Punish Liars, Not Free-Riders*

Brenton Kenkel†

February 17, 2017

This is preliminary work.
Please email me for the latest version before citing.

Abstract

I consider how to encourage efficient public goods provision when contributions are voluntary and potential contributors are unsure of each other's willingness to give. Efficient distribution of costs requires contributors to reveal their private information honestly, but the desire to free-ride gives contributors a short-term incentive to misrepresent. I characterize the conditions under which the threat of communication breakdown in the long run is sufficient to sustain honest information revelation in the short run. The key conditions are that contributors must value the future highly and that there must be a sufficiently great risk that the project will not be completed if a contributor understates her willingness to give. Social welfare is typically greater under this equilibrium, in which contributors are punished for lying rather than for free-riding, than under "tit for tat"–style solutions that mandate contributions every period. The results challenge the conventional wisdom about reciprocity as a key to cooperation and suggests instead communication as a mechanism for enhancing public goods provision in the international arena and other settings.

*Prepared for presentation at the 2017 International Studies Association conference. I thank Rob Carroll, Mike Gibilisco, Brooke Hamilton, and participants in the Vanderbilt Political Science Faculty Workshop for helpful comments.

†Assistant Professor, Department of Political Science, Vanderbilt University. Email: `brenton.kenkel@vanderbilt.edu`.

1 Introduction

How can individuals (or other political actors) cooperate to achieve common goals when there are pervasive incentives to free-ride? This question, which encapsulates the collective action problem (Olson 1965), is a central concern for scholars of political science and political economy. Perhaps the most prominent mechanism proposed to overcome the collective action problem is tit-for-tat reciprocity (Axelrod 1984). Under tit-for-tat, you contribute tomorrow only if I contribute today, and vice versa. It is a simple and intuitively appealing way to prevent free-riding.

But the prevention of free-riding is the not only priority for groups that face collective action problems. The division of labor is often just as critical. For example, think of the ongoing refugee crisis. The question facing each European government is not whether to contribute by admitting refugees—it is *how many* refugees to admit. Out of the many possible ways to split the burden, which one will best promote successful collective action?

In economic terms, the most efficient division of labor is one that matches supply with demand. Those who value the project most highly, or for whom the cost of contribution is least, take on the bulk of the effort. If the costs and benefits are roughly equal across contributors, an even division of labor is efficient. But when the costs or benefits differ significantly, the most efficient allocation of effort might involve no contribution by those for whom the project is most costly or least valuable. In other words, an efficient division of labor may involve free-riding. The tit-for-tat approach to cooperation—or any other approach focused on the minimization of free-riding—is therefore at odds with the efficiency ideal.

Perhaps we could make tit-for-tat compatible with an efficient division of labor if we thought of reciprocity differently. Instead of simply identifying free-riding as non-cooperative behavior, we could say that a potential contributor has failed to cooperate when she contributes less than the efficient solution demands from her. If a potential contributor's benefit-cost ratio for the project were low enough, she would not be punished for free-riding under this modified scheme. However, this raises another complication: the contributors' relative costs and benefits might not be publicly known. In that case, in order to know what division of labor is most efficient and thereby be able to identify non-cooperative behavior, the contributors' private information about their costs and benefits must be publicized. This is a famously difficult incentive problem, as it is in each individual contributor's interest to understate her own demand for the project so as to induce others to take up more of the burden (Samuelson 1954).

In this paper, I establish conditions under which *communication*—ordinary, unverifiable talk among potential contributors—can lead to an efficient division of labor in collective action problems. I ground the argument in a formal model of repeated collective action under uncertainty. In the model, two partners face a new project each period. Every project can be completed by a 50-50 division of labor or by a single player supplying all the effort. The players privately learn the most they are willing to contribute—the whole cost of the project, half its cost, or nothing at all—which varies across projects. Coordination on the most efficient provision scheme is possible only if the players honestly reveal their willingness to contribute.

I characterize an equilibrium of the model in which it is in each player's individual interest to be honest. Not only does meaningful communication take place, but it is the threat of communication breaking down that sustains the equilibrium. Specifically, if a player is caught having lied about her willingness to contribute—if her eventual contribution does not match her earlier claims—then no communication takes place in future periods. The subsequent contribution outcomes, taking place in an environment of uncertainty, are necessarily inefficient. When the shadow of the future looms large enough, this potential loss of efficiency is great enough to keep the players honest.

A counterintuitive feature of the equilibrium is that a player will be punished (through the breakdown of communication) not just for contributing less than she claimed to be willing to, but also for contributing more. Why should these apparently benevolent surprises be punished? Suppose they are not, and imagine a player who values the project so highly that she is willing to contribute its entire cost. If she reveals this, then she will most likely be on the hook for the full cost (unless her partner happens to be equally willing). Alternatively, she could pretend to be willing only to contribute half of the cost. Then there are two possible outcomes: (1) her partner is indeed willing to take up half the cost, so the project is completed at strictly less cost to the player; or (2) her partner is unwilling to contribute at all, in which case the player completes the project on her own, just as she would have done if she were honest. The player is strictly better off in the first case and no worse off in the second, so it is profitable for her to screen her partner by downplaying her true willingness to contribute. To prevent this incentive for dishonesty, unexpected benevolence—which can only result from this sort of screening behavior—must be punished.

This paper has three close relatives in the literature. The first is the analysis by Sartori (2002) of repeated cheap talk in international crisis bargaining. I use a model that is structurally similar to Sartori's, in which players can communicate through cheap talk about period-dependent private information, albeit with a substantively distinct stage game and application. The second is Stone, Slantchev and London (2008), who model infinitely repeated public goods problems in the international system. My analysis differs from theirs insofar as I introduce incomplete information and communication, while setting aside the idea of an international hegemon. The third is Kenkel (2016), who shows that cheap talk communication is inefficacious in a wide range of one-shot public goods problems. This paper shows that introducing the shadow of the future allows for efficient outcomes and influential communication that are impossible in the one-shot setting.

The remainder of the paper proceeds as follows. Section 2 lays out the model and formally defines an efficient division of labor. Section 3 solves for an efficient equilibrium. Section 4 finds that, in terms of social welfare, the mechanism proposed here is usually superior to a simple 50-50 division of labor. Section 5 briefly offers ideas for extending the results.

2 The Model

2.1 Stage Game

In the stage game, there are two players. Let $i \in \{1, 2\}$ denote a generic player and j her partner. At the start of the period, Nature draws each player's *type* $\omega_i \in \Omega_i = \{0, 1, 2\}$. ω_1 and ω_2 are independent and identically distributed, with each having prior probability $p(\omega_i)$. Each player learns her own type but not her partner's; the prior distribution is common knowledge. To avoid trivialities, I assume throughout the main analysis that $p(\omega) > 0$ for each $\omega \in \{0, 1, 2\}$.

In the *contribution stage*, each player simultaneously selects an amount $x_i \in X_i = \{0, 1, 2\}$ to contribute to the provision of a public good. The good, whose value is normalized to 1, is supplied if and only if $x_1 + x_2 \geq 2$. A player's type determines her marginal cost of contribution, $c(\omega_i)$. Payoff functions are

$$u_i(x_i, x_j, \omega_i) = \begin{cases} -c(\omega_i)x_i & x_i + x_j < 2, \\ 1 - c(\omega_i)x_i & x_i + x_j \geq 2. \end{cases} \tag{1}$$

The strategic form of the contribution stage appears in Figure 1. Let $c_\omega = c(\omega)$ and $p_\omega = p(\omega)$ for each $\omega \in \{0, 1, 2\}$. I assume $0 < c_2 < \frac{1}{2} < c_1 < 1 < c_0$, so that a type-$\omega$ player is willing to contribute at most ω toward the success of the project. Contributions $x_i > \omega_i$ are strictly dominated by $x_i = 0$.

		x_2: 0	1	2
	0	0 0	0 $-c(\omega_2)$	1 $1-2c(\omega_2)$
x_1	1	$-c(\omega_1)$ 0	$1-c(\omega_1)$ $1-c(\omega_2)$	$1-c(\omega_1)$ $1-2c(\omega_2)$
	2	$1-2c(\omega_1)$ 1	$1-2c(\omega_1)$ $1-c(\omega_2)$	$1-2c(\omega_1)$ $1-2c(\omega_2)$

Figure 1: Strategic form of the contribution stage.

Between when players learn their types and the contribution stage is the *messaging stage*, in which each player simultaneously sends a message $m_i \in M_i = \{0, 1, 2\}$ about her type. After receiving her partner's message m_j, player i updates her beliefs about ω_j to $\lambda_i(m_j)$, where $\lambda_i : M_j \to \Delta\Omega_j$ is called a *belief system*. Messages are cheap talk (Crawford and Sobel 1982) and have no direct effect on payoffs.

Stage game strategies are as follows. A *messaging strategy* is a function $\mu_i : \Omega_i \to \Delta M_i$ that prescribes a probability distribution over messages for each type of player i. A messaging strategy is *fully separating* if the player always reveals her type exactly. Formally, in a fully separating messaging strategy, $\operatorname{supp} \mu_i(\omega_i) = \{\omega_i\}$ for all $\omega_i \in \Omega_i$, where supp denotes the support of a probability distribution.[1] A *contribution strategy* is a function $\sigma_i : \Omega_i \times M_i \times M_j \to X_i$ that prescribes a contribution for each type of player i in each

[1] More precisely, there is no loss of generality in assuming any fully separating strategy takes this form.

subgame.[2] An *assessment* is a list $(\mu_i, \sigma_i, \lambda_i)_{i=1,2}$ of each player's strategies and belief system; it is *symmetric* if $(\mu_1, \sigma_1, \lambda_1) = (\mu_2, \sigma_2, \lambda_2)$.

Like most cheap talk games, the stage game supports a "babbling equilibrium" in which the players' messages reveal no information about their types. Contribution strategies are then the same as if there were no communication at all. Babbling equilibria will be important in the analysis of the repeated game, where honesty in the short run will be supported by the threat of no (meaningful) communication in the future. The following result states the existence of a babbling equilibrium. Its proof, as well as all subsequent proofs, is in the Appendix.

Proposition 1. *A babbling equilibrium of the stage game exists.*

I am interested in when communication can lead to efficient public good provision outcomes. I define an *efficient provision equilibrium* as an assessment that meets the following conditions:

(C1) It is a perfect Bayesian equilibrium.

(C2) The public good is supplied whenever feasible—namely, whenever $\omega_1 + \omega_2 \geq 2$.

(C3) There are no wasted contributions on the path of play; either $x_1 + x_2 = 0$ or $x_1 + x_2 = 2$.

(C4) The cost of contributions is borne by those most willing to pay: if $x_i > 0$, then $\omega_i = \max\{\omega_i, \omega_j\}$.[3]

Obviously, efficient provision requires some degree of communication. Take, for example, a player of type $\omega_i = 1$, who is willing to contribute at most one unit of effort to the project. If her partner's type is $\omega_j = 0$, then the project is infeasible and efficient provision requires that both give nothing. If her partner's type is $\omega_j = 1$, then supply whenever feasible (C2) requires that each give 1, the most she is willing. Finally, if her partner's type is $\omega_j = 2$, efficient distribution of costs (C4) requires that player i give nothing. Therefore, we cannot have efficient provision without information

[2] Because the contribution game is a potential game (Monderer and Shapley 1996), the restriction to pure strategies is innocuous. See the proof of Proposition 1.

[3] See Palfrey, Rosenthal and Roy (2015) on efficient cost distribution in private-information public goods games.

transmission. In fact, as the following result states, at least one player must fully reveal her type.

Lemma 1. *In any efficient provision equilibrium, at least one player's messaging strategy is fully separating.*

If the players' types were revealed publicly, efficient provision would be possible in equilibrium. But is it always in a player's interest to voluntarily reveal her type? Unfortunately, it is not. To see why, take an efficient provision equilibrium and consider the highest type of player, $\omega_i = 2$. If she reveals her type honestly, she must pay the full cost of provision if her partner's type $\omega_j < 2$. If she were instead to announce that her type were $\omega_i = 1$, then she would still pay the full cost of contribution with a type-0 partner, but would pay only half with a type-1 partner and nothing with a type-2 partner. Either way, provision is assured, but the player bears strictly less of the cost (in expectation) by downplaying her willingness to give. This incentive to misrepresent undercuts the possibility of an efficient provision equilibrium in the stage game.

Proposition 2. *There is no efficient provision equilibrium of the stage game.*

This result accords with previous findings that cheap talk communication can play at best a limited role in one-shot threshold problems: it can separate those who are willing to contribute at all from those who are not, but cannot lead to efficient cost distribution conditional on contribution (Kenkel 2016).

In the one-shot context, efficient provision is not sustainable as equilibrium behavior. It is too tempting for high types to understate their information, so as to bear less of the costs of contribution. In the remainder of the paper, I investigate whether introducing the shadow of the future—and, with it, the possibility of punishment for being caught in a lie—might make efficient provision feasible.

2.2 Repeated Play

Assume the players interact over infinitely many periods, indexed $t = 0, 1, \ldots$, and discount the future according to a common discount factor $\delta \in [0, 1)$. At the beginning of each period, each player's period-specific type $\omega_i^t \in \Omega_i$ is drawn by Nature. I assume that types are independent and identically distributed across players and periods and that a player's period-t type only affects her period-t payoff. As before, players only learn their own types, though the distribution of types is common knowledge.

After types are drawn, play proceeds according to the stage game. First, each player simultaneously chooses a message $m_i^t \in M_i$. Then, the players observe each other's messages and update their beliefs accordingly. Finally, each player simultaneously chooses a contribution $x_i^t \in X_i$. Given a sequence of type realizations $\{\omega_i^t\}_{t=0}^{\infty}$ and sequences of contributions $\{x_i^t\}_{t=0}^{\infty}, \{x_j^t\}_{t=0}^{\infty}$, player i's discounted utility in the repeated game is

$$\sum_{t=0}^{\infty} \delta^t u_i(x_i^t, x_j^t, \omega_i^t). \tag{2}$$

Since the stage payoffs are uniformly bounded and $|\delta| < 1$, the above series converges.

An efficient provision equilibrium in the repeated game is a perfect Bayesian equilibrium[4] in which, with probability 1 along the equilibrium path, play in each period satisfies the efficient provision requirements (C1)–(C4).

3 Results

To sustain an efficient provision equilibrium, I build a strategy profile that relies on a "grim babbling" punishment strategy. In each normal period, each player honestly reveals her type, and then contributions proceed according to the efficient provision requirements. If a player's contribution does not match what she was supposed to give, then the game moves to a punishment stage that never ends (hence "grim"). In the punishment stage, all messages are uninformative (hence "babbling"), and players contribute according to a Nash equilibrium of the game without communication. These contributions, as we already saw in Proposition 2, will necessarily be inefficient. A formal description of this strategy profile appears in Definition 1 in the Appendix.

A subtle but remarkable result here is that as long as the prior probability of the lowest-willingness type is sufficiently high (a condition whose importance will come up again soon), the choice of contribution equilibrium

[4]**Technical note:** Since there is not, to my knowledge, a standard definition of perfect Bayesian equilibrium in infinitely repeated games with period-dependent types, I am self-consciously waving my hands a little bit for now about what exactly this means.

in the punishment stage does not matter. In other words, the efficacy of the punishment does not depend on holding players to an unrealistically low contribution strategy. The loss of efficiency due to hindering future communication is enough to make players prefer to be honest, assuming they value the future sufficiently highly. The following result establishes the conditions under which players *ex ante* (i.e., before their types are drawn) strictly prefer symmetric efficient provision over any babbling equilibrium of the stage game.

Lemma 2. *If $p_0 \geq c_2/(1-c_2)$, the* ex ante *expected utility of symmetric efficient provision in the stage game strictly exceeds that of any babbling equilibrium of the stage game.*

The *ex ante* superiority of efficient provision over any Nash equilibrium is enough to deter any sufficiently patient player from making an observable deviation. Once a player has announced her type honestly, she is effectively committed to follow through. However, there is still the threat of *unobservable* deviations. In particular, a country may announce some other type in the messaging stage and then behave like that type is supposed to in the contribution stage. For efficient provision to be an equilibrium, it cannot ever be in a player's interest to lie and hide it.

The highest-willingness type, $\omega_i^t = 2$, is the most likely to have an unobservable profitable deviation available. Under symmetric efficient provision, a type-2 player must take on all the provision costs if her partner is type 0 or 1, and half if her partner is type 2. If she pretended to be type 1, then she would save half of the costs with a type-1 or type-2 partner. The downside is when she has a type-0 partner. Since her partner will not contribute, in the short run the type-2 player would rather supply the good herself, since $2c_2 < 1$. Doing so, however, would reveal her as a liar and send the game into grim babbling, making it unprofitable if she is sufficiently patient. Consequently, whether it is profitable for a type-2 player to lie (and hide it in the contribution stage) comes down to whether the likelihood of a type-0 partner is high enough to offset the gains she would make from type-1 and type-2 partners. This is true if and only if $p_0 \geq c_2/(1-c_2)$, which gives us the following result.

Proposition 3. *If $p_0 \geq c_2/(1-c_2)$ and players are sufficiently patient, there is an efficient provision equilibrium of the repeated game.*

An interesting feature of this equilibrium is that players might be punished for giving too much, relative to the amount they claimed to be willing to contribute. This is counterintuitive—how can we encourage public goods provision by punishing those who contribute more than expected? The key is that it incentivizes honesty. If players were not punished for contributing more than they said they were willing to, then there would be no incentive for high-willingness players to reveal their types honestly.

In the constructed equilibrium, the punishment phase entails forever playing a babbling equilibrium of the stage game. In effect, once the punishment phase is entered, players neither talk nor listen. This might raise a question about credibility. A player who has caught her opponent in a lie may be able to commit herself not to talk, but can she commit not to listen? In other words, if we allowed for a punishment phase wherein only the liar continued to send informative messages, would the punishment still deter lying? I find that it would. As long as the player who has been caught lying cannot condition her contribution on her partner's type, which is true as long as the partner employs a babbling strategy, then she would *ex ante* strictly prefer efficient provision under the conditions of the proposition.

4 Welfare Comparison

Traditional mechanisms for encouraging international cooperation propose punishing players for failing to contribute (Axelrod 1984; Keohane 1984). The mechanism I have proposed here differs in two important ways. First, players are punished for dishonesty, not for failing to contribute. A player can trigger the punishment phase by contributing more than she said she could. Second, and consequently, completion of the project is not assured in each period. If the players' announced types do not sum to at least 2, then the project is foregone. Under what conditions is this somewhat peculiar mechanism preferable to traditional means?

I consider the class of alternative strategy profiles in which, with probability 1 along the path of play, each player contributes $x_i = 1$ to the project. I call these *egalitarian cooperation* strategies. Under such a strategy profile, the provision threshold is always met along the path of play. We know that this cannot be an equilibrium of the stage game, since contributing $x_i > 0$ is strictly dominated for type-0 players.[5] To make the analysis as favor-

[5]Importantly, this means Lemma 2 does not apply, so it is possible for egalitarian

able as possible to egalitarian cooperation, I do not restrict the analysis to the conditions, if any, under which egalitarian cooperation is sustainable as equilibrium behavior.

My metric for comparison is the *ex ante* expected stage utility under each mechanism. The *ex ante* expected utilities are constant across periods along the path of play for each mechanism, so comparing expected discounted total payoffs would yield the same answer. Since the game is symmetric, so would considering single-period or total *ex ante* social welfare. The *ex ante* expected stage payoff to a player in an efficient provision equilibrium is

$$U_i = p_0(p_2) + p_1(p_2 + p_1(1 - c_1)) + p_2(1 - (2 - p_2)c_2).$$

The *ex ante* expected stage payoff under egalitarian cooperation is

$$U_i' = 1 - p_0 c_0 - p_1 c_1 - p_2 c_2.$$

In the interim, egalitarian cooperation is strictly worse for type-0 players and strictly better for type-2 players. Whether it is better or worse for type-1 players depends on the relative proportions of type-0 and type-2 players; the more of the former, the better it is. Clearly, then, neither mechanism is preferable under all combinations of parameters $(c_0, c_1, c_2, p_0, p_1, p_2)$.

To visualize the welfare comparison, Figure 2 plots which mechanism is *ex ante* preferable as a function of prior probabilities and costs.[6] In the vast majority of the parameter space, efficient provision is preferable to egalitarian cooperation. The main condition for egalitarian cooperation to be preferable is that the probability of type-2 players be negligible (i.e., along the frontier of the triangle of values of p_0 and p_1) and that the cost of provision to type-1 players be relatively low.

Seeing as efficient provision is not always preferable to egalitarian cooperation, one might wonder whether we can have the best of both worlds. Imagine an equilibrium in which project completion is assured, as in egalitarian cooperation, but the costs are distributed efficiently across players, as in efficient provision. More precisely, define an *efficient assured completion* equilibrium as one in which, with probability 1 along the path of play:

- If $\omega_i > \omega_j$, then $x_i = 2$ and $x_j = 0$.

cooperation to be *ex ante* preferable to efficient provision.

[6] The plot fixes $c_0 = 1.1$; the results are qualitatively similar for other values of c_0. As c_0 (which is never paid under efficient provision) increases, the space under which egalitarian cooperation is preferable shrinks.

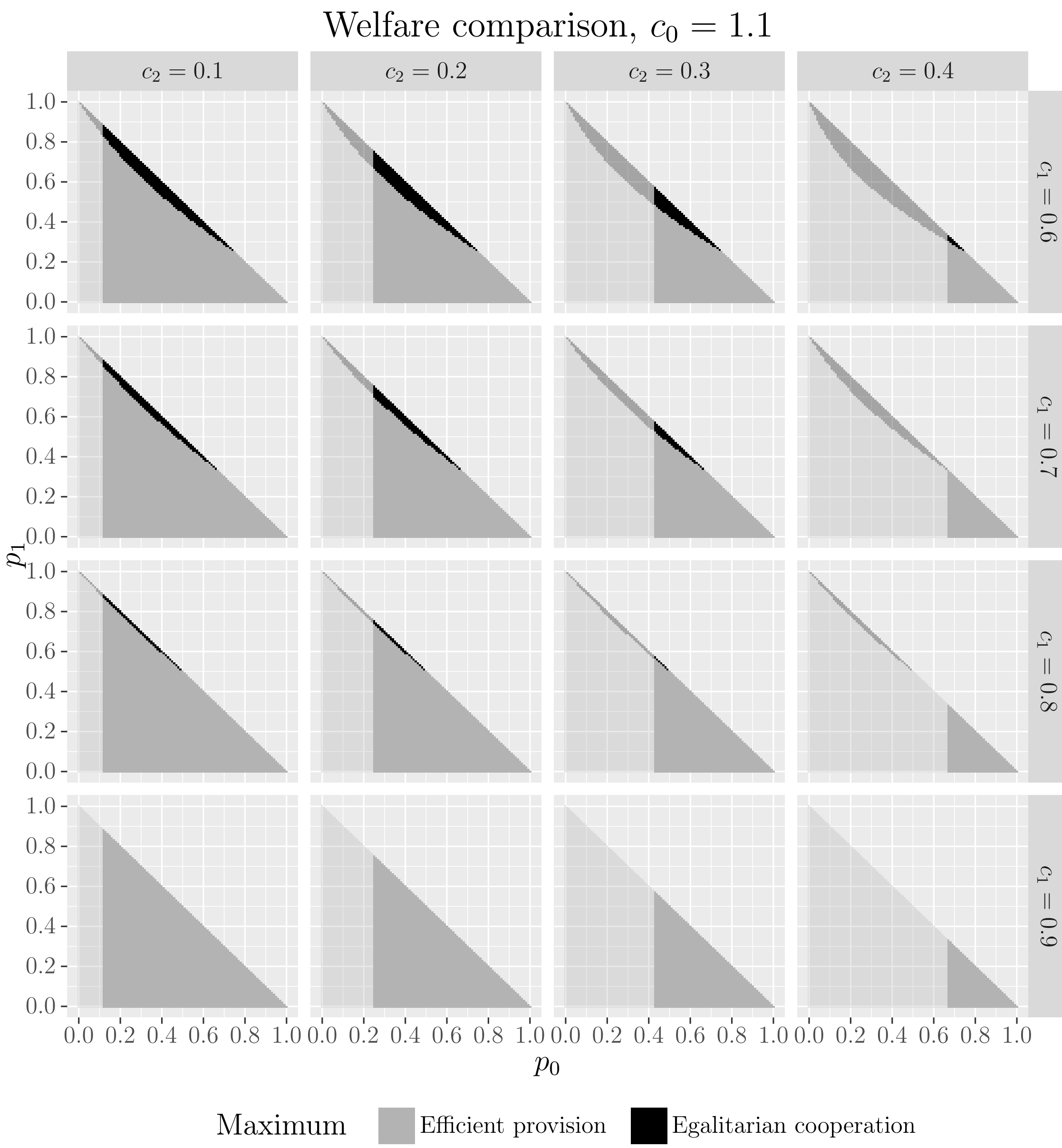

Figure 2: Welfare comparison for the efficient provision equilibrium versus an egalitarian cooperation strategy profile, in which each player contributes $x_i = 1$ each period. Translucent regions are those where an efficient provision equilibrium does not exist (i.e., $p_0 < c_2/(1 - c_2)$).

- If $\omega_i = \omega_j$, then $x_i = x_j = 1$.

Efficient assured completion gives us the best of both worlds, but it is too good to be true. Even in a repeated setting with highly patient players, it is not sustainable as equilibrium behavior, as the following result states.

Proposition 4. *There is no efficient assured completion equilibrium of the repeated game.*

The problem with efficient assured completion is that it gives high-willingness types no incentive to be honest. In the efficient provision equilibrium constructed for Proposition 3, what keeps a high type from lying is the threat that the project will not be completed if she feigns unwillingness and must thereafter mimic a lower type's contribution behavior. If the project's completion is assured, there is no such threat to keep high-type players from lying. In this case, understating one's type shifts more of the cost burden onto one's partner without affecting the chance of provision. Therefore, there is never a best of both worlds equilibrium.

5 Ideas for Future Work

The main idea I have had so far concerns the case where an efficient provision equilibrium does not exist, i.e., $p_0 < c_2/(1-c_2)$. In fact, let $p_0 = 0$, so that egalitarian cooperation is now an equilibrium of the stage game. Is there some way we can attain efficient cost distribution in this setting, or are we doomed by the "no best of both worlds" result?

My idea for enhancing efficiency is with a "countdown" equilibrium. The prior distribution of types is common knowledge, so we know roughly how often there "should" have been a high-type player. For example, if $p_1 = p_2 = 0.5$, then there's only a $1/64$ chance of having three periods in a row where neither has $\omega_i^t = 2$. So imagine a strategy profile where, if we go T periods without someone announcing that they're the high type, then we revert to grim babbling with egalitarian contribution. The lower T is, the less the efficiency gain relative to egalitarian contribution, since then we have all the more likelihood of reaching a terminal period where low types overstate their willingness to avoid moving into a punishment phase. On the other hand, the higher T is, the harder it becomes for incentive compatibility to hold (at least early in the countdown). What would impel honesty in the short run is that you'd rather absorb the costs of being a high type now when you actually are a high type than have to do it to avoid punishment later when you're a low type. So there's probably a Goldilocks solution where you want T to be high enough to drive efficiency gains but not so high that it breaks incentive compatibility.

A Proofs

A.1 Proof of Proposition 1

Proposition 1. *A babbling equilibrium of the stage game exists.*

Proof. Let $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ be a Bayesian Nash equilibrium of the contribution game in which each player's beliefs are given by the prior, where each $\tilde{\sigma}_i : \Omega_i \to X_i$. This is a Bayesian potential game (Monderer and Shapley 1996; van Heumen et al. 1996) with exact potential function $P(x_1, x_2, \omega_1, \omega_2) = \mathbf{1}\{x_1 + x_2 \geq 2\} - c(\omega_1)x_1 - c(\omega_2)x_2$, as in Myatt and Wallace (2009). Since the type and action spaces are finite, a pure strategy equilibrium exists (van Heumen et al. 1996). For each $i \in \{1, 2\}$ and $\omega_i \in \Omega_i$, let $\mu_i(\omega_i) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $\sigma_i(\omega_i, m_i, m_j) = \tilde{\sigma}_i(\omega_i)$, and $\lambda_i = (p_0, p_1, p_2)$. It is straightforward to verify that $(\mu_i, \sigma_i, \lambda_i)_{i=1,2}$ is a babbling equilibrium of the stage game. □

A.2 Proof of Proposition 2

Lemma 1. *In any efficient provision equilibrium, at least one player's messaging strategy is fully separating.*

Proof. Take any efficient provision equilibrium, and let m_i be a message sent with positive probability by some type of player i. Let m_j be any message sent with positive probability by type 1 of player j.

First, suppose $\{0, 1\} \subseteq \operatorname{supp} \lambda_j(m_i)$. If $\sigma_j(1, m_j, m_i) = 0$, then provision whenever feasible (C2) fails when $\omega_i = 1$. Consequently, $\sigma_j(1, m_j, m_i) = 1$, which means no wasted contributions (C3) fails when $\omega_i = 0$. Therefore, $\{0, 1\} \not\subseteq \operatorname{supp} \lambda_j(m_i)$.

Next, suppose $\{1, 2\} \subseteq \operatorname{supp} \lambda_j(m_i)$. By Bayes' rule, $1 \in \operatorname{supp} \lambda_i(m_j)$. Efficient distribution of costs (C4) thus implies $\sigma_i(2, m_i, m_j) = 2$. No wasted contributions (C3) then requires $\sigma_j(1, m_j, m_i) = 0$, in which case provision whenever feasible (C2) is violated when $\omega_i = 1$. Therefore, $\{1, 2\} \nsubseteq \operatorname{supp} \lambda_j(m_i)$.

So far we have seen that, for each $m_i \in M$, $1 \in \operatorname{supp} \lambda_j(m_i)$ implies $\operatorname{supp} \lambda_j(m_i) = \{1\}$. By the same token, $1 \in \operatorname{supp} \lambda_i(m_j)$ implies $\operatorname{supp} \lambda_i(m_j) = \{1\}$. Consequently, the only way the equilibrium may not be fully separating for either player is if there exist $m_i', m_j' \in M$ such that $\operatorname{supp} \lambda_i(m_j') = \operatorname{supp} \lambda_j(m_i') = \{0, 2\}$. In that case, however, provision whenever feasible (C2) requires $\sigma_i(2, m_i', m_j') = \sigma_j(2, m_j', m_i') = 2$, which in turn means no wasted contributions (C3) is violated with positive probability. Therefore, at least one player's messaging strategy must be fully separating. $\square$

Proposition 2. *There is no efficient provision equilibrium of the stage game.*

Proof. Suppose the contrary. By Lemma 1, at least one player's messaging strategy is fully separating; without loss of generality, let this be player 1, so that each $\mu_1(\omega_1)$ places probability 1 on ω_1. From the proof of Lemma 1, we have that type 1 of player 2 separates herself, so for each m_2 sent on the equilibrium path, $\operatorname{supp} \lambda_1(m_2) \in \{\{0\}, \{1\}, \{2\}, \{0, 2\}\}$.

By the conditions of efficient provision, contributions on the path of play given each player's message are as in the following table.[7]

	$m_1 = 1$	$m_1 = 2$
$\operatorname{supp} \lambda_1(m_2) = \{0\}$	$x_1 = 0, x_2 = 0$	$x_1 = 2, x_2 = 0$
$\operatorname{supp} \lambda_1(m_2) = \{1\}$	$x_1 = 1, x_2 = 1$	$x_1 = 2, x_2 = 0$
$\operatorname{supp} \lambda_1(m_2) = \{2\}$	$x_1 = 0, x_2 = 2$	$x_1 = 0, x_2 = 2$
$\operatorname{supp} \lambda_1(m_2) = \{0, 2\}$	$x_1 = 0, x_2 = \omega_2$	$x_1 = 2, x_2 = 0$

Player 2 never contributes less after receiving $m_1 = 1$ than after $m_1 = 2$, and contributes strictly more in case $\omega_2 = 1$. Since $p_1 > 0$, this means it is strictly profitable for type 2 of player 1 to deviate to sending $m_1 = 1$, contradicting the assumption of equilibrium. $\square$

[7]Efficient provision would allow for any contribution profile such that $x_1 + x_2 = 2$ when it is common knowledge that $\omega_1 = \omega_2 = 2$. The one given here is the most favorable to player 1, so changing the proposed contributions for this case would not change the existence of a profitable deviation for player 1.

A.3 Proof of Proposition 3

I begin by formally defining the strategy profile that I will claim constitutes an equilibrium under the conditions of the proposition.

Definition 1 (Automaton representation of efficient provision supported by grim babbling)**.** I define a strategy profile through a finite automaton (see Mailath and Samuelson 2006, 29–31) by introducing a state space $\mathcal{S}$, an initial state $s^0 \in \mathcal{S}$, and a transition function that determines the state at time $t+1$ as a function of the state and actions at time t; and by augmenting the functions that comprise an assessment in the stage game with an argument representing the current state.

- State space $\mathcal{S} = \{h, b\}$ (*h*onest and *b*abbling).
- Initial state $s^0 = h$.
- Messaging strategies

$$\mu_i(s^t, \omega_i^t) = \begin{cases} (1,0,0) & s^t = h, \omega_i^t = 0, \\ (0,1,0) & s^t = h, \omega_i^t = 1, \\ (0,0,1) & s^t = h, \omega_i^t = 2, \\ (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) & s^t = b. \end{cases}$$

- Belief systems

$$\lambda_i(s^t, m_j^t) = \begin{cases} (1,0,0) & s^t = h, m_j^t = 0, \\ (0,1,0) & s^t = h, m_j^t = 1, \\ (0,0,1) & s^t = h, m_j^t = 2, \\ (p_0, p_1, p_2) & s^t = b. \end{cases}$$

- Contribution strategies

$$\sigma_i(s^t, \omega_i^t, m_i^t, m_j^t) = \begin{cases} 0 & s^t = h, m_i^t = 0, \\ 1 & s^t = h, m_i^t = 1, m_j^t = 1, \\ 0 & s^t = h, m_i^t = 1, m_j^t \neq 1, \\ 2 & s^t = h, m_i^t = 2, m_j^t \neq 2, \\ 1 & s^t = h, m_i^t = 2, m_j^t = 2, \\ \tilde{\sigma}_i(\omega_i^t) & s^t = b, \end{cases}$$

where $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ is a Bayesian Nash equilibrium of the contribution game with beliefs equal to the prior, as in the proof of Proposition 1.

- Transition function

$$\tau(s^t, m_1^t, m_2^t, x_1^t, x_2^t) = \begin{cases} h & s^t = h, x_i^t = \sigma_i(h, m_i^t, m_i^t, m_j^t) \text{ for } i = 1, 2, \\ b & \text{otherwise.} \end{cases}$$

$\square$

Lemma 2. *If* $p_0 \geq c_2/(1 - c_2)$, *the* ex ante *expected utility of symmetric efficient provision in the stage game strictly exceeds that of any babbling equilibrium of the stage game.*

Proof. The interim expected stage utility to each type of player i under symmetric efficient provision is

$$U_i(\omega_i) = \begin{cases} p_2 & \omega_i = 0, \\ p_1(1 - c_1) + p_2 & \omega_i = 1, \\ 1 - (2 - p_2)c_2 & \omega_i = 2. \end{cases}$$

Let $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ be a Bayesian Nash equilibrium of the stage game without communication, and let $\tilde{U}_i : \Omega_i \to \mathbb{R}_+$ denote the interim expected utility of each type under this equilibrium. Each $\tilde{\sigma}_i$ must not be strictly dominated, so $\tilde{\sigma}_i(\omega_i) \leq \omega_i$ for all $\omega_i \in \Omega_i$. Therefore, the expected payoff to type 0 of player i satisfies

$$\tilde{U}_i(0) = E[u_i(0, \omega_j, 0)] = p_2 = U_i(0).$$

The expected payoff to type 1 satisfies

$$\begin{aligned} \tilde{U}_i(1) &\leq \max\{E[u_i(0, \omega_j, 1)], E[u_1(1, \omega_j, 1)]\} \\ &= \max\{p_2, p_1 + p_2 - c_1\} \\ &< p_2 + (1 - c_1)p_1 \\ &= U_i(1), \end{aligned}$$

where the strict inequality holds because $c_1 < 1$ and $p_1 < 1$.

For type 2, there are three possible best responses to consider. The condition $p_0 \geq c_2/(1-c_2)$ is equivalent to $c_2 \leq p_0/(1+p_0)$, which in turn implies

$$c_2 \leq \frac{p_0+p_1}{1+p_0+p_1} = \frac{1-p_2}{2-p_2}.$$

This inequality implies

$$\begin{aligned} E[u_i(0,\omega_j,2)] &= p_2 \\ &\leq 1-(2-p_2)c_2 \\ &= U_i(2). \end{aligned}$$

Similarly, $c_2 \leq p_0/(1+p_0)$ implies $c_2 \leq p_0/(p_0+p_1)$, which in turn gives

$$\begin{aligned} E[u_i(1,\omega_j,2)] &= 1-p_0-c_2 \\ &\leq 1-(2-p_2)c_2 \\ &= U_i(2). \end{aligned}$$

Finally, we have

$$\begin{aligned} E[u_i(2,\omega_j,2)] &= 1-2c_2 \\ &\leq 1-(2-p_2)c_2 \\ &= U_i(2). \end{aligned}$$

Therefore, no best response leaves type 2 of player i better off than under symmetric efficient provision:

$$\tilde{U}_i(2) = \max_{x_i \in X_i}\{E[u_i(x_i,\omega_j,2)]\} \leq U_i(2).$$

Since the interim utility is always weakly greater under efficient provision, strictly so for type 1, and there is positive probability of type 1, the *ex ante* utility of efficient provision strictly exceeds that of the given stage game equilibrium. □

Corollary 1. *If $p_0 \geq c_2/(1-c_2)$ and players are sufficiently patient, a one-stage deviation from the contribution strategies in Definition 1 is never profitable.*

Proof. Let U_i^h denote the *ex ante* expected utility to player i from honest stages ($s^t = h$), and let U_i^b be the same for babbling stages ($s^t = b$). By Lemma 2, $U_i^h > U_i^b$. Let v_i denote player i's stage payoff under the given strategy, and let $\hat{v}_i$ denote the stage payoff from a deviation in the contribution stage. The deviation is profitable only if

$$\hat{v}_i + \frac{\delta}{1-\delta} U_i^b > v_i + \frac{\delta}{1-\delta} U_i^h.$$

Since stage payoffs are bounded, this inequality cannot hold if δ is sufficiently close to 1. □

Lemma 3. *There is an unobservable profitable deviation from the messaging strategies in Definition 1 if and only if $p_0 < c_2/(1-c_2)$.*

Proof. Let $U_i(m_i \mid \omega_i)$ denote the expected stage payoff to type ω_i of sending m_i, given that she will follow the prescribed contribution strategies. By sending $m_i = 1$, player i induces each type of player j to give her type—the best feasible outcome for player i. Therefore, we need only check that $U_i(1 \mid 0) \leq U_i(0 \mid 0)$ and $U_i(1 \mid 2) \leq U_i(2 \mid 2)$. We have

$$U_i(1 \mid 0) = p_1 + p_2 - p_1 c_0 \leq p_2 = U_i(0 \mid 0),$$

where the inequality follows because $c_0 > 1$. The condition for a profitable deviation for type 2 is

$$U_i(1 \mid 2) = p_1 + p_2 - p_1 c_2 > 1 - (2 - p_2) c_2 = U_i(2 \mid 2),$$

which holds if and only if $p_0 < c_2/(1-c_2)$. □

Proposition 3. *If $p_0 \geq c_2/(1-c_2)$ and players are sufficiently patient, there is an efficient provision equilibrium of the repeated game.*

Proof. I claim that the strategy profile represented by the automaton in Definition 1 is an efficient provision equilibrium. Obviously, it satisfies the conditions of efficient provision, (C1)– (C4). Under the conditions of the proposition, there is no profitable deviation, per Corollary 1 and Lemma 3. The specified beliefs are consistent with the application of Bayes' rule whenever possible. Therefore, the claim holds. □

A.4 Proof of Proposition 4

Proposition 4. *There is no efficient assured completion equilibrium of the repeated game.*

Proof. Consider type 2 of player i in an efficient assured completion strategy profile. Her interim expected stage payoff, using the notation of the proof of Lemma 3, is

$$U_i(2 \mid 2) = 1 - (2 - p_2)c_2.$$

If she deviated to sending $m_i = 0$ and then followed type 0's contribution strategy, her interim expected stage payoff would be

$$U_i(0 \mid 2) = 1 - p_0 c_2 > U_i(2 \mid 2).$$

Since the deviation is unobservable, it does not affect stage transitions and thus is profitable. Therefore, the strategy profile is not an equilibrium. $\square$

References

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

Crawford, Vincent P and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6):1431–1451.

Kenkel, Brenton. 2016. "The Limits of Communication in Collective Action Problems." Typescript, Vanderbilt University.

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

Mailath, George J. and Larry Samuelson. 2006. *Repeated Games and Reputations: Long-Run Relationships.* Oxford: Oxford University Press.

Monderer, Dov and Lloyd S Shapley. 1996. "Potential Games." *Games and Economic Behavior* 14(1):124–143.

Myatt, David P and Chris Wallace. 2009. "Evolution, Teamwork and Collective Action: Production Targets in the Private Provision of Public Goods." *The Economic Journal* 119(534):61–90.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge: Harvard University Press.

Palfrey, Thomas, Howard Rosenthal and Nilanjan Roy. 2015. "How Cheap Talk Enhances Efficiency in Threshold Public Goods Games." *Games and Economic Behavior* .

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics* 36(4):387.

Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56(1):121–149.

Stone, Randall W, Branislav L Slantchev and Tamar R London. 2008. "Choosing How to Cooperate: A Repeated Public-Goods Model of International Relations." *International Studies Quarterly* 52(2):335–362.

van Heumen, Robert, Bezalel Peleg, Stef Tijs and Peter Borm. 1996. "Axiomatic characterizations of solutions for Bayesian games." *Theory and Decision* 40(2):103–129.