Organizing for Collective Action: Olson Revisited

Marco Battaglini

Cornell University

Thomas R. Palfrey

California Institute of Technology

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.

I. Introduction

Collective action is one of the most basic and ubiquitous forms of strategic interaction in societies. Examples of collective action problems range

Support from the National Science Foundation (SES-2343948) is gratefully acknowledged. This paper has benefited from comments and suggestions by the editor and participants of the 2021 ETH Workshop on Theoretical Political Economy, the 2021 Stanford Institute for Theoretical Economics Workshop on Political Economic Theory, the 2023 Political Economy UK Group, the 2023 Conference on Economic Theory at the Cowles Foundation, and

Electronically published Month XX, 2024

Journal of Political Economy, volume 132, number 9, September 2024.

https://doi.org/10.1086/729580

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 joining 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 significant levels of individual participation that are hard to explain, other than 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 have not yet been fully explored in the literature, as we will argue. 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 with 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 for making 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 m_n out of n agents

the 2023 European Summer Symposium in Economic Theory, as well as seminar audiences at Bocconi University, the California Institute of Technology, Columbia University, CUNEF Universidad, Oxford University, Stanford University, Technische Universität Dortmund, Universidad Carlos III Madrid, the University of California, Berkeley, the University of Bern, the University of Essex, the University of Padova, and the University of Zurich. We are especially grateful to the referees for their detailed comments and suggestions. We also 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. This paper was edited by Emir Kamenica.

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 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 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 very 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 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,

[1] Classic contributions are Palfrey and Rosenthal (1984) in economics and Diekmann (1985) in sociology, who coined the term "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, e.g., Tavoni et al. (2011) and Barrett et al. (2014). Recent contributions in economics include Battaglini and Benabou (2003), Battaglini (2017), Bergstrom (2017), Nöldeke and Peña (2020), and Dziuda, Gitmez, and Shadmeher (2021), among others.

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

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 three 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 one using a simple random mechanism:[3] randomly and anonymously select three 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 (HO) mechanisms was not especially important for limiting results with large groups in the early literature that assumed constant returns to scale; that is, 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

[3] This mechanism is also optimal under some weak conditions, as we see in sec. III. However, this fact is irrelevant for the present discussion.

[4] This 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.

of success for a large enough n even if we adopt a simple random mechanism with no transfers as outlined before. However, such mechanisms violate the obedience constraint, as explained in the example above. It therefore becomes important to understand what can be achieved with an HO mechanism.

Our second theoretical contribution is to show that a simple class of HO mechanisms that we call volunteer-based organizations (VBOs) 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 that 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 HO organizations. This allows us to extend the negative limit results of the previous literature—that is, that the limit probability of success in an optimal IC and IR mechanism with constant returns to scale is zero, both with an organized and with an unorganized group (Rob 1989; Mailath and Postlewaite 1990; Ledyard and Palfrey 1994, 1999, 2002). We show that with HO mechanisms the limiting probability of success is the same with an organized and with an unorganized group for *any* rate of growth of m_n: as with the Bayesian Nash equilibrium for unorganized groups, the optimal HO mechanism achieves a limiting success probability equal to one if m_n grows at a rate slower than $n^{2/3}$, and it achieves a limit probability equal to zero 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 strong mechanisms are successful with probability one even with no transfers, as we will show.

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 an HO 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 with 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 $n \to \infty$ 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 HO 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 VI. That analysis suggests that even when successful collective action is possible only 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 HO 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 IR constraints, ignoring the obedience constraint (Mailath and Postlewaite 1990; Ledyard and Palfrey

[5] Other 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. For important works on these dimensions, see Chamberlin (1974) and more recently Esteban and Ray (2001).

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 one 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 an HO mechanism, results are very different, both with and without constant returns. Allowing for increasing returns, we extend the insight that organizations are not useful in the limit, since we show with HO mechanisms that they can obtain the same limit probabilities of success as unorganized groups only as $n \to \infty$; 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—that is, $vn > m_n\bar{c}$, where $\bar{c}$ represents 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, using both case studies (e.g., Ostrom 1990) and laboratory experiments (van de Kragt et al. 1983; Braver and Wilson 1986; Ostrom and Walker 1991; Palfrey and Rosenthal 1991; Ostrom, Walker, and Gardner 1992; Palfrey, Rosenthal, and Roy 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

[6] Other research on Bayesian mechanism design with public goods analyzes "superstrong" organizations that require IC but allow for unlimited side payments and no participation constraints (d'Aspremont and Gerard-Varet 1979; Crémer and McLean 1988; d'Aspremont, Crémer, and Gérard-Varet 1990; Ledyard and Palfrey 1999, 2002).

[7] The 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, e.g., Healy (2010), Bierbrauer and Hellwig (2016), Bierbrauer and Winkelmann (2020), and Goldlucke and Troger (2020).

coercion, and they report the endogenous emergence of mechanisms similar to the VBO mechanism that we show to be asymptotically optimal.

II. The Collective Action Model

A. Setup

A group with n members, $I = \{1, 2, \ldots, n\}$, desires an outcome generating a total value of W_n, with each member in the group receiving a personal, direct benefit of $v = W_n/n \in (0, 1)$ that 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 independent and identically distributed (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 that $0 < f(c) < \bar{f}$ 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 nondecreasing with $m_n \to \infty$ as $n \to \infty$.[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 $m_n/n \to 0$ 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.

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

[9] The analysis directly extends to more general environments. We discuss alternative assumptions about the cost distribution in sec. VII.

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

The second form is an organized group. We are interested in studying the benefits from 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, cannot exert any form of coercion on the members' actions, and cannot even commit to monetary transfers. We refer to such organizations as *weak organizations*. The organized group can design only 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 HO direct communication mechanisms.[11] In section II.B, we provide a formal characterization of this class of mechanisms and its relationship to the class of IC and IR mechanisms.

In section VI, we describe a stronger form of organization in which only IC and interim IR 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; Hellwig 2003). We refer to these as *strong organizations*.

B. Weak Organizations: HO 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, $\mu : [0, 1]^n \to \Delta(2^I)$, 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 that 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 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 ith component, g_i, is equal to one if $i \in g$ and equal to zero 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 $|g| = \Sigma_i g_i$.

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

Define $I_i = \{g \subseteq I | i \in g\}$ as the subsets of I containing i, and define $I^{m_n} = \{g \subseteq I | |g| \geq m_n\}$ as the set of subsets containing at least m_n members. Given an activity function, μ, the probability that i is active at type profile $\mathbf{c}$ is given by $A_i(\mathbf{c}; \mu) = \Sigma_{g \in I_i} \mu_g(\mathbf{c})$, and the probability that enough members are active that the group is successful is given by $P(\mathbf{c}; \mu) = \Sigma_{g \in I^{m_n}} \mu_g(\mathbf{c})$. 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 that i is active, conditional on i's cost. We assume without loss of generality that the mechanism is symmetric—that is, 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 write these reduced-form functions as simply $p: [0, 1] \to [0, 1]$ and $a: [0, 1] \to [0, 1]$. 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 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 $|g| \geq m_n$ 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 that

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

[12] To see that the restriction to symmetric mechanisms is without loss of generality, consider any HO asymmetric mechanism, μ. For any permutation, ρ, of the member indexes, 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 HO, and it generates the same total surplus as μ.

for any $i = 1, \ldots, n$, $c, c' \in [0, 1]$ 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 (IC) condition:

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

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

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

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

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

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

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

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

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

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 see in sec. III.B).[13]

III. 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

[13] An 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 sec. III.B, where we show conditions under which the optimal IC and INTIR mechanism is EXIR but fails IMH and HO.

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 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.

A. Unorganized Groups

For an unorganized group, the payoff function and distribution of costs described above define a Bayesian game where each member simultaneously chooses whether to be active. 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.

1. Equilibrium with Unorganized Groups

Denote by p the probability that 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 cut point, $\hat{c}_n^U(p)$, 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 cut point must satisfy

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

where $B(m_n - 1, n - 1, p) \equiv \binom{n-1}{m_n-1} (p)^{m_n-1}(1-p)^{n-m_n}$ represents the probability of being pivotal. In equilibrium, it must be that p coincides with the probability that 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 cut point:

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

An equilibrium exists, trivially, because $c_n^U = 0$ is always a solution to equation (3) when $m_n > 1$. It is possible that 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 simply 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 that an unorganized group is successful is

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

It is relatively straightforward to see that in the case with constant returns to scale—that is, $m_n = \alpha n$ 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 $(1 - c_n^U)/(vc_n^U)$ and substitute $\alpha n = m_n$ to obtain

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

The left-hand side of equation (5) is greater than or equal to $(1 - v)/v$ for all n, and the right-hand side converges to $((1 - \alpha)/\alpha)B(\alpha n - 2, n - 1, c_n^U)$, which converges to zero. 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]

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

[14] This is stated and proved more generally as theorem 1 in the next section.

[15] The 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

see that the answer is yes. From equation (3), a cut point 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 $\lim_{n\to\infty} c_n^U = 0$. The probability of success, from equation (4), is

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

so $\lim_{n\to\infty} P_n^U = 1 - ((\lim_{n\to\infty} c_n^U)/v) = 1$. QED

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 (respectively, faster) than a sequence s_n—that is, $m_n \prec s_n$ (respectively, $m_n \succ s_n$) if $m_n/s_n \to 0$ (respectively, $m_n/s_n \to \infty$). We say that m_n grows at the same speed as s_n—that is, $m_n \simeq s_n$ if $m_n/s_n \to \rho$, 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 one in the limit as $n \to \infty$ 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 $v \in (0, 1)$, the following hold:

1) If $m_n \prec n^{2/3}$, then $\lim_{n\to\infty} P_n^U = 1$.
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 section A1.

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—that is, $m_n \succ n^{2/3}$—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 with no organization, the group can achieve a limit success probability of one, 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 line in the figure shows the left-hand side of equation (3), and the two single-peaked curves show the right-hand side of equation (3),

case. The argument presented here generalizes to the case in which m_n is constant and equal to any integer $M > 1$.

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

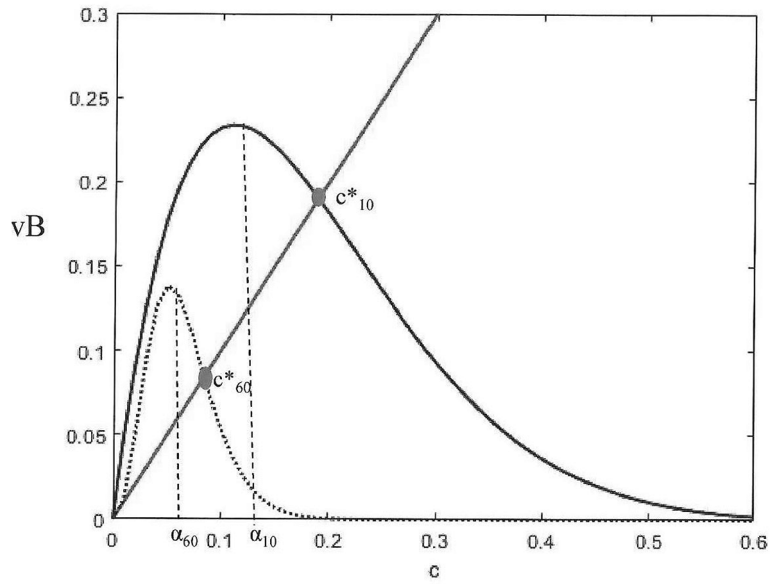

FIG. 1.—Intuition for equilibria with positive limit probability of success in unorganized groups: $v = 0.6$; $n = 10, 60$; $m_n = \lceil 0.5n^{.6} \rceil$.

for $v = 0.6$, $m_n = 0.5n^{0.6}$, 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 one as long as 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 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° line and thus higher than $F^{-1}(\alpha_n)$—that is,[17]

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

[17] The 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 (respectively, decreasing) in c for $c < (\alpha_n n - 1)/(n - 1)$ (respectively, $c > (\alpha_n n - 1)/(n - 1)$).

When this is the case, 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 $n \to \infty$, $c_n^U \to 0$, 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

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

where $\preceq$ here means that the right-hand side converges to zero at the same or a slower rate than the left-hand side. A necessary condition for (7), therefore, is that 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 one. 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 zero for n sufficiently large.

B. *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 IC and INTIR (see Rob 1989; Mailath and Postlewaite 1990; Ledyard and Palfrey 1994, 1999). We refer to a mechanism with transfers requiring IC and INTIR as a strong organization, since in both cases they need to satisfy weaker constraints than in the weak organization defined in section II.B 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 $t(c) = E_{\mathbf{c}_{-i}}[t^i(c, \mathbf{c}_{-i})]$:

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

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

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

To solve (9), consider a relaxed version in which we ignore the (INTIR) constraint. In appendix section A2, 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, $-[(1 - F(c))/f(c)]$. 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 nonincreasing, the best way to do it satisfying feasibility is to keep $a(c)$ and (by IC) $p(c)$ 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 $(m_n/n) \cdot E(c) < v$, we have $p_n = 1$, and a_n is chosen to be the smallest possible, so $a_n = m_n/n$ 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 $(m_n/n) \cdot 1 < v$, 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 ($n \geq n^* = \min\{n|(m_n/n) < v\}$).

THEOREM 2. For all $v \in (0, 1)$, the following hold:

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 one. 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 one. If F satisfies MHRA, the random mechanism is optimal.

Proof. See appendix section A2.

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 $v(1 - \alpha_n(c/v)) > 0$ even if $c > v$; but this guarantees only *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 HO mechanisms: by ignoring moral hazard, the optimal mechanism achieves complete success as $n \to \infty$. To understand why collective action can be only partially successful in more realistic environments, we need to integrate the obedience constraint into the analysis of the optimal mechanism.

Note that the MHRA 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

[18] The 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

assumption that the total benefit of success, nv, is strictly lower than the cost of obtaining it in the worst-case scenario where all members have cost equal to one.[19] In our setting, this assumption reduces to $v < m_n/n = \alpha_n$, which is not satisfied for large n when $m_n \prec n$ or for the case of constant returns if $v \in (\alpha, 1)$. Requiring that $v < \alpha_n$ for all n is essentially the same as assuming constant returns to scale with $\alpha = \liminf_{n\to\infty}\alpha_n$. Indeed, there are many situations in which it is natural to assume that $v > \alpha_n$, such as situations where the "sacrifice" of a small share of population is all that is needed to guarantee group success. As we see in the next two sections, with HO mechanisms, $v > \alpha_n$ does *not* imply success even though strong organizations with sufficiently many members will succeed with probability one. The failure or success of collective action in HO mechanisms with large n depends 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.

IV. The VBO

As discussed in the previous section, strong mechanisms that require only 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 predicts complete failure otherwise. The result for the other benchmark, unorganized groups, is as negative as the results from strong organizations are positive. 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?

---

group benefits from a strong organization are achieved via ex post coercion—by sacrificing the sovereignty of individual high-cost members of the group.

[19] Formally, 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 $nv + n\epsilon < n\alpha_n$.

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 HO, members self-identify as either *volunteers* (low cost) or *free riders* (high cost) and the mechanism coordinates the activity of volunteers 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.

A. *Unique HO 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 the case that the group activates m_n volunteers, all volunteers have the same probability of being included. Using the notation introduced in section II.B, 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; and (3) $k(\mathbf{c}; c_n^O) \geq m_n \Rightarrow P(\mathbf{c}) = 1$ and $A_i(\mathbf{c}) = m_n/(k(\mathbf{c}; c_n^O))$ for all i such that $c_i \leq c_n^O$.

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

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

PROPOSITION 1. For any v, m_n, and n the following hold:

1) The function Y_n is strictly decreasing in c and has a unique fixed point $c_n^O \in (c_n^U, v)$.

[20] The definition requires that $c_n^O > 0$. If c_n^O were exactly equal to zero, 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.

2) A VBO is HO 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 section A3.

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 HO 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 there are indeed 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 $vB(m_n - 1, n - 1, F(c_n^O))$, essentially the same as in the unorganized group in (3). The critical difference for the organized group is that volunteers are called to action not 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 one 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 $F(c_n^O) > F(c_n^U)$.

B. *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 groups,

at least for some parameterizations. 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 HO VBO when n is large requires evaluating how the solution to condition (10), c_n^O, converges as $n \to \infty$. 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 HO VBO, for all $v < 1$ we have the following:

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

Proof. See the 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 with unorganized groups is *infinite* for large enough finite groups. The intuition for the fact that $\lim_{n\to\infty} c_n^O > 0$ 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 HO VBO and constant returns (i.e., $m_n = \alpha n$), 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 $P_\infty^O \equiv \lim_{n\to\infty} P_n^O = 0$.

Proof. See the online appendix.

The second bullet point of proposition 2 is that, when $m_n \prec n$, then, while the equilibrium participation rate converges to zero as $n \to \infty$, it does so at the same (or a slower) rate as the threshold fraction α_n, since $\lim_{n\to\infty}(F(c_n^O)/\alpha_n) \geq 1$. If $\lim_{n\to\infty}(F(c_n^O)/\alpha_n) > 1$, then obviously $P_\infty^O = 1$, but if $\lim_{n\to\infty}(F(c_n^O)/\alpha_n) = 1$, then it depends on exactly how $F(c_n^O)/\alpha_n$ converges to one. 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 arbitrarily 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 HO VBO, for all $v \in (0, 1)$ the following hold:

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

Proof. See appendix section A4.

As we 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 one, 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 point, $P_n^U/P_n^O = 0$. This can imply a significant benefit of adopting a simple VBO compared with having an unorganized group. As we 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 $\alpha_n \le F(c_n^O)$. Since $c_n^O \to 0$, we have $F(c_n^O) \simeq f(0)c_n^O$, implying that

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

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

$$a_n(\alpha_n) \lesssim vf(0) \cdot \frac{B(\alpha_n n - 1, n - 1, \alpha_n)}{\alpha_n}. \tag{11}$$

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. However, it is 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, which is the minimal requirement for success.

C. *The Value of a VBO with Finite n*

While the limit probability of success is the same for a VBO as for 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. Alternatively, $c_n^U = 0$ for finite-sized unorganized groups. As a result, a VBO performs infinitely better than an unorganized group, as shown above.

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 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 $P_n^*/e^{-\gamma n} \to \infty$ for any $\gamma > 0$. We have the following:

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 section A5.

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:

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

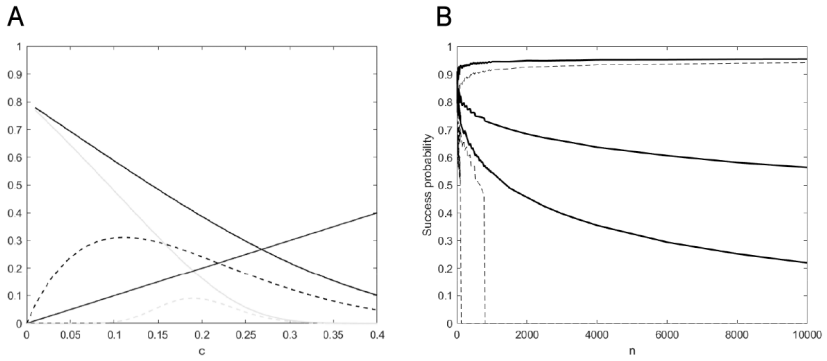

FIG. 2.—Comparison of VBO mechanism (solid lines) and unorganized group equilibrium (dashed lines). $v = 0.8$; $m_n = 0.2n^{\beta}$; $F(c)$ uniform. *A*, Equilibrium cut points. $\beta = 1$. Equilibrium cut points are located at the intersection of each curve ($n = 10$ [darker shade] and $n = 80$ [lighter shade]) with the diagonal. *B*, Probability of success. $n = 10, \ldots, 10{,}000$; $\beta = 0.65$ (top curve), 0.80 (middle curve), 0.85 (bottom curve).

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

$$Z_n(c_n^U, \alpha_n, v) = vB(m_n - 1, n - 1, c_n^U). \tag{13}$$

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

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

Figure 2*A* 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 $Y_{80}(c, 0.2, 08)$, corresponding to the right-hand side of equation (12) for the VBO. The respective equilibrium cut points, 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]

Figure 2*B* illustrates the equilibrium probability of success for organized (solid curves) and unorganized (dashed curves) groups for group sizes up to 10,000, with $v = 0.8, F \sim \mathcal{U}[0, 1]$, and $m_n = 0.2n^{\beta}$ for three values

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

of $\beta < 1$. One can see that for $\beta = 0.65 < 2/3$, both unorganized and organized groups can achieve success with probability that is high even for small groups and converges to one. For $\beta > 2/3$, the probability of success instead 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 a 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 $n \to \infty$, 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 zero.

V. Asymptotic Optimality of the VBO

The previous section illustrated that 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 HO 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 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 zero if $m_n \succ n^{2/3}$ and equal to one 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).

A. *Optimal HO Mechanisms*

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

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

[22] Whenever 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^*.

where (HO) is the honest and obedient constraint specified in section II.B. As discussed in section II, the (HO) constraint implies (IC), (INTIR), and condition (1). In the following, we first study the solution of a relaxed problem in which only (IC) and (1) are considered; we then prove that this solution satisfies the omitted constraints and thus solves (14). By standard methods, one can rewrite the problem as

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

where p_2 represents the (constant) expected probability of success for all types $c > c^*$. In (15), we derived the objective function using the (IC) constraint in a way similar to that in (9). The constraint 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 IC, 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 above 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—that is, 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 and 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

[23] For details on how it follows from (1) and IC, see the proof of proposition 5.
[24] Indeed, $U(c) \ge U(c^*) \ge vp_2 \ge 0$ for all $c \in [0, c^*]$.

type with $c > v$ will ever accept to be activated even if it is 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 IC.

B. 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 $n \to \infty$. 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 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 $n \to \infty$. Third, we show that the optimal binary HO mechanism has a limiting probability of success equal to zero if $m_n \succ n^{2/3}$ and equal to one if $m_n \prec n^{2/3}$.

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 zero 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 zero 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 III.B 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) $k(c; c_n^G) = k_n^G \Rightarrow P(c) = 1$ and $A_i(c) = m_n/(k(\mathbf{c}; c_n^G))$ for all i such that $c_i \leq c_n^G$; and (4) $k(c; c_n^G) = k_n^G \Rightarrow P(c) = q_{k_n^G} \in (0, 1]$ and $A_i(\mathbf{c}) = q_{k_n^G}(m_n/(k(\mathbf{c}; c_n^G)))$ 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 HO binary mechanism.

Proof. The proof is carried out in two steps. In step 1, we establish that the optimal binary mechanism is nonwasteful, meaning that it activates only zero 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. For details, see appendix section A6. QED

One might conjecture that the unique VBO characterized by equation (10) would be the optimal HO binary mechanism—that is, 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 though 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., $k_n^G = m_n + 1$) 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.

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, only 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 members (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 HO mechanism. For any two types c' and c'' with $c'' > c' > 0$, $p_n^*(c') - p_n^*(c'') \to 0$ and $a_n^*(c') - a_n^*(c'') \to 0$ as $n \to \infty$.

Proof. See the online appendix.

To see the intuition of this result, note that by IC, $p_n^*(c)$ is nonincreasing 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 $p_n^*(c \geq c'')$ represent the interim probabilities of success conditioning on $c \leq c'$ and $c \geq c''$, respectively. Then we have

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

where $\tau_{0,n-1}$ represents the probability that, out of the remaining $n - 1$ agents, there is at least one type $\tilde{c} \geq c''$; P_B^n represents 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 represents 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 $\tau_{1,n-1}$ represents the probability that, out of the remaining $n - 1$ agents, there is at least one type $\tilde{c} \leq c'$ and P_1^n represents the expected probability of success conditioning on the presence of at least a type $\tilde{c} \geq c''$ and the absence of a type $\tilde{c} \leq c'$. But then we have

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

As $n \to \infty$, both $\tau_{0,n-1}$ and $\tau_{1,n-1}$ converge to one. Since P_0^n, P_1^n, and P_B^n are all bounded, we have that for any $\varepsilon > 0$ there is an 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 nonincreasing function that 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^* represent the expected

welfare generated in, respectively, the optimal binary HO mechanism (generalized VBO) and the optimal HO mechanism when the number of agents is n. Putting this all together, we have the following:

PROPOSITION 5. $\lim_{n\to\infty} V_n^G = \lim_{n\to\infty} V_n^*$.

Proof. See appendix section A7.

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* HO mechanism. We 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.

3. The Irrelevance of Organizations in the Limit as $n \to \infty$

In our earlier analysis of the simple VBO mechanism (theorem 3), we proved that in the limit large groups succeed with probability one if $m_n \prec n^{2/3}$ and large groups fail with probability one if $m_n \succ n^{2/3}$. However, that left open the question of whether the optimal HO 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 HO mechanism.

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

THEOREM 4. For any $v \in (0, 1)$, the following hold:

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

Proof. The first bullet point follows immediately, since the VBO achieves success with probability one 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. For details, see appendix section A8. QED

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 IV.C, 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—that is, in the limit. If mechanisms must satisfy HO 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.

VI. 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

[25] See the discussion in chaps. 1 and 2 of Olson (1965).

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

On the other hand, whether the group might become organized depends on the underlying fundamentals of the economy and 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 that 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 respectively represent the expected utilities with and without an organization and $\kappa(n)$ represents the per-person cost of forming and operating an organization with these n agents. We assume that $\kappa(2) > 0$, and $\lim_{n\to\infty}\kappa(n) = \kappa > 0$.[26]

The expected utility in a VBO with cutoff c_n^O can be written as

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

where $p_{1,n}(c_n^O)$ and $p_{2,n}(c_n^O)$ represent the probability of success for an agent, conditioning on being a volunteer and not being a volunteer, respectively; and $E(c; c')$ represents 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 utility with an optimal organization can therefore be written as

[26] This simple model is intended to 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.

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

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

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

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 brackets 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 $c_n^O \to 0^+$, we deduce that no organization will ever be formed for arbitrarily large groups:[27]

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

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 the following:

Proposition 7. There is an $n^* > 0$ 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 section A9.

Propositions 6 and 7 are interesting because they suggest why the two factors highlighted by Olson (size and individual incentives) matter for a group's effectiveness. It is not simply 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 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 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 the large, unorganized, and ineffective groups on the other hand. What

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

[28] The 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.

creates the dichotomy is the decision to organize that transforms a continuous effect in a discrete drop in effectiveness.

VII. Variations and Discussions

A. *On High-Value Environments ($v \geq 1$)*

An assumption that we have maintained throughout the analysis is that $v < 1$, where one 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 one, so that $v < \alpha$ (see, e.g., 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 zero. The probability of success equals one.

When $v \geq 1$, however, the limit probability of success in the symmetric equilibrium of an unorganized group remains zero 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 only strategic problem faced by the members is coordination. Coordination can be easily solved by an HO mechanism but is unsolvable in a symmetric equilibrium without an organization.[29]

B. *On the Divisibility of Tasks*

In the previous analysis, we 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, signs a petition or not, joins a union or a committee or not. However, there are cases in which the contribution can be split up. For example, suppose that an agent has up to 1 day to donate

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

to a cause—say, the organization of a charity. However, if the agent cannot donate 1 day, perhaps the agent can donate less—say, 1 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 now a contribution can be divided into λ parts: when $\lambda = 1$, the contribution is, say, 1 day; when $\lambda = 24$, it is 1 hour; and so on. Now the mechanism can ask each agent to contribute any discrete amount $x \in \{0, 1/\lambda, 2/\lambda, \ldots, 1\}$—say, from 0 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 possibly be larger than one; now we can require m_n agents providing one unit or up to λm_n agents providing 1 hour. If we can choose λ so large that $\bar{c}/\lambda < v$ and $\lambda m \leq n$, 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 simply 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 one 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 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 simply 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 nonincreasing 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 is nonincreasing in c. The analysis is completely analogous. Indeed, the same logic as in section V suggests that, for finite n, $a(c)$ will be nonincreasing 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 $n \to \infty$, 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.

C. *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—that is, $f(0) > 0$. 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 $c \to 0$.

A second assumption is that $F(0) = 0$—that is, 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—that is, 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 $c_L/v < (1 - (\phi/\alpha_n))/(1 - \phi)$, where $\phi \in (0, 1)$ is the probability of the low type.

VIII. 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 IC and IR. Mechanisms that are only IC and IR 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 map vectors of type profiles only to recommendations: allocations are the decentralized results of the members' individual actions.

Strong mechanisms that satisfy interim IC and IR 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 HO 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 HO 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 chooses not only to build a bridge but also the bridge's quality and 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, e.g., Matthews 2013; 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 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 fieldwork. 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

A1. 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 $\lim_{n\to\infty} P_n^U = 1$. 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 $B(\alpha_n n - 1, n - 1, \alpha_n)/\alpha_n \to \infty$; if $m_n \succ n^{2/3}$, then $B(\alpha_n n - 1, n - 1, \alpha_n)/\alpha_n \to 0$.

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

$$\begin{aligned}\binom{n-1}{\alpha n - 1} &= \frac{(n-1)!}{(\alpha_n n - 1)!\,(n - \alpha_n n)!} = \frac{\alpha_n n}{n}\frac{n!}{\alpha_n n \cdot (\alpha_n n - 1)!\,(n - \alpha_n n)!} \\ &= \alpha_n \binom{n}{\alpha_n n}.\end{aligned}$$

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. Thus, an approximation of $B(\alpha_n n - 1, n - 1, \alpha_n)$ is given by

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

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 $B(\alpha_n n - 1, n - 1, \alpha_n)/\alpha_n \to \infty$ if $(m_n/n)^3(1 - (m_n/n))n \to 0$ as $n \to \infty$, a condition satisfied if $m_n/n^{2/3} \to 0$ or $m_n \prec n^{2/3}$; we have $B(\alpha_n n - 1, n - 1,$

$\alpha_n)/\alpha_n \to 0$ if $(m_n/n)^3(1 - (m_n/n))n \to \infty$ as $n \to \infty$, a condition satisfied if $m_n/n^{2/3} \to \infty$ or $m_n \succ n^{2/3}$. QED

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 $n \to \infty$, $vB(\alpha_n n - 1, n - 1, \alpha_n) \to 0$, 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 $o(B(\alpha_n n - 1, n - 1, \alpha_n))/\alpha_n \to 0$ as $n \to \infty$. It follows that

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

We conclude that whenever $B(\alpha_n n - 1, n - 1, \alpha_n)/\alpha_n$ converges to zero or diverges at ∞, so does $F[vB(\alpha_n n - 1, n - 1, \alpha_n)]/\alpha_n$. This implies that when $m_n \prec n^{2/3}$, then $F[vB(\alpha_n n - 1, n - 1, \alpha_n)]/\alpha_n \to \infty$, implying that $\alpha_n < F[vB(\alpha_n n - 1, n - 1, \alpha_n)]$, 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 one. We proceed in two substeps.

Step 2.1.—Assume first that $F(c_n^U)/\alpha_n \to 1$. We have

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

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

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

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

Step 2.2.—It follows from step 2.1 that we can assume that $F(c_n^U)/\alpha_n \to \zeta > 1$. 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

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

where in the second line we used $\alpha_n/F(c_n^U) < 1$, so $(1 - (\alpha_n/F(c_n^U))) \ge \vartheta$ for some $\vartheta > 0$; in the fourth line we define $\sigma_{c_n^U}(k/n) = \sqrt{F(c_n^U)(1 - F(c_n^U))}/\sqrt{n}$ and used Chebyshev's inequality; and in the last step of the fourth line ($\to$), we used the fact that $n\alpha_n = m_n \to \infty$.

Step 3.—We now show that if $m_n \succ n^{2/3}$, then $\lim_{n\to\infty} P_n^U = 0$. We first establish that $\lim_{n\to\infty}(p_n^U/\alpha_n) = 0$. Assume not, so $\lim_{n\to\infty}(p_n^U/\alpha_n) = \zeta$ 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

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

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

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

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 $(1/\alpha_n)\sqrt{1/(2\pi\alpha_n(1-\alpha_n)n)} \rightarrow 0$ since $m_n \succ n^{2/3}$ and $(\xi_n)^n \leq 1$. This is a contradiction, implying that $\lim_{n\to\infty}(p_n^U/\alpha_n) = 0$. We next use this to show that $p_n^U = 0$ for large n. By definition, we have

$$p_n^U = F(vB(\alpha_n n - 1, n - 1, p_n^U)) = \Psi(\alpha_n, n, p_n^U). \tag{A1}$$

Note that since we do not 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 an 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

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

where the equilibrium condition (A1) 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 $p_n^U/\alpha_n \rightarrow 0$ 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 $p_n^U = 0$. QED

A2. 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 one. 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

$$\begin{gathered}
\max_{p(\cdot),a(\cdot)} \left\{ vp(0) - E\left[a(c)\cdot\frac{1 - F(c)}{f(c)}\right]\right\} \\
\text{s.t. } a(c) \text{ is nonincreasing with } a(c) \in [0, 1] \\
\text{and } p, a \text{ feasible}
\end{gathered} \tag{A2}$$

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 that it is IC and feasible. In step 2, we show that such a perturbation strictly improves the relaxed problem (A2) 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, here we omit the subscript n for simplicity. Let μ represent any feasible and IC 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 $\mu(\tilde{\mathbf{c}})$ with probability γ, where $\tilde{c}$ 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 IC is satisfied since $a^\gamma(c_i)$ is nonincreasing 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 $G(x) = -((1 - F(x))/f(x))$ and $\bar{a}$ is $\int_0^1 a(x)f(x)\,dx$. Note that $a^\gamma(x)f(x)/\bar{a}$ is a density since $a^\gamma(x)f(x)/\bar{a} \geq 0$ and $\int_0^1 (a^\gamma(x)f(x)/\bar{a})\,dx = 1$. By MHRA, $G(x)$ is monotone nondecreasing in x, so the result is proven if we prove that an increase in γ implies a first-order stochastic dominance improvement in $(a^\gamma(x)/\bar{a})f(x)$. Define $\Gamma^\gamma(t) = \int_0^t (a^\gamma(x)/\bar{a})f(x)\,dx$. We prove the result if $\partial\Gamma^\gamma(t)/\partial\gamma < 1$ for all $t < 1$. We have

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

where the last inequality follows from the fact that $a(x)$ is nonincreasing 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 nonempty coalition of size smaller than m_n. In the first case, simply 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 case, 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 selects 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 $1 - \alpha_n(c/v)$. We conclude that the optimal solution of the relaxed problem is $p_n^S = 1$ and $a_n^S = \alpha_n$.

A3. 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 IC if and only if the volunteer cutoff is c_n^O, and $c_n^O \in (c_n^U, v)$. Finally, we establish that the VBO is HO. The following lemma will prove useful.

LEMMA A2. $B(\alpha_n n - 1 + j, n - 1, p)/B(\alpha_n n - 1, n - 1, p) = \prod_{k=1}^{j}((n - \alpha_n n - 1 - k)/(\alpha_n n - 1 + k)) \cdot (p/(1-p))^j$.

Proof. See the online appendix. QED

We now proceed in three steps.

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

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

We can rewrite it as

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

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.—IC 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 $p_n^O(c)$, $a_n^O(c)$ represents 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, $p_{1,n}^O$, $p_{2,n}^O$, and $a_n^O$are given by the following formulas:

$$a_n^O = \sum_{k=\alpha_n n-1}^{n-1} \frac{\alpha_n n}{k+1} B(k, n-1, F(c_n^O)), \tag{A3}$$

$$p_{1,n}^O = \sum_{k=\alpha_n n-1}^{n-1} B(k, n-1, F(c_n^O)), \tag{A4}$$

$$p_{2,n}^O = \sum_{k=\alpha_n n}^{n-1} B(k, n-1, F(c_n^O)). \tag{A5}$$

The equilibrium condition for IC is

$$ac_n^O = v(p_{1,n}^O - p_{2,n}^O). \tag{A6}$$

By substituting equations (A3), (A4), and (A5) into equation (A6) for any n, we obtain an expression for $c_n^O(v)$, the VBO volunteer threshold cost as a function of group size and the value of success:

$$c_n^O(v) = v\frac{B(\alpha_n n - 1, n - 1, F(c_n^O))}{\sum_{k=\alpha_n n-1}^{n-1}(\alpha_n n/(k+1))B(k, n-1, F(c_n^O))}, \tag{A7}$$

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 that a volunteer is activated. It is easy to see that (A7) 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 HO, 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 zero 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 $c_i \le v$ 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 zero if they disobey the recommendation to activate and $v - c_i > 0$ if they obey. It follows that all types $c_i \le v$ 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 $c_i' \ne c_i$. Formally: $E_{\mathbf{c}_{-i}}[\Sigma_{g\in I}\mu_g(\tilde{c}_i, \mathbf{c}_{-i})u_g^i(c)] \ge E_{\mathbf{c}_{-i}}[\Sigma_{g\in I}\mu_g(\tilde{c}_i, \mathbf{c}_{-i})u_{\xi_i(g)}^i(c)]$. This condition and (IC) imply that

$$E_{\mathbf{c}_{-i}}\left[\sum_{g\in I}\mu_g(c_i, \mathbf{c}_{-i})u_g^i(c)\right] \ge E_{\mathbf{c}_{-i}}\left[\sum_{g\in I}\mu_g(\tilde{c}_i, \mathbf{c}_{-i})u_g^i(c)\right] \tag{A8}$$

$$\ge E_{\mathbf{c}_{-i}}\left[\sum_{g\in I}\mu_g(\tilde{c}_i, \mathbf{c}_{-i})u_{\xi_i(g)}^i(c)\right] \tag{A9}$$

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

A4. 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

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

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

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

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, e.g., the proof of theorem 1), we have

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

where $\sigma_{c_n^O}(k/n) = \sqrt{F(c_n^O)(1-F(c_n^O))}/\sqrt{n}$. This proves the result. QED

A5. Proof of Proposition 3

We can bound the probability of success in an optimal HO mechanism as follows. Let $D((k/n)\|p) = (k/n)\log(k/(np)) + (1-(k/n))\log((1-k/n)/(1-p))$ represent the Kullback–Leibler divergence, or relative entropy, from p to k/n. We can write

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

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, $\lim_{n\to\infty}(p_n^O/(m_n/n)) \geq 1$, so for any $\epsilon > 0$, there exists n_ϵ such that for $n > n_\epsilon$, we have $D((m_n/n)\|p_n^*) \leq \epsilon/2$. Thus,

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

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. QED

A6. Proof of Proposition 4

We omit n here as a subscript for simplicity whenever it does not create confusion. Let μ^b represent an optimal binary mechanism, let c^b represent the volunteer cut

point associated with μ^b, and let q_k represent the corresponding probability of success when there are k volunteers. We 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 $q_{k^b} \in (0, 1)$. Thus, the optimal binary mechanism is a k^b-VBO except at most for an event with probability that converges to zero as $n \to 0$—that is, when there are exactly k^b volunteers.

We proceed in two steps. In step 1, we establish that the optimal binary mechanism is nonwasteful, 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 HO binary mechanism must be nonwasteful 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 IC. 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 $\tilde{a}^{\tilde{\mu}'} \tilde{c}^b = v(p_1^{\tilde{\mu}'} - p_2^{\tilde{\mu}'})$. (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), members are better off because for each $k \geq m$ for which $q_k > 0$ there is 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 $(v - (m/k)c)q_k > 0$. (Such members receive the same conditional expected utility for all other profiles.) Hence, $u(c; \tilde{c}^b) > u(c; c^b)$ for all $c \leq c^b$ and so 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 nonwasteful.

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 need to prove the result only 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 represent the minimum cost above which an agent is activated with probability zero. Then IC 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 that

$$\begin{aligned} & c^b \left[-\frac{m}{k} B(k-1, n-1, F(c^b)) + \frac{m}{k+1} B(k, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \right] dq_k \\ &= \left[\begin{array}{l} -(B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b))) \\ + B(k, n-1, F(c^b)) - B(k+1, n-1, F(c^b)) \frac{dq_{k+1}}{dq_k} \end{array} \right] dq_k. \end{aligned} \tag{A10}$$

Note that we can write

$$\begin{aligned} p_1^b - p_2^b &= B(n-1, n-1, F(c^b)) q_n \\ &\quad + \sum_{k=m}^{n-1} [B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b))] q_k = \sum_{k=m}^{n} \Theta_k q_n, \end{aligned} \tag{A11}$$

where we denote

$$\begin{aligned} \Theta_n &= B(n-1, n-1, F(c^b)), \\ \Theta_k &= B(k-1, n-1, F(c^b)) - B(k, n-1, F(c^b)) \text{ for } k = n-1, \ldots, m. \end{aligned}$$

We can rewrite the previous expression as

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

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 (A10) gives

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

or

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

It follows that

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

where

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

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

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

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. QED

A7. Proof of Proposition 5

As discussed in section V.A, we derive the objective function in the relaxed problem (15) using the (IC) constraint in a way similar to that in (9). The constraint in the second line of (15) is only 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

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

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

Let $V_n^{**}(c)$ denote the expected value for a type c in an optimal mechanism that solves the relaxed problem (15), and let $V_n^{**} = E_c\{V_n^{**}(c)\}$ denote the value of the objective function in (15). Note that $V_n^{**} \geq V_n^*$, where V_n^* represents the optimal mechanism in an HO 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 $\tilde{V}_n^{\varepsilon}$, 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 $\tilde{c}_n^{\varepsilon}$ 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, IC 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 $n > n_\eta \Longrightarrow V_n^b \geq V_n^{**} - \eta$ 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) $\lim_{n\to\infty} D_n = 0$. In this case,

$$\lim_{n\to\infty} D_n = \lim_{n\to\infty}\left(V_n^{**} - V_n^b\right) = \lim_{n\to\infty}\left(V_n^{**} - V_n^G\right) = 0,$$

where V_n^G represents the value in the optimal generalized VBO and we are finished.

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

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

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

Suppose that $\varepsilon(\eta) > 0$. Then for any $\varepsilon' \leq \varepsilon(\eta)$, there is an 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 an 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 $\lim_{n\to\infty} D_n = D > 0$. 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 $\lim_{n\to\infty}\epsilon(n, \eta) = 0$, then for any arbitrarily small $\varepsilon \in (0, \eta/2)$ there is an n_ϵ such that for $n > n_\epsilon$, we have $\tilde{V}_n^{\epsilon(n,\eta)} \leq \tilde{V}_n^0 + \varepsilon$.

Proof. See the online appendix. QED

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

Step 2.—We can now put together step 1 and proposition 4 to argue that a VBO is approximately optimal for large n. Since $V_n^b \leq V_n^*$, step 1 implies that $|V_n^b - V_n^*| \to 0$ as $n \to \infty$. Proposition 4, moreover, shows that the optimal binary mechanism is a generalized VBO with threshold k_n^*. Let VBO(k_n^*) represent 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 $|V_n^b - V_n^{\text{VBO}(k_n^*)}| \to 0$. Hence, we have that for any η there is an n_η such that for $n > n_\eta$ $V_n^{\text{VBO}(k_n^*)} \geq V_n^* - \eta$ for some threshold k_n^*, which implies the result. QED

A8. Proof of Theorem 4

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

Step 1.—We first prove that $\lim_{n\to\infty}(p_n^{\theta_n}/\theta_n) < 1$, where $p_n^{\theta_n} = F(c_n^{\theta_n})$ and $c_n^{\theta_n}$ represents the cutoff for participation with θ_n. In equilibrium, we must have

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

Consider the right-hand side. By the mean value theorem, we can write

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

where $\xi \in [0, vB(\theta_n n - 1, n - 1, p_n^{\theta_n})/(\sum_{j=\theta_n n-1}^{n-1}(m_n/(j+1))B(j, n-1, p_n^{\theta_n}))]$ 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

$$F\left(\frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1}(m_n/(j+1))B(j, n-1, p_n^{\theta_n})}\right) \leq (v\bar{f}) \cdot \frac{vB(\theta_n n - 1, n - 1, p_n^{\theta_n})}{\sum_{j=\theta_n n-1}^{n-1}(m_n/(j+1))B(j, n-1, p_n^{\theta_n})}, \quad \text{(A12)}$$

since $F(0) = 0$. We now prove that there is a constant $\vartheta < 1$ such that $\lim_{n\to\infty}(p_n^{\theta_n}/\theta_n) < \vartheta$. Assume not, so $\lim_{n\to\infty}(p_n^{\theta_n}/\theta_n) = \zeta \geq 1$. Following an argument

very similar to that in step 1 of theorem 3, we can prove that there is a constant $a_\infty > 0$ such that $\Sigma_{j=\theta_n n-1}^{n-1}(\theta_n/(j+1))B(j, n-1, p_n^{\theta_n}) \geq a_\infty$. 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 an argument very similar to that in step 2 of theorem 3, we can prove that

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

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

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

where the last step follows since $(\alpha_n)^2\theta_n n = (\alpha_n)^3(\theta_n/\alpha_n)n > (\alpha_n)^3 n$, 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 $\lim_{n\to\infty}(p_n^{\theta_n}/\theta_n) < \vartheta$. 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 converges to the same value. For this we use proposition 5—that the generalized 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 represent the expected per capita welfare in the optimal mechanism and in the optimal VBO, respectively. Note that for any ε, there is an $n_{1,\varepsilon}$ 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 $P_n^G \to 0$. Moreover, for any ε, there is an $n_{2,\varepsilon}$ 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 $c \leq v$, $a_n^*(c)(c/v) \leq p_n^*(c) - p_n^*(v) \to 0$, as proved earlier, and (*b*) $P_n^* \to P^* > 0$. It follows that for any ε, there is an $n_\varepsilon = \max\{n_{1,\varepsilon}, n_{2,\varepsilon}\}$ such that for $n > n_\epsilon$, $W_n^* - W_n^G > vP^* - \epsilon$.

By proposition 5, for any arbitrarily small $\eta > 0$, there is an n_η such that for $n > n_\eta$, $|W_n^* - W_n^G| < \eta$, where W_n^* and W_n^G represent the expected per capita

welfare in the optimal mechanism and in the optimal VBO, respectively. It follows that for large n, $\eta + \varepsilon > |W_n^* - W_n^G| \geq vP^*$, which is a contradiction since η and ε are both arbitrarily small and vP^* is bounded away from zero. QED

A9. Proof of Proposition 7

If $m_n \succ n^{2/3}$, $c_n^O > 0$ for all n and $c_n^U = 0$ for sufficiently large n, 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 brackets of (17) and (18) converge to zero faster than the second terms, and 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. QED

References

Archetti, M., and I. Scheuring. 2012. "Game Theory of Public Goods in One-Shot Social Dilemmas without Assortment." *J. Theoretical Biology* 299:9–20.

Ash, R. 1990. *Information Theory*. New York: Dover.

Barrett, S., and A. Dannenberg. 2014. "Sensitivity of Collective Action to Uncertainty about Climate Tipping Points." *Nature Climate Change* 4 (1): 36–39.

Battaglini, M. 2017. "Public Protests and Policy Making." *Q.J.E.* 132 (1): 485–549.

Battaglini, M., and R. Benabou. 2003. "Trust, Coordination, and the Industrial Organization of Political Activism." *J. European Econ. Assoc.* 1 (4): 851–89.

Battaglini, M., and B. Harstad. 2016. "Participation and Duration of Environmental Agreements." *J.P.E.* 124 (1): 160–204.

Battaglini, M., S. Nunnari, and T. Palfrey. 2014. "Dynamic Free Riding with Irreversible Investments." *A.E.R.* 104 (9): 2858–71.

Battaglini, M., and T. Palfrey. 2023. "Organizing Collective Action: Olson Revisited." Working Paper no. 30991, NBER, Cambridge, MA.

Bergstrom, T. 2017. "The Good Samaritan and Traffic on the Road to Jericho." *American Econ. J.: Microeconomics* 9 (2): 33–53.

Bierbrauer, F., and M. Hellwig. 2016. "Robustly Coalition-Proof Incentive Mechanisms for Public Good Provision Are Voting Mechanisms and Vice Versa." *Rev. Econ. Studies* 83:1440–64.

Bierbrauer, F., and J. Winkelmann. 2020. "All or Nothing: State Capacity and Optimal Public Goods Provision." *J. Econ. Theory* 185:1–15.

Braver, S. L., and L. A. Wilson. 1986. "Choices in Social Dilemmas: Effects of Communication within Subgroups." *J. Conflict Resolution* 30 (1): 51–62.

Chamberlin, J. R. 1974. "Provision of Collective Goods as a Function of Group Size." *American Polit. Sci. Rev.* 68:707–16.

Crémer, J., and R. P. McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* 56:1247–57.

d'Aspremont, C., J. Crémer, and L.-A. Gérard-Varet. 1990. "Incentives and the Existence of Pareto-Optimal Revelation Mechanisms." *J. Econ. Theory* 51 (2): 233–54.

d'Aspremont, C., and L.-A. Gérard-Varet. 1979. "Incentives and Incomplete Information." *J. Public Econ.* 11:25–45.
Diekmann, A. 1985. "Volunteer's Dilemma." *J. Conflict Resolution* 29:605–10.
Dixit, A., and M. Olson. 2000. "Does Voluntary Participation Undermine the Coase Theorem?" *J. Public Econ.* 76 (3): 309–35.
Dziuda, W., A. Gitmez, and M. Shadmeher. 2021. "The Difficulty of Easy Projects." *A.E.R.: Insights* 3 (3): 285–302.
Esteban, J., and D. Ray. 2001. "Collective Action and the Group Paradox." *American Polit. Sci. Rev.* 95 (3): 663–72.
Goldlucke, S., and T. Troger. 2020. "The Multiple-Volunteers Dilemma." Manuscript, Univ. Mannheim, Mannheim, Germany.
Healy, P. J. 2010. "Equilibrium Participation in Public Goods Allocations." *Rev. Econ. Design* 14:27–50.
Hellwig, M. F. 2003. "Public Good Provision with Many Participants." *Rev. Econ. Studies* 70:589–614.
Ledyard, J. O. 1984. "The Pure Theory of Large Two-Candidate Elections." *Public Choice* 44:7–41.
Ledyard, J. O., and T. R. Palfrey. 1994. "Voting and Lottery Drafts as Efficient Public Goods Mechanisms." *Rev. Econ. Studies* 61 (2): 327–55.
———. 1999. "A Characterization of Interim Efficiency with Public Goods." *Econometrica* 67 (2): 435–48.
———. 2002. "The Approximation of Efficient Public Good Mechanisms by Simple Voting Schemes." *J. Public Econ.* 83 (2): 153–72.
Mailath, G. J., and A. Postlewaite. 1990. "Asymmetric Information Bargaining Problems with Many Agents." *Rev. Econ. Studies* 57:351–67.
Matthews, S. 2013. "Achievable Outcomes of Dynamic Contribution Games." *Theoretical Econ.* 8 (2): 365–403.
Myerson, R. B. 1982. "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems." *J. Math. Econ.* 10:67–81.
———. 2000. "Large Poisson Games." *J. Econ. Theory* 94:7–45.
Nöldeke, G., and J. Peña. 2020. "Group Size and Collective Action in a Binary Contribution Game." *J. Math. Econ.* 88:42–51.
Olson, M. 1965. *The Logic of Collective Action.* Cambridge, MA: Harvard Univ. Press.
Ostrom, E. 1990. *Governing the Commons.* Cambridge: Cambridge Univ. Press.
Ostrom, E., and J. Walker. 1991. "Communication in a Commons: Cooperation without External Enforcement." In *Contemporary Laboratory Research in Political Economy*, edited by Tom Palfrey, 287–322. Ann Arbor, MI: Univ. Michigan Press.
Ostrom, E., J. Walker, and R. Gardner. 1992. "Covenants With and Without a Sword: Self-Governance Is Possible." *American Polit. Sci. Rev.* 86 (2): 404–17.
Palfrey, T. R., and H. Rosenthal. 1984. "Participation and the Provision of Public Goods: A Strategic Analysis." *J. Public Econ.* 24:171–93.
———. 1991. "Testing for Effects of Cheap Talk in a Public Goods Game with Private Information." *Games and Econ. Behavior* 3:183–220.
Palfrey, T. R., H. Rosenthal, and N. Roy. 2017. "How Cheap Talk Enhances Efficiency in Threshold Public Goods Games." *Games and Econ. Behavior* 101: 234–59.
Passarelli, F., and G. Tabellini. 2017. "Emotions and Political Unrest." *J.P.E.* 125 (3): 903–46.
Rob, R. 1989. "Pollution Claim Settlements with Private Information." *J. Econ. Theory* 47:307–33.

Tavoni, A., A. Dannenberg, G. Kallis, and A. Löschel. 2011. "Inequality, Communication, and the Avoidance of Disastrous Climate Change in a Public Goods Game." *Proc. Nat. Acad. Sci. USA* 108 (29): 11825–29.

van de Kragt, A., J. Orbell, and R. Dawes. 1983. "Organizing Groups for Collective Action." *American Polit. Sci. Rev.* 80 (4): 1171–85.