)^n \leq 1$. This is a contradiction, implying that $\lim_{n \rightarrow \infty} \frac{p_n^O}{\alpha_n} = \lim_{n \rightarrow \infty} \frac{F(c_n^O)}{\alpha_n} < \vartheta$ for some $\vartheta < 1$.

Proof of Lemma 1

We show that for any ε , there is a n_ε such that for $n > n_\varepsilon$ we have:

$$p_n^*(0) - p_n^*(c_n^*) < \varepsilon$$

where $c_n^* = \sup \{c \in [0, 1] \mid a_n^*(c) > 0\}$. Since $p_n^*(c) - p_n^*(c_n^*) = 0$ for $c \geq c_n^*$, this implies the statement of Lemma 1. There are two cases to consider: (1) $\lim_{n \rightarrow \infty} c_n^* = 0$; (2) $\lim_{n \rightarrow \infty} c_n^* = c_\infty^* \in (0, v)$.

Case 1: $\lim_{n \rightarrow \infty} c_n^* = 0$. By IC, we have:

$$\begin{aligned} vp_n^*(c_n^*) - a_n^*(c_n^*)c_n^* &\geq vp_n^*(0) - a_n^*(0)c_n^* \\ &\Leftrightarrow v[p_n^*(0) - p_n^*(c_n^*)] \leq [a_n^*(0) - a_n^*(c_n^*)]c_n^* \rightarrow 0 \end{aligned}$$

It follows that for any $\varepsilon > 0$, there is a n_ε such that $p_n^*(0) - p_n^*(c_n^*) < \varepsilon$ for $n > n_\varepsilon$.

Case 2: $\lim_{n \rightarrow \infty} c_n^* = c_\infty^* \in (0, v)$. We first show that for any $\varepsilon > 0$ and for every $\delta \in (0, c_\infty^*)$, there is a $n_{\varepsilon, \delta}$ such that $p_n^*(\delta) - p_n^*(c_\infty^*) < \frac{\varepsilon}{2}$ if $n > n_{\varepsilon, \delta}$.

Suppose by contradiction that this is not true, then, for some $\delta \in (0, c_\infty^*)$, it must be that $p_n^*(\delta) - p_n^*(c_\infty^*) > \varepsilon$ for all n . Define $\tau_{0, n-1}$ as the probability that there is at least one

member out of $n - 1$ with cost $c > c_n^*$; similarly let $\tau_{1,n-1}$ denote the probability that there is at least one member out of $n - 1$ with cost $c < \delta$. Denote by P_0^n, P_1^n, P_B^n the probability of success conditioning on, respectively, the presence of a type $c \leq \delta$ and no type $c \geq c_n^*$; the presence of a type $c \geq c_n^*$ and no type $c \leq \delta$; and the presence of both a type $c \leq \delta$ and a type $c \geq c_n^*$. Then, using this notation, we have:

$$\begin{aligned} E[p_n^*(c)|c \leq \delta] &= \tau_{0,n-1}P_B^n + (1 - \tau_{0,n-1})P_0^n \\ p_2 &= \tau_{1,n-1}P_B^n + (1 - \tau_{1,n-1})P_1^n \end{aligned}$$

and furthermore we have: $p_n^*(\delta) \leq E[p_n^*(c)|c \leq \delta]$ and $p_n^*(c_\infty^*) \geq p_2$. So we have:

$$\begin{aligned} 0 \leq p_n^*(\delta) - p_n^*(c_\infty^*) &\leq E[p_n^*(c); c \leq \delta] - p_2 \\ &\leq (\tau_{0,n-1} - \tau_{1,n-1})P_B^n + (1 - \tau_{0,n-1})P_0^n - (1 - \tau_{1,n-1})P_1^n \end{aligned}$$

As $n \rightarrow \infty$, both $\tau_{0,n-1}$ and $\tau_{1,n-1}$ converge to 1. Since P_0^n , P_1^n and P_B^n are all bounded, we have that for any $\varepsilon > 0$ and $\delta > 0$, there is a $n_{\varepsilon,\delta}$ such that $p_n^*(\delta) - p_n^*(c_\infty^*) < \frac{\varepsilon}{2}$ for all $n > n_{\varepsilon,\delta}$.

Finally, by incentive compatibility: $v[p_n^*(0) - p_n^*(\delta)] \leq [a_n^*(0) - a_n^*(\delta)]\delta \leq \delta$. Hence, if we set $\delta/v = \varepsilon/2$:

$$p_n^*(0) - p_n^*(v\frac{\varepsilon}{2}) \leq \frac{\varepsilon}{2}. \quad (3)$$

Furthermore, for any ε , there is a n_ε such that:

$$n > n_\varepsilon \implies p_n^*(v\frac{\varepsilon}{2}) - p_n^*(c_\infty^*) \leq \frac{\varepsilon}{2} \quad (4)$$

Combining (3) and (4) implies that for any ε , there is a n_ε such that for $n > n_\varepsilon$: $p_n^*(0) - p_n^*(c_\infty^*) \leq \varepsilon$. ■

Proof of Lemma A5

Proof. We first prove that, for any sequence ε_l such that $\varepsilon_l \rightarrow 0$ as $l \rightarrow \infty$ and $\varepsilon_l \leq \varepsilon_{l-1}$, we have:

$$\lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} = \lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^{\varepsilon_l}$$

To this goal define for convenience $\tilde{V}_n^j = \tilde{V}_n^{\varepsilon_j}$ and:

$$a_n^j = \tilde{V}_n^{j-1} - \tilde{V}_n^j \text{ with } \tilde{V}_n^0 = 0,$$

noting that by construction $a_n^j \geq 0$. We can write:

$$\lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^l = - \lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \sum_{j=1}^l a_n^j$$

since $\tilde{V}_n^j = - \sum_{j=1}^l a_n^j$. So we have:

$$\lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^j = - \lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \sum_{j=1}^l a_n^j = - \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} a_n^j = - \lim_{n \rightarrow \infty} \int a_n$$

where the $\int a_n = \sum_{j=1}^{\infty} a_n^j = \lim_{l \rightarrow \infty} \sum_{j=1}^l a_n^j$ is the Lebesgue integral of a_n with respect to the counting measure.

Define now the sequence $(g^j)_{j=0}^{\infty}$ such that $g^j = \sup_n a_n^j$. Note that $g^j \geq a_n^j \geq 0$ for all n by definition. Moreover, g^j is integrable with respect to the counting measure since a_n^j is integrable and integrability is passed by the *sup* operator. Finally, we have

$$\int g_n = \lim_{l \rightarrow \infty} \sum_{j=1}^l g^j = \lim_{l \rightarrow \infty} \sum_{j=1}^l \sup_n a_n^j = \lim_{l \rightarrow \infty} \sup_n \sum_{j=1}^l a_n^j = \lim_{l \rightarrow \infty} \sup_n \tilde{V}_n^l \leq v$$

where v is the value of the collective good. We conclude that g^j is an integrable function that dominates a_n^j for any n . We can therefore apply the Dominated Convergence Theorem

as follows:

$$\begin{aligned}\lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^l &= - \lim_{n \rightarrow \infty} \int a_n = - \int \lim_{n \rightarrow \infty} a_n \\ &= - \lim_{l \rightarrow \infty} \sum_{j=1}^l \lim_{n \rightarrow \infty} a_n^j = - \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \sum_{j=1}^l a_n^j = \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{V}_n^l\end{aligned}$$

So we have $\lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^l = \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{V}_n^l$.

Now note that:

$$\begin{aligned}\lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} &= \lim_{n \rightarrow \infty} V_n^{**} \Leftrightarrow \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} = \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} V_n^{**} \\ &\Leftrightarrow \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} = \lim_{n \rightarrow \infty} V_n^{**} \geq \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon(n, \eta)}\end{aligned}$$

since we have proven in Lemma 2 that the optimal (HO) mechanism becomes approximately flat for large n , so the constraint becomes slack for n large enough. Moreover, $\tilde{V}_n^{\varepsilon(n, \eta)} \geq \lim_{l \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} = \tilde{V}_n^0$. It follows that:

$$\begin{aligned}\lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} &\leq \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon(n, \eta)} \leq \lim_{l \rightarrow \infty} \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} = \lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} \\ &\Leftrightarrow \lim_{n \rightarrow \infty} \tilde{V}_n^{\varepsilon(n, \eta)} = \lim_{n \rightarrow \infty} \lim_{l \rightarrow \infty} \tilde{V}_n^{\varepsilon_l} = \lim_{n \rightarrow \infty} \tilde{V}_n^0\end{aligned}$$

It follows that for any $\epsilon > 0$, there is an n_ϵ such that for $n > n_\epsilon$, we have $\tilde{V}_n^{\varepsilon(n, \eta)} \leq \tilde{V}_n^0 + \epsilon$, which proves the result. \blacksquare