Solving the generalized volunteers' dilemma: a comparison of simple ex-ante versus ex-post mechanisms

Tim Friehe* Avraham Tabbach†

Preliminary draft from September 2015

Please do not cite without the authors' approval

Abstract

This paper explores a class of social dilemmas in which the participation of a prespecified number of individuals is required to achieve a social end (e.g., the prevention of a loss). We describe the first- and second-best outcomes before we study whether or not simple policy instruments can *solve* the social dilemma, distinguishing between policies that focus on the individuals' activity (i.e., ex-ante mechanisms) and policies that focus on the outcome (i.e., ex-post mechanisms). For the domain of simple policies, we establish that using a policy that only takes account of the individual action is superior to a policy that focuses only on whether the end was achieved or not, such that ex-ante regulation dominates ex-post mechanisms.

Keywords: Social Dilemma, Volunteer's Dilemma, Punishment, Rewards

JEL Code: H23, H41

*University of Marburg, Public Economics Group, Am Plan 2, 35037 Marburg, Germany. CESifo, Munich, Germany. E-mail: tim.friehe@uni-marburg.de.

†Tel Aviv University, Faculty of Law, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: adtabbac@post.tau.ac.il.

“Once, all the mice agreed that life would be better if the cat wore a bell to warn of his coming– until a wise old mouse asked, “Who will bell the cat?” Æsop. (Sixth century B.C.) “The Mice in Council”. Fables.

1 Introduction

Motivation and main results

Social dilemmas are situations in which the optimal decision of an individual conflicts with the optimal decision for society at large. The presumably most prominent example is the so-called Prisoners' Dilemma (PD). In the PD, a group of individuals needs to *cooperate* to achieve some goal, but each individual finds it strictly preferable to *defect* (i.e., not cooperate) independent of what the other individuals do (Tucker, 1950; Luce & Raiffa, 1957; Hardin, 1968; Dawes, 1980). From a social point of view, the PD can be solved relatively easily: a social planner can reward cooperators and/ or punish defectors, thereby inducing players to cooperate and uniquely attain the efficient outcome.[1]

However, not all social dilemmas are PD situations. Instead, in many cases, it is desirable that only some of the players act. For example, consider a group of bystanders watching a person in peril. The person will be saved when a single bystander provides aid, but suffer harm if no bystander volunteers. However, helping is costly such that each bystander prefers that a fellow observer, rather than he himself, renders assistance to the imperiled person. This example reflects the classic Volunteer's Dilemma (VD) first analyzed by Diekmann (1985), where one player suffices for the provision of the public good. The *Belling the Cat* fable, referred to by our introductory quote, is another illustration of the VD,[2] as is the famous biblical story of David and Goliath.[3] To

[1] Without a social planner, N person prisoner dilemmas may be solved by the threat of punishment in repeated interactions (Axelrod and Hamilton, 1981; Nowak, 2006).

[2] “Bell the Cat” is the nickname given the Scottish nobleman, Archibald Douglas, 5th Earl of Angus. In 1482, at a meeting of nobles who wanted to depose and hang James III's favourite, Robert Cochrane, Lord Gray remarked, “Tis well said, but wha daur bell the cat?” The challenge was accepted and successfully accomplished by the Earl of Angus. In recognition of this, he was always known afterwards as Archie Bell-the-cat (see Wikipedia, https://en.wikipedia.org/wiki/Belling_the_cat).

[3] The story of David and Goliath tells about the battle between the Philistines and the Israelites that took place in valley of Elah. The Philistines stood on a mountain on one side and the Israelites stood on a mountain on the

analyze other realistic settings, the VD has been generalized to the case in which more than one volunteer is necessary to provide the public good. For example, defeating a common enemy like in hostage dramas or in *Belling the Cat* scenarios may require that more than one individual takes action. Similarly, helping a person stuck in a crashed car after an accident typically requires that more than one individual will lend a hand. Another example of a generalized volunteers' dilemma due to Myatt and Wallace (2009) is attendance in voluntary committees that require a quorum, where attendance involves a private cost and generates a public benefit. The quorum may be likened to a production target such that a collective action (the committee's decision) goes ahead only if a threshold (the quorum) is reached. Finally, there are also examples in the environmental sphere where natural resources have tipping points such that using resources beyond this point may have disastrous consequences for the environment and welfare in general (Lenton et al. 2008). When the impact of an actor is more or less fixed, then the tipping point translates into a number of economic actors not refraining from the use of the natural resource.

From a social perspective, situations strategically equivalent to a VD are less problematic than scenarios that are best captured by the PD because, in VD games, players need not have a dominant strategy to defect, and some level of cooperation can be present without any intervention.[4] On the other hand, situations best described by the VD pose serious challenges when compared to PD situations, because it is more difficult to attain the desirable outcome using simple policies in these kind of games, as we will illustrate.[5]

This paper studies social solutions to a generalized volunteers' dilemma. We consider a game

other. Then, there went out a champion out of the camp of the Philistines, named Goliath of Gath, and he cried unto the armies of Israel: "Choose you a man for you, and let him come down to me. If he be able to fight with me and to kill me, then will we be your servants: but if I prevail against him and kill him, then you shall be our servants, and serve us." When Saul and all the Israelites heard those words of the Philistine, they were dismayed and greatly afraid. When David arrived to the scene he asked what will be done to the man who will kill the Philistine and the people say "the king will give great wealth to the man who kills him. He will also give him his daughter in marriage and will exempt his family from taxes in Israel." (See 1 Shmuel 17 New International Version).

[4]Volunteer's dilemma are usually characterized by a player's preference to take the action if no other player takes the action. However, this need not be the case. It may well be that all players have a dominant strategy not to volunteer. From our perspective, the crucial element of volunteers' dilemmas are the fact that a subset of players are required to provide the public good. As will become apparent, the game turns out to be a volunteer's dilemma in its classical sense by the intervention of a social planner.

[5]This phenomenon is also captured, to some extent, by the moral of the fable Belling the Cat: "it is one thing to say that something should be done, but quite a different matter to do (or suggest how it can be done)."

in which the action of a specific number of players is required to achieve a social end, which is representative of many real-life circumstances as elaborated above. Players have private information about their costs of taking action (i.e., volunteering costs) and simultaneously decide between participating and not participating. The private information about volunteering costs is a realistic assumption as well. Consider the committee example where specific private or professional circumstances may make attending the meeting particularly costly or the accident example in which some individuals may have particularly high costs of helping due to an accident history in their own family, for example. Individuals with heterogenous volunteering costs in organizational team settings, for example, may also reflect heterogeneous ability to perform the task, because the higher ability allows to achieve the individual contribution at lower costs.

For our framework, we first characterize the first-best outcome which mandates that the required number of players with the *minimum* realized costs of volunteering will act (provided that the sum of costs is lower than the harm resulting from the adverse event). However, absent verifiable information regarding the realized costs (i.e., the type of players) and absent a mechanism to elicit this kind of information given the nature of the emergency of the situations we analyze, the first-best outcome is practically unattainable. Therefore, we concentrate on inducing a second-best outcome, and in particular a symmetric second-best outcome in which all players are treated equally ex-ante. This special interest is due to the fact that it is hard to distinguish among players ex-ante and that equal treatment prescriptions in reality make other policies practically irrelevant.[6] Our research interest is thus with the performance of simple solutions, that is, policy instruments that allow the attainment of the second-best outcome without a great number of prerequisites. In our setting, attaining the feasible second-best outcome means that individuals should participate or not depending on whether or not their level of volunteering costs falls short of or exceeds a cutoff level. With regard to the implementation of this outcome, we find that making rewards and/or sanctions contingent only on the individuals' act outperforms the alter-

[6]Sometimes, it is possible to identify certain individuals with certain characteristics. The law in the US, for example, follows this path by imposing a duty to protect against unreasonable risk of physical harm on a person who has a certain relationship with the potential victim such as innkeeper or custodian (Restatement (Second) of Torts Section 314A (1965)).

native of making these payments contingent only on the eventual outcome. This ranking is due to the fact that ex-post regulation may induce an outcome in which all individuals choose not to participate, or participate excessively, whereas ex-ante regulation uniquely induces the desirable second-best outcome. The reason is simple. Regulation based on the action of players removes the strategic interaction among players, and therefore the possibility of multiple equilibria, whereas ex-post regulation maintains strategic interaction. On the other hand, ex-post regulation requires very little information (namely the level of social harm), whereas ex-ante regulation demands that the social planner be able to calculate the socially optimal cut-off level and to distinguish players who act from those who do not. The decision between ex-ante and ex-post regulation recurs in numerous contexts (e.g., Shavell 1993), making the additional results we derive here relevant for policy.

With regard to the payments required to implement the second-best outcome, ex-post regulation makes use of the level of harm while ex-ante regulation uses a lower level. Our analysis is thus roughly consistent with the observation that volunteers are typically rewarded for performing a task, at a level that does not reflect the harm or the social value of the public good, but instead reflect much lower levels. The following examples all have in common that there are presumably some individuals with the required knowledge to act but who may not act when there are no regulatory incentives to act due to the costs that are incurred by taking action. Moreover, these cases exemplify that the magnitude of incentives fall short of what is at stake from a social point of view, as is the case with ex-ante regulation. The US False Claims Act serves as an example. It permits a person with knowledge of fraud against the government of the United States to file a lawsuit on behalf of the government against the individual or firm that has committed the fraud. If the legal action is successful, the plaintiff is rewarded with a percentage of the recovery. Another example comes from the realm of tax enforcement. The IRS rewards those who provide information leading to tax investigations where money is successfully recovered. The reward is a function of the tax revenue recovered due to the tip by the informant (see, e.g., Yaniv 2001). A final contemporary example: in many countries, rewards are granted to anyone able to provide information leading to the arrest of individuals guilty of very serious offenses. Two prominent examples are the 'Rewards for Justice Program' in the fight against terror and the 'Narcotics Rewards Program' in the fight against drugs; both programs of the United States Department of State. Moreover, under ex-ante regulation, payments concern only a subset of the players whereas all players are concerned under ex-post regulation, which also seems to be closer to what we find in reality.

Related literature

Our paper is closely related to the literature on the VD that started with Diekmann (1985).[7] In the classic volunteers' dilemma, symmetric individuals simultaneously choose between helping or not, knowing that only one volunteer is required to provide the public good. The symmetric equilibrium is in mixed strategies and has worrisome features, as help is not always provided. The basic VD has been extended in several directions. For example, Diekmann (1993) introduces asymmetric preferences among volunteers, Weesie (1993) considers a sequential game in which the magnitude of the social harm increases with time, Weesie (1994) and Xu (2001) examine a game whereby players know their own cost-benefit ratios but only the distribution of other players' ratios, and Weesie and Franzen (1998) examine the potentially salutary effect of cost sharing among players. Our paper differs from this strand of the literature in that we are interested in simple social policies that improve upon the unregulated equilibrium in a generalized VD, comparing ex-ante and ex-post interventions. Leshem and Tabbach (forthcoming) consider policies for the classic VD setup (in which only one helper is required) with a focus on the comparison of rewards and sanctions when there are administrative costs associated with them. The distinction between a setup in which one volunteer suffices for achieving the social end and one in which more than one volunteer is needed proves to be important for the comparison of ex-ante and ex-post mechanisms.

Our research interest is with the relative performance of simple ex-ante and ex-post policies in the generalized VD. Thus, our paper also relates to the strand of the literature analyzing these two policy responses in other setups. In the law and economics literature, the comparison between ex-

[7] The structure of the VD is in fact very similar to the so-called threshold public good game. Palfrey and Rosenthal (1984) provide an important early analysis in this field in which numerous experimental studies have considered outcomes under different circumstances (e.g., Bagnoli and McKee 1991, Erev and Rapoport 1990). Myatt and Wallace (2008, 2009) represent two recent theoretical contributions. The close relationship between the VD and the threshold public good game is lost when each individual may not only decide between participating or not but also about the level of the contribution (e.g., Normann and Rau forthcoming).

post liability and ex-ante regulation started with Wittman (1977) and Shavell (1984). The latter argues in a setup in which there is imperfect enforcement and potential judgment proofness of the tortfeasor that the simultaneous use of the two instruments may be welfare-enhancing. Schmitz (2000) considers a small modification of Shavell's framework and arrives at the conclusion that either of the two policies ought to be used. Kolstad et al. (1990) provide a rationale relating to the tortfeasor's uncertainty about whether the court will hold him liable. Shleifer (2010) foresees an advantage for regulation vis-a-vis liability due to shortcomings of the courts, and Helland and Klick (2013) examine their relationship empirically. In a very recent contribution to the discussion, Shavell (2013) argues that negligence has an enforcement cost advantage relative to regulation, because the test of whether or not actual behavior was in line with prescribed behavior does not occur in all states of the world. The point that we would like to emphasize about ex-ante and ex-post regulation by using the generalized VD has not been considered yet. The paper that stems from this strand of the literature and is the closest to our research interest is Garoupa and Dari-Mattiacci (2009). They consider a setting in which one tortfeasor interacts with a victim and the accident technology is such that care-taking by one party suffices to rule out that an accident may happen, where the standard prescription is that the least-cost avoider should bear liability to have adequate care incentives. When the care costs of the opponent are unknown, however, it may be that no party takes care, an outcome that can be averted by the use of fines that apply independent of the outcome. We similarly argue that the adverse outcome of inaction may be prevented by using ex-ante instead of ex-post mechanisms. However, in contrast to their argument, in our setup, the possibility of inaction is not due to the asymmetric information on the costs of volunteering because the scenario in which no individual volunteers is always a Nash equilibrium whenever more than one volunteer is required. In other terms, a coordination problem is at the heart of the problem of ex-post mechanisms in our setup.

The paper proceeds as follows. Section 2 introduces the model. Section 3 discusses first- and second-best outcomes. Section 4 considers different potential *social solutions* to the problem at hand. Section 5 concludes.

2 The model

We consider n risk-neutral individuals playing an incomplete information, simultaneous-move game with action set $a \in \{0, 1\}$. Taking action $a = 0$ is meant as no participation and thus associated with no costs. Instead, taking action $a = 1$ is associated with a cost $c > 0$. This cost is a random variable drawn according to a distribution function $F(c)$ on the support $[0, C]$. We consider this cost to be rather specific to the context, such that all individuals are symmetric ex ante but asymmetric ex post. In all likelihood, the level of the realized participation cost in a specific context is practically unverifiable.

When m $(1 \leq m \leq n)$ individuals choose to *participate* (i.e., take action $a = 1$), the social end is attained. We will usually refer to this social end as the prevention of an adverse event, but other interpretations fit the framework as well. When less than m individuals participate, the adverse event occurs. The adverse event makes each individual from the group of n suffer monetarized losses in the level of $l \geq 0$. In addition, there is a negative effect on members of the society that are external to the considered group of n players, such that the total loss from the adverse event is $h = nl + e$.

3 Social optimum

In this section, we will elaborate on first- and second-best outcomes in our setting. When we turn to simple solutions to the dilemma in the next section, we will argue that practically feasible instruments must treat individuals equally, which rules out ex-ante asymmetric treatments and thereby some policy alternatives.

We assume that the welfare implications of the participation decisions can be measured by the level of social costs, summing over the expected costs of participation and the expected harm.

3.1 First-best outcome

In the first-best outcome, every individual in the set of m individuals for whom the realized costs are the smallest is required to participate (i.e., take action $a = 1$) if the sum of their costs is lower than the harm, thereby avoiding the adverse event.

The expected cost of participation for the m individuals with the smallest cost of taking action is in this case synonymous with the level of social costs (as the adverse event is avoided) and can be calculated as follows. Let $\mathbf{c} = (c_1, c_2, ..., c_n)$ be the vector of the n realized random variables. Then, the order statistic $c_{(k)}$ of rank k is the kth smallest value from this set. For example, we have $c_{(1)} = \min\{c_1, c_2, ..., c_n\}$ and $c_{(n)} = \max\{c_1, c_2, ..., c_n\}$. The distribution function of the order statistic $c_{(k)}$, $G_k(c)$, gives the probability that the kth smallest level is less than c and is defined as:

$$G_k(c) = \sum_{i=k}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) F(c)^n (1 - F(c))^{n-1}, \tag{1}$$

yielding a density function

$$g_k(c) = f(c) \left(\begin{array}{c} n \\ k-1 \end{array} \right) F(c)^{k-1} (1 - F(c))^{n-k}. \tag{2}$$

Using (2), the expected costs of participation for the subset of m individuals with the smallest cost is:

$$EC = \sum_{i=1}^{m} \int_0^C c \left[f(c) \left(\begin{array}{c} n \\ i-1 \end{array} \right) F(c)^{i-1} (1 - F(c))^{n-i} \right] dc. \tag{3}$$

We are now in the position to state optimal policy in the first-best outcome.

Proposition 1 *The minimum expected costs for the prevention of the adverse event are EC as specified in* (3)*. When* $EC < (>)\, h$*, it is ex ante socially desirable (not) to achieve the social end. From an ex post point of view, prevention is efficient only if* $\sum_{i=1}^{m} c_{(i)} < h$*, where* $c_{(i)}$ *is the ith order statistic.*

3.2 Symmetric second-best outcome

In the minimum of social costs, either no one participates or the m individuals with the lowest cost participate. In technical terms, in the first-best outcome, the social planner made some individuals participate with probability one and others refrain from participating contingent on the whole vector of realized levels of participation cost, that is, used a function $p_i^*(\mathbf{c}) \in \{0, 1\}$

for each individual i. This is due to the fact that an individual with a given participation cost realization will be called on for taking action depending on the other cost realizations.

In this section, we will consider the scenario in which the individual probability of participating is no longer a function of the whole vector of participation cost realizations, but instead of the individual cost realization c and (possibly multiple) cost cutoff levels $\tilde{c}$. This procedure allows for redundancy in the sense that more than m individuals participate, since there is no coordinating effort contingent on the whole vector $\mathbf{c}$. This procedure also means that there usually remains a positive probability for the adverse event occurring. The individual participation probability now uses only the individual cost realization and exogenous cutoffs. As a result, it may be that the cost realizations of all individuals are greater than the highest cutoff, implying that the participation probability is zero for all individuals, even though it still could be ex post efficient to prevent the adverse event. But without a coordinating effort, such that social costs comprise the expected costs of participation and the expected harm given the arrangement. We focus on the scenario in which the participation probability is symmetric. This is the practically relevant scenario. In all likelihood, it is hard to distinguish among players ex-ante. Moreover, equal treatment prescriptions make other policies practically irrelevant in reality.[8]

As a first step, we will argue that the individual participation probability is either equal to zero or equal to one. In other terms, there is no range of participation cost realizations where individuals are made to participate with a probability strictly between zero and one. To see this, suppose that all individuals are confronted with k cutoff levels $\tilde{c}_j$, $j = 1, ..., k$, where $k \geq 2$ and $\tilde{c}_1 < ... < \tilde{c}_k$ and with associated probabilities of taking action p_j when $c \in [\tilde{c}_{j-1}, \tilde{c}_j)$. Then, the social costs are

$$SC = n \sum_{j=1}^{k} p_j \int_{\tilde{c}_{j-1}}^{\tilde{c}_j} cf(c)dc + P\left(n, m-1, \tilde{c}_1, ..., \tilde{c}_k\right) h \tag{4}$$

where

$$P\left(n, m-1, \tilde{c}_1, ..., \tilde{c}_k\right) = \sum_{i=0}^{m-1} \left(\begin{array}{c} n \\ i \end{array} \right) \left(\sum_{j=1}^{k} p_j \int_{\tilde{c}_{j-1}}^{\tilde{c}_j} f(c)dc \right)^{i} \left(1 - \left(\sum_{j=1}^{k} p_j \int_{\tilde{c}_{j-1}}^{\tilde{c}_j} f(c)dc \right) \right)^{n-i}, \tag{5}$$

[8] In the appendix, we consider the scenario in which the policy maker may utilize different cutoff levels at the same time. It turns out that the consideration of the symmetric case need not be a restriction.

where $\sum_{j=1}^{k} p_j \int_{\tilde{c}_{j-1}}^{\tilde{c}_j} f(c)dc$ gives the probability that one individual will participate.

Referring to (4) and (5), the intuitive explanation of our result is that it is possible to increase p_j while lowering p_{j+1} such that the level of the probability of the adverse event P stays the same while the level of social costs decrease. In other terms, the change in p_j and p_{j+1} maintains the probability that individuals will participate in preventing the adverse event but makes it more likely that this will be achieved at lower participation costs.

We summarize the result (that is proved in detail in the appendix) as:

Proposition 2 *When participation probability functions are symmetric, the optimal rule involves a unique cutoff value such that individuals for whom the realized participation costs are less than the cutoff participate and others do not.*

When participation probability functions are symmetric, the policy maker employs only one participation cost cutoff level effectively. We will now derive a characterization of this cutoff level for the level of participation costs. With only one cutoff, the representation of social costs simplifies to

$$SC = n \int_0^{\tilde{c}} cf(c)dc + P\left(n, m-1, \tilde{c}\right) h \tag{6}$$

where

$$P\left(n, m-1, \tilde{c}\right) = \sum_{i=0}^{m-1} \binom{n}{i} F(\tilde{c})^i \left(1 - F(\tilde{c})\right)^{n-i} \tag{7}$$

is the probability that there will be at most $m-1$ individuals taking action $a = 1$ out of the set of n individuals when the cutoff value of costs is $\tilde{c}$.

The social planner seeks to minimize (6) using the level of the cutoff $\tilde{c}$. One possible solution is the corner solution where $\tilde{c}^* = 0$ and social costs are therefore equal to h. For an interior solution, the necessary first-order condition is

$$\frac{dSC}{d\tilde{c}} = hnf(\tilde{c}) \left(\frac{\tilde{c}}{h} - \binom{n-1}{m-1} F(\tilde{c})^{m-1}(1 - F(\tilde{c}))^{n-m} \right) = 0. \tag{8}$$

The term in parentheses is decisive and $\tilde{c}/h$, which reflects marginal costs, is linearly increasing with the level of the cutoff. The second term in the parentheses, which we will denote by Γ

$$\Gamma\left(n-1, m-1, \tilde{c}\right) = \binom{n-1}{m-1} F(\tilde{c})^{m-1}(1 - F(\tilde{c}))^{n-m} \tag{9}$$

and which reflects marginal benefits has a maximum with respect to $\tilde{c}$ where $F(\tilde{c}) = (m-1)/(n-1)$. It follows that there are either no, one or two interior levels of $\tilde{c}$ solving the first-order condition (8), which always is fulfilled at the corner solution $\tilde{c} = 0$. When two interior levels solve the first-order condition, then the higher one will reflect a local minimum, because $d^2SC/d\tilde{c} > 0$ will hold. This level, $\tilde{c}^*$, is implicitly defined by

$$\frac{\tilde{c}}{h} = \binom{n-1}{m-1} F(\tilde{c})^{m-1}(1 - F(\tilde{c}))^{n-m}. \tag{10}$$

The corner cutoff level will result (i.e, $\tilde{c}^* = 0$) when trying to prevent the adverse by the means of symmetric participation probabilities is too expensive given the level of harm. An interior cutoff for participation costs implies that the individual will participate only in some states of the world. An increase in the level of the cutoff increases the level of costs actually incurred and the probability that there will be some redundancy as more than m individuals participate. The upside of the increase in the level of the cutoff lies with decreasing the probability that the adverse event occurs, which goes to zero as the cutoff level goes to C. The socially optimal cutoff level will never be equal to the upper bound C, because $dSC/d\tilde{c} > 0$ at that level.

Proposition 3 *When participation probability functions are symmetric and specify an interior participation cost cutoff level, this level fulfills* (10).

4 Simple Social Solutions

In this section, we examine institutions that implement the outcome in which all individuals participate only when their participation costs are less than the socially optimal cutoff level. We consider the possibility of utilizing punishments and/or rewards, and distinguish social solutions according to the information relied upon. First, we address potential solutions that make payments contingent only on the observation of the outcome (i.e., on whether or not the adverse event occurred). Next, we consider instruments that make payments contingent on the individual's choice (i.e., on whether or not the player took action or not). Our distinction reflects that between ex-post liability (based on outcome) and ex-ante regulation (based on actions).

4.1 Ex-post punishment and rewards

The outcome of the interaction is usually easily observable, because the adverse event either occurred or did not occur. As a result, when the n players who may act are easily defined and identifiable ex post (not necessarily a trivial assumption), it is possible to hand out rewards (punishments) to all n individuals when the adverse event was prevented (occurred).

Let us assume that the social planner can punish players in the event that the adverse event occurred with a per capita punishment of magnitude $S \geq 0$ or can reward players in the event that the averse event did not occur with a per capita reward of magnitude $R \geq 0$. We are searching for the levels of S and R, respectively, that implement the second-best outcome.

In light of the symmetric nature of the game, we focus on symmetric pure strategy equilibria. The individual's strategy in this game is a participation probability function that maps from the space of costs $[0, C]$ into a binary choice set $\{0, 1\}$. The critical ingredient of this function is the cutoff utilized by the individual. Denote by $\tilde{c}_{-i} = (\tilde{c}_1, ..., \tilde{c}_{i-1}, \tilde{c}_{i+1}, ..., \tilde{c}_n)$ the cutoffs (i.e., strategies) of all players but player i. Player i's payoff from participating is thus

$$u_{a=1}^{EP} = (1 - P(n-1, m-2, \tilde{c}_{-i})R - c - P(n-1, m-2, \tilde{c}_{-i})(S + l) \tag{11}$$

and that from not participating

$$u_{a=0}^{EP} = (1 - P(n-1, m-1, \tilde{c}_{-i})R - P(n-1, m-1, \tilde{c}_{-i})(S + l), \tag{12}$$

leading to

$$\Delta^{EP} = \Gamma(n-1, m-1, \tilde{c}_{-i})(R + S + l) - c. \tag{13}$$

In the symmetric equilibrium, the participation probability function is symmetric across players and the resultant cutoff for the level of participation costs must be such that (13) is equal to zero. The term in (13) is actually quite similar to the first-order condition for the socially optimal cutoff level, which allows us to derive the result that the socially optimal cutoff may be implementable when $R + S = h - l$. Quite intuitively, it is up to the instruments reward and sanction to close the gap between the level of harm that is socially relevant and the one that is of importance for the individual at hand. For example if $l = 0$, then the sum of the punishment and rewards should equal to the social harm h.

Making payments contingent on the outcome can induce the second-best outcome. However, the use of the described solution allows for other equilibria as well. In fact, the case in which no one participates (i.e., where all individuals set $\tilde{c} = 0$) is also always an equilibrium under this regime when $m > 1$. The reasoning is simple. If all other players never participate, then a player's best response is to never participate, because taking action $a = 1$ involves costs but cannot generate any benefit. Moreover, the condition of having (13) equal to zero is also fulfilled at the other interior solution which does not minimize costs.

Proposition 4 *Under an institution that pays a reward when the adverse event was prevented and punishes individuals with a sanction when the adverse event occurred with $R + S = h - l$, the socially optimal cutoff is an equilibrium. However, such an institution also allows for other equilibria when $m > 1$. In particular, the equilibrium in which all individuals choose not to participate is an equilibrium in that case.*

4.2 Ex-ante regulation

In this section, we will discuss solutions that make payments dependent only on actions. In our framework, individuals may participate or not and have private information about their level of participation costs.

The solution we propose is actually very simple. A social planner who aims at implementing a cutoff value $\tilde{c}^* > 0$ may do so by offering players a reward R if they take action $a = 1$ and punish them with a sanction S if they take action $a = 0$. The payoff from participating is given by

$$u^{EA}_{a=1} = R - c - P(n-1, m-2, \tilde{c}_{-i})l \tag{14}$$

and that from not participating by

$$u^{EA}_{a=0} = -S - P(n-1, m-1, \tilde{c}_{-i})l. \tag{15}$$

The difference can be calculated as

$$\Delta^{EA} = R + S + \Gamma(n-1, m-1, \tilde{c}_{-i})(l) - c. \tag{16}$$

It is the objective of the policy maker to ensure that the level of c that sets (16) equal to zero is $\tilde{c}^*$.

Proposition 5 *Suppose that $l \to 0$. Under an institution that pays a reward for participation and punishes individuals for not participating with $R + S = \tilde{c}^*$, the socially optimal cutoff is a unique equilibrium.*

When the individual suffers almost no harm from the adverse event (i.e., when $l \to 0$), then individuals are no longer in a strategic interaction. Instead, each individual determines whether or not to participate depending on the payments specified. It is important to note that the deficiency of the ex-post regulation, namely that it permits the equilibrium in which no individual participates, can be ruled out under ex-ante regulation even if there remains strategic interaction due to a strictly positive personal loss l.

The important advantage of an ex-ante policy is that it implements the second-best solution as a unique equilibrium. However, it may come at a costs. First, it might be more costly to observe and verify which player did or did not take action $a = 1$. In addition, it requires transfers of rewards or punishment independent of the outcome, while an ex-post mechanism, if it works may involve a very infrequent use of transfers/punishments (e.g., Shavell 2013).

5 Conclusion

Social dilemmas – situations in which self-interest is at odds with collective interests – are a pervasive aspect recurring in many spheres of social interaction. These circumstances are often interpreted as Prisoner's Dilemmas in which players have dominant strategies and for which simple policy solutions have long been discussed. This paper studies simple policy solutions for a social dilemma of a different kind, namely circumstances in which only some of the players ought to act prosocially. For example, a given number of people is required to free a hostage or to save somebody from drowning.

In the second-best outcome that treats all symmetric subjects the same, the policy maker makes players' participation contingent on whether or not the level of realized costs exceeds a cutoff level.

In this way, the policy maker trades off the possibility that there will be an insufficient number of players acting against the expected costs in terms of possible redundancy (when more than the required number of players incur the costs of taking action) and higher cost levels.

In terms of simple policy solutions, we explore the relative performance of an ex-post regulation that depends only on the outcome (i.e., did the effort succeed or fail?) and an ex-ante regulation that focuses on whether a player participates or not. When appropriately designed, both regulations can induce the second-best outcome. However, the ex-post regulation cannot induce the desired outcome with certainty because the case in which no player takes action is also an equilibrium when more than one volunteers are required. This does not arise under ex-ante regulation since the regulation puts away with the strategic interdependence and strongly argues in favor of ex-ante regulation. However, ex-ante regulation may put a higher informational burden on the policy maker.

References

Bagnoli, M., and M. McKee, 1991. Voluntary contribution games: Efficient provision of public goods. Economic Inquiry 29, 351-366.

Dari-Mattiacci, G., and N. Garoupa, 2009. Least-cost avoidance: The tragedy of the commons. Journal of Law, Economics, and Organization 25, 235-261.

Dawes, R.M. 1980. Social dilemmas. Annu. Rev. Psychol. 31: 169-193.

Diekmann, A., 1985. Volunteer's dilemma. Journal of Conflict Resolution 29, 605-610.

Diekmann, A., 1993. Cooperation in an asymmetric volunteer's dilemma game. Theory and experimental evidence. International Journal of Game Theory 22, 75-85.

Erev, I., and A. Rapoport, 1990. Provision of step-level public goods: The sequential contribution mechanism. Journal of Conflict Resolution 34, 401-425.

Hardin, J. 1968. The tragedy of the commons. Science 162: 1243-1248.

Helland, E., and J. Klick, 2013. Regulation and Litigation: Complements or Substitutes. In: Buckley, F.H. (Ed.), The American Illness: Essays on the Rule of Law. Yale University Press.

Kolstad, C.D., Ulen, T.S., and G.V. Johnson, 1990. Ex post liability for harm vs. ex ante regulation of safety: Substitutes or complements? American Economic Review 80, 888-901.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and H. J. Schellnhuber, 2008. Tipping elements in the Earth's climate system. Proceedings of the National Academy of Sciences 105, 1786-1793.

Leshem, S., and A. Tabbach, forthcoming. Solving the volunteer's dilemma: The efficiency of rewards versus punishments. American Law and Economics Review.

Luce, R.D. and Raiffa, H. 1957. Games and Decisions: Introduction and Critical Survey. Wiley & Sons, New York.

McAdams, R., 2009. Beyond the prisoners' dilemma: Coordination, Game Theory, and the Law. Southern California Law Review 82, 209-258.

Myatt, D.P., and C. Wallace, 2008. When does one bad apple spoil the barrel? An evolutionary analysis of collective action. Review of Economic Studies 75, 499-527.

Myatt, D.P., and C. Wallace, 2009. Evolution, teamwork, and collective action: Production targets in the private provision of public goods. Economic Journal 119, 61-90.

Normann, H.T., and H. Rau, forthcoming. Simultaneous and sequential contributions to step-level public goods: One vs. two provision levels. Journal of Conflict Resolution.

Nowak, M. 2006. Five rules for the evolution of cooperation. Science 314: 1560-1565.

Palfrey, T.R., and H. Rosenthal, 1984. Participation and the provision of discrete public goods: A strategic analysis. Journal of Public Economics 24, 171-193.

Shavell, S., 1984. A model of the optimal use of liability and safety regulation. Rand Journal of Economics 15, 271-280.

Shavell, S., 1993. The optimal structure of law enforcement. Journal of Law and Economics 36, 255-287.

Shavell, S., 2013. A fundamental enforcement cost advantage of the negligence rule over regulation. Journal of Legal Studies 2013, 275-302.

Shleifer, A., 2010. Efficient regulation. In: Kessler, D. (Ed.), Regulation Vs. Litigation. NBER and University of Chicago Press.

Tucker, A. 1950. A two-person dilemma. In: Readings in Games and Information (E. Rasmusen, ed., 2001), pp. 7-8. Blackwell, Oxford.

Weesie, J. 1993. Asymmetry and timing in the volunteer's dilemma. Journal of Conflict Resolution 37, 569-590.

Weesie, J., 1994. Incomplete information and timing in the volunteer's dilemma: A comparison of four models. Journal of Conflict Resolution 38, 557-585.

Weesie, J., and A. Franzen, 1998. Cost sharing in the volunteer's dilemma. Journal of Conflict Resolution 42, 600-618.

Wittman, D., 1977. Prior regulation versus post liability: The choice between input and output monitoring. Journal of Legal Studies 6, 193-212.

Xu, X., 2001. Group size and the private supply of a best shot public good. European Journal of Political Economy 17, 897-904.

Yaniv, G., 2001. Revenge, tax informing, and the optimal bounty. Journal of Public Economic

Theory 3, 225-233.

Appendix

Proof to Proposition 2

We will show first that p_i must be such that $p_1 > p_2 > ... > p_k$. Suppose to the contrary that there exists $p_j < p_{j+1}$. Then we can increase p_j and decreases p_{j+1} such that P remains constant. That implies that the change in p_{j+1} should satisfy:

$$\frac{dp_{j+1}}{dp_j} = -\frac{\int_{c_{j-1}}^{c_j} f(c)dc}{\int_{c_j}^{c_{j+1}} f(c)dc}$$

The effect of such a change on the social costs is given by:

$$\int_{c_{j-1}}^{c_j} cf(c)dc - \int_{c_j}^{c_{j+1}} cf(c)dc * \frac{\int_{c_{j-1}}^{c_j} f(c)dc}{\int_{c_j}^{c_{j+1}} f(c)dc}$$

which can be rewritten as:

$$\frac{1}{\int_{c_j}^{c_{j+1}} f(c)dc}\left(\int_{c_{j-1}}^{c_j} cf(c)dc * \int_{c_j}^{c_{j+1}} f(c)dc\right) - \left(\int_{c_j}^{c_{j+1}} cf(c)dc * \int_{c_{j-1}}^{c_j} f(c)dc\right)$$

writing the conditional expectation of $\int_{c_{j-1}}^{c_j} cf(c)dc$ as $\widetilde{c_j} \int_{c_{j-1}}^{c_j} f(c)dc$, we obtain

$$\frac{1}{\int_{c_j}^{c_{j+1}} f(c)dc}\left(\widetilde{c_j}\int_{c_{j-1}}^{c_j} f(c)dc * \int_{c_j}^{c_{j+1}} f(c)dc\right) - \left(\widetilde{c_{j+1}}\int_{c_j}^{c_{j+1}} f(c)dc * \int_{c_{j-1}}^{c_j} f(c)dc\right) < 0$$

Because $c_{j-1} < \widetilde{c_j} < c_j$ and $c_j < \widetilde{c_{j+1}} < c_{j+1}$. Thus it must be that $p_1 > p_2 > ... > p_k$.

We will now show that $c_j < c_{j+1}$ for any given j. Consider increasing c_j and decreasing c_{j+1} while keeping the probability of the adverse effect, and therefore of the expected harm constant. This requires that

$$\frac{dc_2}{dc_1} = -\frac{f(c_j)(p_j - p_{j+1})}{f(c_{j+1})(p_{j+1} - p_{j+2})}. \tag{17}$$

The effects such a change on the costs of taking the action a=1 is (after rearranging and simplifying):

$$nf(c_1)(p_j - p_{j+1})(c_j - c_{j+1}) < 0 \tag{18}$$

The last inequality follows because $c_j < c_{j+1}$.

The use of heterogeneous cutoffs for individuals identical ex ante

Our analysis in the body of the paper has focused on the practically relevant scenario in which the participation probability is symmetric, because all n individuals are ex ante symmetric so that equal treatment demands symmetric probability functions. At least theoretically, the social planner may differentiate the cost cutoffs across individuals. At one extreme, there is one cutoff level for each of the n individuals. At the other extreme, all individuals face the same cutoff level. For example, one possible configuration would be to have the cutoff set equal to C for m individuals and equal to zero for the remaining $n-m$ individuals. In this case, the probability of the adverse event would be equal to zero and the possibility of redundancy is ruled out, but the actually incurred participation costs may turn out to be quite high.

To illustrate, we will consider an example where $m = 2$ and $n = 3$. The level of social costs results as

$$SC = \sum_{j=1}^{3} \int_0^{\tilde{c}_j} c dF(c) + Ph, \tag{19}$$

where

$$P = \prod_{j=1}^{3}(1 - F(\tilde{c}_j)) + \prod_{i,j,k=1}^{3} F(\tilde{c}_i)(1 - F(\tilde{c}_j))(1 - F(\tilde{c}_k)) \tag{20}$$

with $i \neq j \neq k$.

The first-order conditions for the three cutoff levels result as

$$\frac{\partial SC}{\partial \tilde{c}_1} = f(\tilde{c}_1)\left[\tilde{c}_1 - h\left\{F(\tilde{c}_2)(1 - F(\tilde{c}_3)) + F(\tilde{c}_3)(1 - F(\tilde{c}_2))\right\}\right] = 0 \tag{21}$$

$$\frac{\partial SC}{\partial \tilde{c}_2} = f(\tilde{c}_2)\left[\tilde{c}_2 - h\left\{F(\tilde{c}_1)(1 - F(\tilde{c}_3)) + F(\tilde{c}_3)(1 - F(\tilde{c}_1))\right\}\right] = 0 \tag{22}$$

$$\frac{\partial SC}{\partial \tilde{c}_3} = f(\tilde{c}_3)\left[\tilde{c}_3 - h\left\{F(\tilde{c}_2)(1 - F(\tilde{c}_1)) + F(\tilde{c}_1)(1 - F(\tilde{c}_2))\right\}\right] = 0. \tag{23}$$

The conditions show that the increase in expected participation costs is traded off against the impact on the probability of the adverse event, where only the states where the individual at hand is pivotal are relevant (i.e., where one individual does not participate whereas one other individual does). It is clear from this that the symmetric outcome solves the set of first-order conditions, but also that there may be alternative configurations such as $\tilde{c}_1 \neq \tilde{c}_2 = \tilde{c}_3$.