Anonymous Rituals

David Hugh-Jones* and David Reinstein†

December 10, 2009

Abstract

Religion and ritual have been characterized as costly ways for conditional cooperators to signal their type, and thus identify and interact with one another. But an effective signal may be prohibitively expensive: if the cost of participation is too small, free-riders may send the signal and behave selfishly later. However, if the ritual reveals only the *average* level of signaling in a group, free-riders can behave selfishly without being detected, and even a low cost signal can separate types. While individuals cannot be screened out, members can learn the group's profile of types. Under specified conditions, this information gain leads to greater cooperation and hence increases expected welfare. Furthermore, if crowding is unimportant relative to the conditional cooperation term, anonymous rituals will be preferred to ones which reveal individuals' behavior. Examples of anonymous institutions include church collections, voting, music, dance, and military customs.

Keywords: signaling, anonymity, public goods, religion, social capital, trust.

JEL codes: H41, Z12, D82.

*Max Planck Institute for Economics, Kahlaische Strasse 10, 07745 Jena, Germany. Tel: +49 3641 686 646. Fax: +49 3641 686990. Email: hugh-jones@econ.mpg.de.

†Corresponding author. Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom. Tel: +44 1206 873518. Fax: +44 1206 872724. Email: drein@essex.ac.uk.

1 Introduction

Societies and organizations often encourage their members to take part in activities – such as religious services, rain dances and football matches – that have little or no direct material benefit to either the individual or the group. Why do these ”rituals” persist, and why are they so widespread? One recent explanation is that rituals are a form of costly signaling (Weber, 1946; Ruffle and Sosis, 2003; Bird and Smith, 2005). People take part to show that they are committed to the social group, or to a particular collective project. This has two benefits: first, less committed people can be excluded from the group or project; second, group members learn about each other, and this can reduce uncertainty and allow them to coordinate. Both these benefits help explain findings like those of Sosis and Bressler (2003), who show that 19th century religious communes that imposed stricter requirements on members' behavior survived for longer, and Sosis and Ruffle (2003), who find that members of religious *kibbutzim* (communes) cooperate more than others in an anonymous game.

For a ritual to yield an informative signal, it must “separate” the committed and the uncommitted; each must behave differently. Often, people gain from being perceived as committed even if they are not, so ritual behavior must be costly enough to dissuade this “pooling”. In particular, if not participating in the ritual gets one excluded from the group, the ritual must cost enough to deter uncommitted individuals from taking part, even despite this punishment. However, these costs are then a burden on all group members. While a commune with too lax rules might be invaded by free-riders, a commune with rules that are too strict may be unpleasant for everyone.

But even if non-participants are not excluded, rituals may still benefit the group. Signaling provides two kinds of information: individuals' types are revealed, but so is the composition of the group as a whole. Some rituals may reveal aggregate behavior but keep individuals' actions hidden. Although this rules out screening (exclusion), the information gain may still lead to greater cooperation and hence higher expected welfare. At the same time, the ritual need not be so costly, since not participating does not automatically lead to exclusion. Our paper is about this kind of anonymous rituals.

To understand the benefit of this information gain, we need to delve into the theory of signaling in the context of collective action. As extensively modeled by economists and game theorists, cooperation in social dilemmas may be assured by binding agreements, or by an equilibrium in a repeated game (Fudenberg and Tirole, 1991). But it is also widely noted that such agreements may not be enforceable – for instance, in illegal enterprises, in stateless societies or those where legal institutions are inefficient, in chaotic situations (such as wars and humanitarian disasters), or when performance cannot reasonably be verified by third parties. Similarly, when the environment is uncertain, the temptation of immediate defection may regularly outweigh the "shadow of the future", so that high levels of cooperation cannot be sustained by a repeated game. Or, actions may be imperfectly observable, again preventing repeated games from achieving efficiency. In these situations, the choice to cooperate can be thought of as a one-shot game, which we will sometimes call the "basic collective action problem (CAP)".[1]

However, defection need not be a dominant strategy. Instead, there may be heterogeneity, with some selfish players and some who are willing to cooperate if they expect others to do so. Both heterogeneity and conditional cooperativeness may have material foundations: some players may have better outside options from defection, cooperation may have increasing returns, or the CAP may directly reward "investors." Alternatively, these preferences may be part of human psychology (Berg et al., 1995; Kreps et al., 2001; Schram, 2000; Ostrom, 2000a); this "heterogeneous reciprocity" has been well documented in economic experiments (Keser and van Winden, 2000; Simpson and Willer, 2008; Fischbacher et al., 2001).

In this environment, conditionally cooperative players will wish to learn the other players' preferences (technically, their "type"), since it will help them decide how much to cooperate. However, even where individuals cannot be punished directly, players will not honestly reveal their types through cheap talk: selfish types might pretend to be cooperative so as to induce

[1]The one-shot structure should not be taken too literally. The basic CAP could also represent an ongoing situation in which group members' actions are unobservable, so that players cannot condition on others' choices.

more cooperation from the others. Instead, honest type revelation requires a *costly* action which only cooperative types prefer to take. The benefit is the resulting increase in others' cooperation; if cooperative types gain more from this than selfish types, separation can be achieved (as in Austen-Smith and Banks (2000)).[2]

So, under these conditions – uncertainty, limited external enforcement, heterogeneous reciprocity, and when conditional cooperators also value the collective good more – rituals may be useful as costly signaling devices even where exclusion is impossible. If the information these rituals provide about types increases cooperation enough on average, they will be socially beneficial. Then groups with these rituals will achieve higher welfare than groups without them, and we might expect rituals either to be deliberately introduced, or to evolve by chance and to spread via imitation or via differences in group survival. [3]

In this context, the *anonymity* of a ritual may serve as a commitment device, and help resolve a time-inconsistency problem. As suggested above, in addition to reducing others' cooperation, a selfish player may suffer directly from revealing his type. Members may exclude selfish types from the basic CAP to avoid crowding. For example, those not conforming to the rules of a religious commune may be expelled, and might face bleak prospects in the outside labor market. Individuals may also simply enjoy punishing bad people (Bolton and Zwick, 1995; Fehr and Gachter, 2000; Ostrom et al., 1992), or they may avoid selfish types in one-to-one interactions (Coricelli et al., 2004). The higher the penalty for not sending the signal, the more costly the signal needs to be in order to separate types. (And ritual practices can indeed be very costly. For example, the followers of Saint Simeon Stylites would spend months or years living on top of a pillar; this practice survived for centuries.)

Thus, while group members may be better off *ex ante* by not punishing or excluding the

[2]This condition, though often plausible, need not hold. Selfish players could benefit more from others' cooperation, while simultaneously getting less utility from cooperating themselves. Cf. Lemma 1 below. Separation may also be achievable if (as in the well known Spence (1973) model) the good types find it less costly (or more directly beneficial) to take this "ritual" action (see [*Authors' other paper*]).

[3]As noted by Ostrom (2000b), rituals may also be important in groups organized to preserve a common resource. She notes that membership in such groups may "involve complex rituals and beliefs that help solidify individual beliefs about the trustworthiness of others."

uncommitted, so that types can separate at lower cost, *ex post* they may still want to do so. To reconcile this, societies can tie their hands by using an *anonymizing technology* that reveals the overall level of cooperation (in the ritual), but keeps individual behavior hidden. The cost of revealing one's type is then limited to its effect on others' cooperation. The social welfare comparison between anonymous and non-anonymous rituals (essentially, between allowing and disallowing exclusion) will depend on the trade-off between the signaling cost and the benefits of revealing individuals' types. If crowding is extremely costly, or if individuals benefit greatly from avoiding uncommitted types, then anonymity will not be worthwhile.

In section 3 we develop a model of anonymous rituals and show how and when they might survive, in terms of some key parameters: the benefit from reducing uncertainty about the overall composition of the group; the benefit from learning about particular individuals' types, and the cost to individuals who are exposed as uncommitted.[4] Before this, we motivate our model with some empirical examples, and survey the related literature.

2 Examples of anonymity-preserving institutions

Costly signaling explanations for religious behavior have been widely advanced, so we look first here for examples of anonymity-preserving institutions. Notably, donations to churches are often anonymous. Supporters for anonymity find a Biblical justification in Matthew 6:2-4, which references the different motivations of public and private givers:

> Therefore when thou doest thine alms, do not sound a trumpet before thee, as the hypocrites do in the synagogues and in the streets, that they may have glory of men.... But when thou doest alms, let not thy left hand know what thy right hand doeth....

Some churches use "pledge cards" and collection plates, which are generally thought to increase the visibility of donations, but many denominations do not (Hoge et al., 1996). As a modern Baptist says:

[4] In a companion paper ([authors' other paper]) we test our model in a laboratory public goods experiment. This is summarized briefly in the Appendix.

> Many of our lay leadership like the idea of anonymity.... Maybe people are not giving as much as their neighbors, and they would be embarrassed.... In small rural churches, it is a difficult thing to get commitment cards signed.
>
> (quoted in Hoge, ibid.)

Churches with pledge cards appear to raise more donations (ibid.) so why the resistance? We theorize that a key good provided by churches is mutual support.[5] Donations not only raise funds, but signal individuals' commitment to the congregation, and thereby their likely future behavior. When individual donations are private but total levels are public, as in many churches, the overall level of "community spirit" (or "social capital") is known. Public donations would lead either to a pooling equilibrium in which this information would be lost, decreasing church members' mutual trust and solidarity, or to a separating equilibrium with higher donations – perhaps excessively high from the perspective of church members.

The tradition of anonymous giving has puzzled several economists (e.g., Andreoni and Petrie, 2004), since much existing work has focused on the disadvantages of anonymity: generosity and pro-social behavior increase when reputation is at stake (Harbaugh, 1998; Glazer and Konrad, 1996; Milinski et al., 2002; Cooter and Broughman, 2005; Andreoni and Petrie, 2004). Although anonymity may reduce giving by reputation-seekers, it may help to create mutual trust. Furthermore, early anonymous donations may possess greater signaling value in encouraging others to come forward (see List and Lucking-Reiley, 2002; List and Rondeau, 2003; Karlan and List, 2007; Reinstein and Riener, 2009 for related experimental evidence).

Voting can be a reliable signal of public support (Stigler, 1972; Londregan and Vindigni, 2006). Strike ballots are particularly likely to be a signal of resolve in industrial disputes. In the UK, legislation requiring a secret ballot before a strike was introduced during the 1980s. Unions found an unexpected advantage to the secret ballot: as voting is a costly activity, negotiators could use the level of turnout in such elections to persuade management of the strength of feeling among their members. The secret ballot may have provided more credible

[5] That is, churches are clubs (Iannaccone, 1992). For an in-depth examination of one church from the club goods perspective, see McBride (2007).

information than public demonstrations of support, in which some workers might have felt pressured to take part. It may also have increased members' willingness to take action after a successful ballot (Martin et al., 1991), and allowed union leaders to call off a strike when it was unlikely to succeed.[6]

As noted by Akerlof and Kranton (2000), "in the military, it is hard to observe effort, especially when it is most crucial—in battle" (and even if effort is observed, it is hard to reward a soldier if the whole regiment is massacred). But for a soldier who values being hailed as a hero, fighting bravely may be rational – but only if he expects others to join the fight in numbers strong enough to make victory likely. Hence, soldiers who value personal glory may be conditional cooperators, while those who do not are our model's "selfish" types. Kellett (1982) notes "Israelis regard fighting as very much a social act based on collective activity, cooperation, and mutual support." According to King (2006) (citing Randall (2004)) "British forces have deliberately sought to engage in ritualized forms of movement to encourage collective action." Such movements (marches, chants, drills) can be seen as anonymous signals. While an individual soldier's physically costly effort in a military drill may not be observable to others, the general cohesiveness and power of the exercise signals the unit's overall level of commitment to themselves and to others. According to an Israeli Defense Force spokesperson, a successful company commander must have "the ability to create mutual trust between the sub-commanders and the soldier."[7] An officer who is confident that his men are all "conditional cooperators" may employ such exercises to make this type distribution common knowledge.

Several cultural "technologies" seem well-designed to preserve anonymity. Hagen and Bryant (2003) suggest that song and music evolved as "coalitional signaling". Many traditional rituals include communal singing and dancing. Communal singing can be judged by

[6]Card and Olson (1995) note that "an increase in the fraction of the firm's employees involved in the strike increases both the probability of success and the wage increase conditional on a success." They also point out that in the early days of the labor movement, union leaders seemed to want "to prevent a strike when the likelihood of success was 'too low'."

[7]Soldiers' Rights Commissioner, IDF Spokesperson's Unit, "The Human Being's Human Doctrine," http://www.idf.il/english/organization/nakhal/kavod.stm.

its overall volume, harmony, and enthusiasm, while it is difficult to discover if any individuals are out of tune or keeping quiet. In communal dances, if one person makes a mistake, the entire group may lose the rhythm, and the guilty party can not always be identified. In a modern context, applause, cheers and jeers are all reliable signals of collective appreciation, since contribution levels cannot be distinguished.

We are not claiming that information gain is the sole reason that anonymity is preserved in the above examples. For instance, people may be concerned to protect low contributors from embarrassment, not only to avoid pooling as our theory suggests, but also out of empathy or a desire not to drive others away from the group. We are merely proposing a previously unexplored motive which may complement other explanations. Hence it is often difficult to categorize particular institutions as pure "anonymous signaling rituals." For example, consider a group of people going out to a restaurant, who agree to contribute "what they think they owe." This avoids working out each individual's debt. But, although it is not designed to measure the cooperativeness of the group, when contributions are counted up and compared to the total bill, the diners still learn something about whether their companions are generous or stingy on average. When all present learn that their peers are "nice," this may lead to future dinners out, and the group may eventually become a close circle of friends.

The advantages of anonymity have been discussed in the literature on principal-agent relationships (Holmstrom, 1999), as well as in a legislative context (Prat 2005; Levy 2007b; 2007a), though not to our knowledge in the context of large elections. With incomplete contracts and "career concerns" principals may benefit from committing not to learn too much; this will better align incentives and induce the agents to make more productive choices (Acemoglu, 2007). In our model the benefit of anonymity is more indirect: revelation of types can be achieved at a lower cost, leading in some cases to more efficient decisions by other agents.

As well as explaining existing institutions, we believe our theory has practical implications for the design of new ones. As an example, consider the finding by Karlan (2005) that behavior in a trust game predicts default by borrowers in a microcredit scheme. Bank lenders might wish to use this result to predict default risk. But if it became known that trustworthiness in a trust game was the way to get a loan, the game would quickly lose its predictive power (or would have to be played for such high stakes that it deterred many borrowers). However, many microcredit schemes rely on informal group enforcement to avoid default. So, groups of potential borrowers could be asked to play an anonymous public goods game. If the group size and incentives were rightly balanced, this would be an incentive-compatible way to revealing participants' character (in the aggregate). The game's overall results could be announced, and this could both aid banks in deciding how much to lend, and allow trust to develop among group members in cases where that trust is warranted.

3 Model

The society consists of N agents. Agent are of two types, which we call "committed" and "uncommitted", or "good" and "bad". A type $\tau \in \{G, B\}$ agent's prior belief is that there are g other good types with probability $\Pi^{\tau}(g) > 0$ for $g \in \{0, ..., N-1\}$. Sometimes we wish to increase N while holding the structure of the problem constant. We assume that as N grows large, Π^G approaches[8] Π^B, i.e. one's own type is not very informative about others' types, and the *proportion* of good types is distributed with continuous cdf $F(\gamma)$, supported on $\gamma \in [0,1]$.[9]

Let x_i represent the share of wealth i contributes to the CAP, $P \in [0,1]$ the proportion of agents included in the collective good, $X = \sum x_j$ total contributions, and $X_{-i} = \sum_{j \neq i} x_j$ represents contributions from players other than i. Unless otherwise specified the choice set is $x_i \in \mathbb{R}_+$. In the basic CAP, agents may contribute their wealth to a collective good, and

[8] In the space of the $N-1$-dimensional probability simplex.

[9] If agents' types are independent, then Π^G will equal Π^B for any N. But independence may not hold, for example because people's preferences are affected by those around them, or by collective culture or education.

individual i of type τ receives utility

$$w^{\tau}(x_i, X_{-i}, P). \tag{1}$$

Allowing additive separability and a flexible restriction on the nature of crowding, we specialize this as

$$w^{\tau}(x_i, X_{-i}, P) = \frac{\phi(X)}{P^{\delta}} - x_i + \beta_{\tau}\psi(x_i, X_{-i}). \tag{2}$$

This form yields some clearer proofs and intuition, without any interesting loss of generality. Nonetheless, where possible, we will give our results for the more general case. ϕ is a bounded, weakly increasing production function for the collective good, with $\phi(0) = 0$. $\delta \geq 0$ parametrizes the degree of crowding; when $\delta = 0$ the good is purely public. The ψ part of utility represents what is only received by good types: that is, $\beta_B = 0$ and $\beta_G = \beta > 0$. ψ, which we label "cooperation utility," could represent either psychological benefits from contributing or material benefits that only committed types receive.(For this reason, the labels "good" and "bad" should not be taken too seriously; similarly, we sometimes call $\phi(X) - x_i$ "material welfare" for convenience.) We assume ψ is differentiable and concave, and to ensure an interior equilibrium, that $\psi_1(x, X) \to \infty$ as $x \to 0$, $\psi_1(x, X) \to 0$ as $x \to \infty$.[10] Define the "target level" for a given contribution by others, $T(X_{-i})$, as the maximizer of $\psi(\cdot, X_{-i})$.[11]

We focus on the good type's welfare rather than on total welfare, as the former is likely to be more important to explaining the persistence of anonymous institutions, and our focus is positive rather than normative. Our argument here is informal, but motivated by evolutionary theory, in the vein of Frank (1987). Groups full of bad types are not likely to survive as they will produce little surplus. Hence the groups that survive will need to attract good types, and hence increase good types' welfare. Furthermore, within an institution's political process

[10] w_a represents the partial derivative of $w(x, X, P)$ with respect to the a'th argument. For example $w_2^{\tau} \equiv \partial w_2^{\tau}(x, X, P)/\partial X$. Cross-partials are represented by double subscripts. The conditions on ψ_1 could be relaxed but proofs would be more complex with little insight added.

[11] If there is more than one maximizer, then $T(\cdot)$ is set-valued.

changes that favor bad types are not likely to be proposed, as they reveal the advocate as a bad type, making her vulnerable to exclusion. For convenience, we write $w \equiv w^G$, and $w^\tau(x,X) \equiv w^\tau(x,X,1)$ for welfare when no agent is excluded.

We first examine the case without exclusion, so that $P = 1$ always. Before introducing signaling institutions, we examine equilibria in the basic CAP, first when g is unknown, second when it is common knowledge.

3.1 Equilibrium without exclusion

Suppose ϕ is concave with $\phi' < 1$. Then bad types never contribute. A good type's interior best response x_i will satisfy the first order condition $\int w_1(x,X_{-i})d\Phi(X_{-i}) = 0$, where $\Phi(\cdot)$ is the (continuous or discrete) probability distribution of others' total contributions. Plugging in the derivative of (2) we have

$$\int \left\{\phi'(X_i + x_i) + \beta\psi_1(x_i,X_{-i})\right\} d\Phi(X_{-i}) = 1,$$

or in the more general notation $\int w_1(x_i,X_{-i})d\Phi(X_{-i}) = 0$. When Φ puts 100% probability on X_{-i}, write $b(X_{-i}) \equiv x$ satisfying $\beta\psi_1(b(X_{-i}),X_{-i}) + \phi'(X_{-i} + b(X_{-i})) = 1$. If $w_{12} > 0$ then $b(\cdot)$ is weakly increasing by monotone comparative statics: other people's contributions make good types want to contribute more.

A symmetric interior equilibrium, in which good types contribute x^*, satisfies

$$\sum_{g=0}^{N-1} \Pi^G(g)\left\{\phi'((g+1)x^*) + \beta\psi_1(x^*,gx^*)\right\} = 1, \tag{3}$$

recalling that $\Pi^G(\cdot)$ gives a good type's distribution of the number of *other* good types.

Suppose now that the number of good types is known to be $g+1$. An interior equilibrium in which good types contribute x_g has average "others"' contributions of $gx_g/(N-1)$ and must simultaneously satisfy the best response condition $x = b(X)$; hence

$$\phi'((g+1)x_g) + \beta\psi_1(x_g, gx_g) = 1 \tag{4}$$

or equivalently $w_1(x_g, gx_g) = 0$. Again if $w_{12} > 0$ then x_g will be weakly increasing in g by strategic complementarities (Milgrom and Roberts, 1990).

We are interested in the conditions under which knowing the number of good types is valuable, since this is the purpose of signaling here. Knowledge is double-edged. It enables good players to satisfy their cooperation utility better, since they will have more accurate expectations of others' donations. And when there are many good types, this knowledge will also raise total contributions. On the other hand, when there are few good types, knowledge will lower all good types' contributions and exacerbate the collective action problem. Thus knowledge is certainly beneficial when cooperation utility is important enough to outweigh any possible decrease in contributions. As an illustration,[12] consider the additively separable case where good types' extra utility ψ is purely a function of the difference between own donations and some target function of others' contributions – representing perhaps guilt, reciprocity, or inequity aversion.

Example 1. *Suppose that ϕ is concave with $\phi' < 1$, and that $\psi(x_i, X_{-i}) = \hat{\psi}(x_i - T(X_{-i}))$ with $\hat{\psi}$ strictly concave and single peaked at 0. If β is high enough (holding other parameters fixed), and T is strictly increasing, knowledge increases good type welfare.*[13]

Knowledge will also be beneficial (for both types) if, by reducing uncertainty, it increases contributions.The next example shows this. We set $\phi(X) = \alpha X$, $\psi(x_i, X_{-i}) = X_{-i}x_i$, $0 < \alpha < 1$, with $x_i \in [0,1]$. Thus, players may make a contribution to a linear public good, but good types get an extra bonus if many others also contribute. When expectations of the number of good types are low enough, contributions will be 0 without knowledge; knowledge allows good types to contribute when there are enough of them.[14]

[12]A more general proof of the claim is provided in an Appendix.

[13]All proofs are in the Appendix.

[14]Due to the linearity there may be multiple equilibria; we assume that the equilibrium with most contribu-

Example 2. *If* $\phi(X) = \alpha X$*,* $\psi(x_i, X_{-i}) = X_{-i}x_i$*,* $0 < \alpha < 1$*, and* $x_i \in [0,1]$*, then knowledge improves good type welfare iff either* $\ddot{g} \leq \frac{1-\alpha}{\beta}$ *or* $\hat{g} \leq \frac{1-\alpha}{\alpha+\beta}$*, and increases bad type welfare iff* $\ddot{g} \leq \frac{1-\alpha}{\beta}$*, where* $\ddot{g} \equiv \sum_{g=0}^{N-1} \Pi^G(g)g$ *is the mean number of good types conditional on at least one good type, and* $\hat{g} \equiv \sum_{g=0}^{\left\lfloor \frac{1-\alpha}{\beta} \right\rfloor} \Pi^G(g)/Pr(g \leq \frac{1-\alpha}{\beta})$ *is the conditional mean given* $g < \frac{1-\alpha}{\beta}$ *and at least one good type.*

3.2 Signaling institutions

While complete information may increase welfare, individuals would not typically reveal their types via cheap talk, as explained in the introduction. An institution or mechanism to reveal types must therefore involve some costs. For instance, if an action was very costly for bad types but had no cost for good types, then this action could separate the types at zero cost in equilibrium. However, it is more reasonable to assume that types' costs are not completely independent, so that any mechanism must involve good types paying some cost. To provide a simple welfare metric, we assume that both types face the same cost for actions. Thus, in a signaling institution, players may choose to make a fixed payment $\sigma \geq 0$ before the basic game, and this payment is publicly observed.[15] We seek conditions for a separating equilibrium in which only good types pay the cost. This will hold if increasing the number of apparent good types increases others' contributions, and – more substantively – if good types value contributions more.

Lemma 1. *Suppose that* w^G *and* w^B *are bounded and strictly concave with, for all* x_i, X_{-i}*,* $w_1^B(x_i, X_{-i}) < 0$*,* $w_2^G(x_i, X_{-i}) \geq 0$ *and* $w_{12}^G(x_i, X_{-i}) \geq 0$*. Suppose also that* $||\Pi^G - \Pi^B|| < \varepsilon$ *for* ε *low enough.*[16] *Then a separating equilibrium exists if* $w_2^G(b(X_{-i}), X_{-i}) > w_2^B(0, X_{-i})$ *for all*

tions is always selected.

[15]We focus on conditions for a separating equilibrium. Under a full pooling equilibrium, signaling institutions are either irrelevant or impose a deadweight cost on all agents.

[16]I.e. treating Π^G, Π^B as points in the probability simplex. If $P^B(g)$ differs substantially from $P^G(g)$, the conclusion may not hold. For example, let $N = 3$. There is 1 good type with probability 1/2, and 0, 2 or 3

$X_{-i} \in [0, Nx_N]$.

Intuitively, when N is large, any individual's contribution will make little difference to the total and therefore have little effect on others' behavior, so the cost of signaling need not be high to deter bad types. This will hold, for example, when *welfare is a function of average contributions*: that is, $w^\tau(x_i, X_{-i}) = \hat{w}^\tau(x_i, X_{-i}/N)$, with the $\hat{w}^\tau$ constant as N varies.

Lemma 2. *If welfare is a differentiable function of average contributions with $w_{12}(x_i, X_{-i}) >$ 0, and $\Pi^G(g) \to \Pi^B(g) \to F(g/N)$ as $N \to \infty$ with F continuous, then for high enough N, if a signaling equilibrium is possible, the cheapest possible signal approaches* 0.

When the conditions for these Lemmas hold, and knowledge is welfare-improving, then a signaling institution can be beneficial, since it allows learning at a low cost. For example, if welfare is as in Lemma 2 then signaling can be beneficial: in an equilibrium with contributions, $w_2^G = \alpha + \beta > \alpha = w_2^B$, and $w_{12}^G = \beta > 0$.

The conditions need not hold in general. For instance if $\psi(x_i, X_{-i}) = \phi(X_{-i} + x_i)$ so that good types simply benefit more from the public good, and ϕ is concave, then higher others' contributions lead good types to reduce their contributions, i.e. $\psi_{12} < 0$ (this is the standard "crowding out" story, as in the framework of Warr (1983)). Similarly, if $\psi(x_i, X_{-i}) =$ $-(x_i - T(X_{-i}))^2$ with $T(\cdot)$ increasing, as in a model of inequality aversion, then good types benefit less from an increase in X_{-i} than bad types: in equilibrium, good type contributions are less than the target $T(X_{-i})$, and an increase in $T(X_{-i})$ increases their "guilt". On the other hand if $\psi(x_i, X_{-i}) = \alpha_G \phi(x_i + X_{-i}) - (x_i - T(X_{-i}))^2$ so that good types also receive a higher level of material benefit, the conditions of Lemma 1 may be fulfilled.

good types with probability 1/6 each. Suppose that if there are at least two good types, all good types wish to contribute $a > 0$; otherwise they wish to contribute nothing. Now, in a separating equilibrium a good type who pays σ will only be decisive when there is 1 other good type, which occurs with probability $P^G(1) = 1/5$ by Bayes' rule. A bad type will influence contributions with a 3/5 probability, $P^B(1) = 3/5$. Then if the relevant payoffs of the two types are not too different, the bad type will be prepared to pay more than the good type, and no separating equilibrium exists.

3.3 Exclusion and punishment

Suppose that after the signaling mechanism, but before the basic game, it is possible to exclude some or all players. Here crowding, indexed by δ in (2), becomes relevant. There are many possible exclusion mechanisms that capture our intuition. For simplicity we assume that the exclusion mechanism maximizes good types' welfare, given the information provided by the signaling institution.[17] Thus, if the institution separates the types, bad types will be excluded, since they would increase crowding but make no contributions in the basic game.[18]

We consider two kinds of signaling institution; the difference is only relevant if exclusion is possible, otherwise both lead to the results described in the previous section. In an *anonymous* signaling institution, only the number of agents who paid σ is visible. In a *public* signaling institution each agent's choice to pay σ or not is visible, and refusal to pay the signaling cost will result in exclusion from the collective good and a payoff of 0. Since sending a public signal means avoiding exclusion, the minimum signaling cost that meets the bad type's incentive constraint must be higher in the revealed case. While anonymous signaling institutions are less costly, as anonymity makes targeted exclusion impossible, public signaling can weed out bad types and reduce crowding; the optimal choice of institution depends on the tradeoff between these concerns.

Proposition 1. *Suppose that ϕ is strictly concave with $\phi' < 1$, and w^G is continuously differentiable with $w^G_2(x_i, X_{-i}, P) \geq 0$, $w^G_{12}(x_i, X_{-i}, P) \geq 0$. Then, for δ close enough to 0, if there*

[17] For example, suppose a player is chosen at random to exclude or include all players. All players will include themselves, but when N is large we can ignore this; otherwise, all players will include only good types. Alternatively, inclusion might be decided on a case-by-case basis by a majority vote. In this case, if there is a separating equilibrium, there will be unanimous agreement to include every good type, so long as their expected contributions outweigh crowding, and a $N-1$ against 1 vote to exclude every bad type, since even bad types wish to exclude other bad types.

[18] Some good types might also be excluded, if their contributions would not compensate for the increased crowding. This issue introduces cumbersome technicalities and is not central to our argument, so we simply assume that inclusion decisions cannot be probabilistic: either all those who pay the cost are included, or none are.

is a revealed signaling institution that separates the types, then there is also an anonymous signaling institution that separates the types and gives strictly higher good-type welfare.

The proof works taking $\delta \to 0$. However, even with full crowding ($\delta = 1$) anonymity may be helpful. We offer an example with a linear public good and full crowding, where good types value the public good more but also suffer a psychological loss from not hitting a target. Here we give only a simulation, providing some cases where anonymity "works."

Example 3. *Let $\phi(X) = \alpha_\tau X$ for $\tau \in \{B, G\}$, $1 > \alpha_G > \alpha_B > 0$, and let $\psi(x_i, X_{-i}) = -(x_i - [q\frac{X_{-i}}{N-1} + (1-q)])^2$ for $q \in [0,1]$, with $\beta = 1/2$. Let $\Pi^B(g) = \Pi^G(g) = 1/(N - (\underline{g} - 1))$ so the number of good types is uniformly distributed between $\underline{g}$ and N. Setting $N = 20$, $\underline{g} = 12$, $\alpha_B = .5$ and $q = .7$, Figure 1 shows good type welfare in the different institutions as α_G varies. Setting $N = 20$, $\alpha_B = .5$, $\alpha_G = .8$ and $q = .7$, Figure 2 shows good type welfare in the different institutions as $\underline{g}$ varies.*

Here, q measures how much the target depends on others' donations. If $q = 1$ donations will be 0 since good types who gave a positive amount, i.e., more than the average of all types, would strictly benefit from lowering their donations. $\underline{g}$ parametrizes the uncertainty about good types.

[Figure 1 about here.]

[Figure 2 about here.]

4 Conclusion

In societies without large-scale markets, and in areas that the market does not reach (e.g., inside the firm itself), others' character and intentions towards us may be vital for our success and survival. In these contexts it is crucial to be able to gauge others' character, so we can choose who to interact with and how much to invest in these interactions. This in turn gives some individuals a strong incentive to conceal their true character. Certain rituals can be seen as institutions which provide a forgiving environment, getting people to reveal their character by letting their identity be hidden. In this paper, we developed a model showing how anonymous rituals can foster greater cooperation than revealed ones.

Further empirical work is needed to establish whether this model explains the survival of *particular* rituals and institutions. Still, we note that cultural forms such as song seem especially well-suited for preserving the anonymity of participants and shirkers. We hope this approach will be of interest to anthropologists and sociologists of religion.

We also believe there are lessons for policy-makers and managers who wish to predict whether cooperation will be sufficient to undertake an ambitious project, or to reduce agents' uncertainty and thus foster voluntary cooperation. As mentioned at the end of section 2, anonymous contribution games and small "pre-loan" projects may be particularly relevant to joint micro-credit schemes. In designing these "trust-measuring" and "trust-building" exercises it may be important not to expose individual performance too much, since this can lead to uninformative pooling or pandering behavior. Incentive-compatible mechanisms which allow for collective achievement may work better. Indeed, a large non-academic literature on team-building emphasizes just this kind of group work (e.g. Newstrom and Scannell 1998).

References

Acemoglu, D., Dec. 2007. Incentives in markets, firms and governments. Journal of Law, Economics and Organization, 1–34.
URL `http://0-jleo.oxfordjournals.org.serlib0.essex.ac.uk/cgi/reprint/ewm055v1.pdf`

Akerlof, George A., Kranton, Rachel E., aug 2000. Economics and identity. The Quarterly Journal of Economics 115 (3), 715–753.

Andreoni, J., Petrie, R., 2004. Public goods experiments without confidentiality: a glimpse into fund-raising. Journal of Public Economics 88 (7-8), 1605–1623.

Austen-Smith, D., Banks, J., 2000. Cheap talk and burned money. Journal of Economic Theory 91 (1), 1–16.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games and Economic Behavior 10 (1), 122 – 142.

Bird, R. B., Smith, E. A., Apr. 2005. Signaling theory, strategic interaction, and symbolic capital. Current Anthropology 46 (2), 221–248.
URL http://www.journals.uchicago.edu/doi/abs/10.1086/427115

Bolton, G., Zwick, R., 1995. Anonymity versus punishment in ultimatum bargaining. Games and Economic Behavior 10 (1), 95–121.

Card, D., Olson, C., 1995. Bargaining power, strike durations, and wage outcomes: An analysis of strikes in the 1880s. Journal of Labor Economics, 32–61.

Cooter, R., Broughman, B., 2005. Charity, Publicity, and the Donation Registry. The Economists' Voice 2 (3), 4.

Coricelli, G., Fehr, D., Fellner, G., Jun. 2004. Partner selection in public goods experiments. Journal of Conflict Resolution 48 (3), 356–378.
URL http://jcr.sagepub.com/cgi/content/abstract/48/3/356

Fehr, E., Gachter, S., 2000. Cooperation and punishment in public goods experiments. The American Economic Review 90 (4), 980–994.

Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71 (3), 397–404.

Frank, R., 1987. If homo economicus could choose his own utility function, would he want one with a conscience? The American Economic Review, 593–604.

Fudenberg, D., Tirole, J., 1991. Game Theory. Mit Press.

Glazer, A., Konrad, K., 1996. A signaling explanation for charity. American Economic Review 86(4), 1019–1028.

Hagen, E. H., Bryant, G. A., 2003. Music and Dance as a Coalition Signaling System. Human Nature 14 (1), 21–51.

Harbaugh, W. T., May 1998. The prestige motive for making charitable transfers. The American Economic Review 88 (2), 277–282.

Hoge, D. R., Zech, C. E., Donahue, M. J., McNamara, P. H., 1996. Money matters: Personal giving in American churches. Westminster John Knox Pr.

Holmstrom, B., 1999. Managerial incentive problems: A dynamic perspective. The Review of Economic Studies 66 (1), 169–182.
URL `http://www.jstor.org/stable/2566954`

Iannaccone, L. R., Apr. 1992. Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. The Journal of Political Economy 100 (2), 271–291, ArticleType: primary_article / Full publication date: Apr., 1992 / Copyright Â© 1992 The University of Chicago Press.
URL `http://www.jstor.org/stable/2138608`

Karlan, D., 2005. Using Experimental Economics to Measure Social Capital and Predict Financial Decisions. American Economic Review 5: 25.

Karlan, D., List, J., 2007. Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment. American Economic Review 97 (5), 1774–1793.

Kellett, A., 1982. Combat motivation: The behavior of soldiers in battle. Kluwer-Nijhoff.

Keser, C., van Winden, F., 2000. Conditional Cooperation and Voluntary Contributions to Public Goods. Scandinavian Journal of Economics 102 (1), 23–39.

King, A., 2006. The word of command: communication and cohesion in the military. Armed Forces & Society 32 (4), 493.

Kreps, D. M., Milgrom, P., Roberts, J., Wilson, R., 2001. Rational cooperation in the finitely repeated prisoners' dilemma. Readings in Games and Information.

Levy, G., 2007a. Decision making in committees: Transparency, reputation, and voting rules. The American Economic Review 97 (1), 150–168.

Levy, G., 2007b. Decision-Making procedures for committees of careerist experts. American Economic Review 97 (2), 306–310.

List, J., Lucking-Reiley, D., 2002. The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. Journal of Political Economy 110 (1), 215–233.

List, J., Rondeau, D., 2003. The impact of challenge gifts on charitable giving: an experimental investigation. Economics Letters 79 (2), 153–159.

Londregan, J., Vindigni, A., 2006. Voting as a credible threat.
URL `http://www.princeton.edu/ pegrad/papers/londvind.pdf`

Martin, R., Fosh, P., Morris, H., Smith, P., Undy, R., 1991. The decollectivisation of trade unions? ballots and collective bargaining in the 1980s. Industrial Relations Journal 22 (3), 197–208.

McBride, M., Nov. 2007. Club mormon: Free-Riders, monitoring, and exclusion in the LDS church. Rationality and Society 19 (4), 395–424.
URL `http://rss.sagepub.com/cgi/content/abstract/19/4/395`

Milgrom, P., Roberts, J., Nov. 1990. Rationalizability, learning, and equilibrium in games with strategic complementarities. Econometrica 58 (6), 1255–1277, ArticleType: primary_article / Full publication date: Nov., 1990 / Copyright © 1990 The Econometric Society.
URL `http://www.jstor.org/stable/2938316`

Milinski, M., Semmann, D., Krambeck, H., 2002. Reputation helps solve the'tragedy of the commons'. Nature 415 (6870), 424–6.

Newstrom, J. W., Scannell, E. E., 1998. The Big Book of Team Building Games: Trust-Building Activities, Team Spirit Exercises, and Other Fun Things to Do. McGraw-Hill Professional.

Ostrom, E., 2000a. Collective action and the evolution of social norms. The Journal of Economic Perspectives, 137–158.

Ostrom, E., 2000b. Collective action and the evolution of social norms. The Journal of Economic Perspectives, 137–158.

Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: Self-Governance is possible. The American Political Science Review 86 (2), 404–417.

Prat, A., 2005. The wrong kind of transparency. The American Economic Review 95 (3), 862–877.

Randall, C., 2004. Interaction Ritual Chains.

Reinstein, D., Riener, G., 2009. Reputation and influence in charitable giving: An experiment.

Ruffle, B. J., Sosis, R. H., 2003. Does it Pay to Pray? Evaluating the Economic Return to Religious Ritual. SSRN, SSRN eLibrary.
URL `http://ssrn.com/paper=441285`

Schram, A., 2000. Sorting out the seeking: The economics of individual motivations. Public Choice 103 (3), 231–258.

Simpson, B., Willer, R., 2008. Altruism and Indirect Reciprocity: The Interaction of Person and Situation in Prosocial Behavior. Social Psychology Quarterly 71 (1), 37–52.

Sosis, R., Bressler, E. R., 2003. Cooperation and commune longevity: A test of the costly signaling theory of religion. Cross-Cultural Research 37 (2), 211.

Sosis, R., Ruffle, B. J., 2003. Religious ritual and cooperation: Testing for a relationship on israeli religious and secular kibbutzim. Current Anthropology 44 (5), 713–722.

Spence, M., 1973. Job Market Signaling. Quarterly Journal of Economics 87 (3), 355–374.

Stigler, G. J., 1972. Economic competition and political competition. Public Choice 13 (1), 91–106.
URL `http://dx.doi.org/10.1007/BF01718854`

Warr, P., 1983. The private provision of a public good is independent of the distribution of income. Economics Letters 13 (2), 207–211.

Weber, M., 1946. The Protestant Sects and the Spirit of Capitalism. Oxford, New York.

Appendix 1: Proofs

Proof of Example 1

Proof. Rewriting (4) as $\phi((g+1)x_g)/\beta + \hat{\psi}'(x_g - T(gx_g)) = 1/\beta$, as β increases x_g approaches the solution to $\hat{\psi}'(x_g - T(gx_g)) = 0$, in other words x_g approaches $T(gx_g)$. On the other hand, contributions without knowledge approach x^* satisfying $\sum_{g=0}^{N-1} \Pi^G(g)\hat{\psi}'(x^* - T(gx^*)) = 0$. Since $\hat{\psi}$ is single-peaked, x^* will give weakly lower cooperation utility than x_g for all g, and since T is single-valued and strictly increasing, x^* will give strictly lower cooperation utility than x_g for at least some $g \in \{1,...,N\}$ (which will occur with positive probability $\Pi^G(g) > 0$). For β high enough the resulting loss will dominate any change in the amount of the public good provided, since ϕ is bounded. $\square$

Proof of Example 2

Proof. $x_g = 1$ is possible in equilibrium if $\alpha - 1 + \beta g \geq 0 \Leftrightarrow g \geq \frac{1-\alpha}{\beta}$; otherwise $x_g = 0$. Ex ante good type welfare is $\sum_{g=\left\lceil \frac{1-\alpha}{\beta} \right\rceil}^{N-1} \Pi^G(g)\left\{\alpha(g+1) - 1 + \beta g\right\}$. When g is unknown, by the FOC $x^* = 1$ if $\ddot{g} \geq \frac{1-\alpha}{\beta}$, and 0 otherwise; the corresponding good type welfare is either $\sum_{g=0}^{N-1} \Pi^G(g)\left\{\alpha(g+1) - 1 + \beta g\right\}$ or 0. In the latter case, knowledge always improves good type welfare since for $g \geq \frac{1-\alpha}{\beta}$, $\alpha(g+1) - 1 + \beta g = (\alpha+\beta)g + \alpha - 1 \geq (\alpha+\beta)\frac{1-\alpha}{\beta} + \alpha - 1 = \frac{\alpha(1-\alpha)}{\beta} > 0$. (The increase in contributions also improves bad type welfare.) In the former case, the difference in favour of ignorance is $\sum_{g=0}^{\left\lfloor \frac{1-\alpha}{\beta} \right\rfloor} \Pi^G(g)\left\{\alpha(g+1) - 1 + \beta g\right\} = (\alpha+\beta)\sum_{g=0}^{\left\lfloor \frac{1-\alpha}{\beta} \right\rfloor} \Pi^G(g)g + Pr(g \leq \frac{1-\alpha}{\beta})(\alpha - 1)$. Rearranging gives that this is positive iff $\hat{g} > \frac{1-\alpha}{\alpha+\beta}$. On the other hand, bad type welfare is always lowered since contributions are decreased □

Proof of Lemma 1

Proof. Since $w_1^B < 0$ everywhere, bad types contribute nothing, so in an equilibrium with g good types, $w_1(x_g, gx_g) = 0$. (The Inada condition on ψ ensures the equilibrium is interior.) Define $V_\tau(X_{-i})$ as individual i of type τ's value from known contributions of X_{-i}, thus $V_G(X_{-i}) = w^G(b(X_{-i}), X_{-i})$, and since bad types contribute nothing, $V_B(X_{-i}) = w^B(0, X_{-i})$. Recall that x_g is good type equilibrium contributions when there are $g+1$ good types in total. The condition for good types to prefer to pay the costly signal is

$$\sum_{g=0}^{N-1} \Pi^G(g) V_G(gx_g) - \sigma \geq \sum_{g=0}^{N-1} \Pi^G(g) V_G(gx_{g-1}).$$

That is, when there are g other good types, by signaling one induces them to play the correct equilibrium for a total of $g+1$ good types; by not signaling one misleads them. Similarly, the condition for bad types to prefer not to pay the signal cost is

$$\sum_{g=0}^{N-1} \Pi^B(g) V_B(gx_{g-1}) \geq \sum_{g=0}^{N-1} \Pi^B(g) V_B(gx_g) - \sigma. \tag{5}$$

where the honest signal leads good types to believe correctly that there are only g good types.

For these both to hold for some $\sigma \geq 0$ it must be firstly that $\sum_{g=0}^{N-1} \Pi^G(g) V_G(gx_g) \geq \sum_{g=0}^{N-1} \Pi^G(g) V_G(gx_{g-1})$, and secondly that

$$\sum_{g=0}^{N-1} \Pi^G(g) \left\{ V_G(gx_g) - V_G(gx_{g-1}) \right\} \geq \sum_{g=0}^{N-1} \Pi^B(g) \left\{ V_B(gx_g) - V_B(gx_{g-1}) \right\}. \qquad (6)$$

The first condition will hold if V_G is increasing and total contributions gx_g are increasing in g. By the Envelope Theorem, $V_G'(X_{-i}) = w_2^G(b(X_{-i}), X_{-i})$ which is non-negative by assumption. That total contributions are increasing in g can be shown as follows. Suppose $h > g$ and $gx_g \leq hx_h$. Then $x_g < x_h$. And x_g solves $w_1(x_g, gx_g) = 0$. But $w_1(x_h, hx_h) < w_1(x_g, hx_h) \leq w_1(x_g, gx_g) = 0$, the first inequality by strict concavity of w (which follows from concavity of ϕ and ψ), and the second by $w_{12} \geq 0$. This contradicts optimality of x_h.

To show the second condition holds we prove

$$\sum_{g=0}^{N-1} \Pi^G(g) \left\{ V_G(gx_g) - V_G(gx_{g-1}) \right\} > \sum_{g=0}^{N-1} \Pi^G(g) \left\{ V_B(gx_g) - V_B(gx_{g-1}) \right\}.$$

WhenΠ^G is close enough to Π^B, since the terms in curlies are bounded, the above will suffice to show (6). For a proof, simply write $V_\tau(gx_g) - V_\tau(gx_{g-1}) = \int_{gx_{g-1}}^{gx_g} V_\tau'(y)dy = \int_{gx_{g-1}}^{gx_g} w_2^\tau(b^\tau(y), y)dy$, for all g and for $\tau \in \{G, B\}$, where $b^G = b$ and $b^B = 0$, and apply the assumption that $w_2^G > w_2^B$ along the equilibrium path. $\square$

Proof of Lemma 2

Proof. When welfare is a function of average contributions, x_g solves $\hat{w}_1(x_g, gx_g/N) = 0$, or writing $\gamma = g/N$ and $\xi_\gamma = x_{\gamma N}$, $\hat{w}_1(\xi_\gamma, \gamma\xi_\gamma) = 0$. Thus as N increases, equilibrium contributions when the number of good types g is a fixed proportion of N will be a constant ξ_γ and average contributions will be the constant $\gamma\xi_\gamma$.

The minimum signaling cost is given by σ satisfying (5) with equality, i.e. by

$$\sum_{g=0}^{N-1} \Pi^B(g) \left\{ V_B(gx_g) - V_B(gx_{g-1}) \right\}. \tag{7}$$

To show that this goes to 0 when N grows large, observe that it approaches

$$\int_0^1 \left\{ \hat{w}^B(0, \gamma\xi_\gamma) - \hat{w}^B(0, \gamma\xi_{\gamma-1/N}) \right\} dF(\gamma).$$

Now since $\xi_\gamma = x_{\gamma N}$ is increasing in γ by $w_{12} \geq 0$, it is continuous almost everywhere. Thus the term inside curlies goes to 0 almost everywhere as $N \to \infty$, and since F is continuous the whole expression goes to 0. □

Proof of Proposition 1

Proof. The analysis of anonymous signaling proceeds exactly as in Lemma 1, since nothing there depended on the identity of any participant being known. In particular, good type welfare from the cheapest possible anonymous signaling institution is

$$\sum_{g=0}^{N-1} \Pi^G(g) \left\{ \frac{\phi((g+1)x_g)}{1} - x_g + \beta \psi(x_g, gx_g) \right\} - \sum_{g=0}^{N-1} \Pi^B(g) \left\{ \phi(gx_g) - \phi(gx_{g-1}) \right\} \tag{8}$$

where the second term is the lowest possible signaling cost that separates the types, from (7).

In a separating equilibrium in a revealed institution, bad types who do not pay the cost will be excluded from the basic CAP. For, in the basic CAP with $g+1$ good types and P included players, good type donations solve

$$\frac{\phi'((g+1)x)}{P^\delta} + \beta \psi_1(x, gx) = 1$$

and the solution is decreasing in P (as an application of the Implicit Function Theorem shows). Thus, included bad types decrease welfare by increasing crowding, and by decreasing contributions (by assumption, $w_2^G(x_i, X_{-i}, P) \geq 0$ and in equilibrium $w_1^G(x_g, gx_g, P) = 0$).

After all bad types are excluded, and $g+1$ known good types remain, equilibrium good type contributions are given by $\tilde{x}_g$ solving

$$\frac{\phi'((g+1)\tilde{x}_g)}{[(g+1)/N]^{\delta}} + \beta \psi_1(\tilde{x}_g, g\tilde{x}_g) = 1 \tag{9}$$

(compare equation (4)).

If, in a separating equilibrium, a bad type were to pay the signaling cost and be included in the basic CAP, his expected welfare would be

$$\sum_{g=0}^{N-1} \Pi^B(g) \frac{\phi(g\tilde{x}_g)}{[(g+1)/N]^{\delta}}.$$

(With probability $\Pi^B(g)$ there are g good types; the misleading signal leads them to believe there are $g+1$ good types and so to play $\tilde{x}_g$; $g+1$ players are included so $P = (g+1)/N$.) This is therefore the lowest possible revealed signaling cost, since the bad type's alternative is to pay nothing, be excluded and receive a subsequent payoff of 0. Ex ante good type welfare is thus

$$\sum_{g=0}^{N-1} \Pi^G(g) \left\{ \frac{\phi((g+1)\tilde{x}_g)}{[(g+1)/N]^{\delta}} - \tilde{x}_g + \beta \psi(\tilde{x}_g, g\tilde{x}_g) \right\} - \sum_{g=0}^{N-1} \Pi^B(g) \frac{\phi(g\tilde{x}_g)}{[(g+1)/N]^{\delta}}. \tag{10}$$

Note also that this expression must be non-negative for a separating equilibrium to be possible, since the first term gives welfare from inclusion for good types, the second term welfare for bad types; if the second term is larger than good types would rather be excluded and get 0 than pay the cost and be included. If this does not hold, then revealed signaling fails to separate the types and gives weakly lower welfare than ignorance of g. Comparing (10) to (8) shows that it will suffice for our proof if $\tilde{x}_g \to x_g$ as $\delta \to 0$. For then the first term of each expresion will be the same and the only difference will be the larger signaling cost in the second term – since the bad type must be paid his total loss from exclusion, rather than the marginal loss from revealing his type and lowering x_g. Observe that $\tilde{x}_g = x_g$ when

$\delta = 0$, so it will suffice to show that $\tilde{x}_g$ is continuous in δ at $\delta = 0$. This can be done by the Implicit Function Theorem applied to the left hand side of (9). The requirements are that the LHS is continuously differentiable in $\tilde{x}_g$ (which holds by assumption), that its slope in $\tilde{x}_g$ is non-zero at $\delta = 0$ (which follows from strict concavity of ϕ and concavity of ψ), and that some $(\tilde{x}_g, \delta)$ satisfying (9) exist in an open interval around $(\tilde{x}_g = x_g, \delta = 0)$, which holds since x_g is interior. □

Appendix 2 - Proof that when "cooperation utility" matters enough, knowledge improves good type welfare

Suppose w is strictly concave and differentiable with $w_2 > 0$ and $w_{12} > 0$, and let $b(X) \equiv x$ solving $w_1(x,X) = 0$ be positive for all X. Set $w^{i\tau}(x,X) = w^{\tau}(b(X),X) + i[w^{\tau}(x,X) - w^{\tau}(b(X),X)]$ for $i > 0$ (and set $w^i \equiv w^{iG}$). Then for i high enough, good type welfare is higher when g is known than when g is unknown.

Proof. Set $w^i(x,X) = w(b(X),X) + i[w(x,X) - w(b(X),X)]$ for $i > 0$. It is easy to show that $w^i(b(X),X) = w(b(X),x)$, and that $w^i_1(x,X) = iw_1(x,X)$, for all x and X. Therefore, the sets of solutions x_g and x^* to $w^i_1(x_g, gx_g) = 0$ and $\sum_{g=0}^{N} \Pi^G(g) w^i_1(x^*, gx^*) = 0$, i.e. of equilibria with and without knowledge, do not vary with i, and expected good type welfare when g is known is also unchanged. Also $w^i_{12}(x,X) = w_{12}(b(X),X) + i[w_{12}(x,X) + w_{12}(b(X),X)] = iw_{12}(x,X) > 0$ for all i.

We wish to show that

$$\sum_{g=0}^{N-1} P^G(g)[w^i(x_g, gx_g) - w^i(x^*, gx^*)] > 0 \tag{11}$$

for high enough i. Fix i and define $b(\cdot)$ temporarily as the best response function for w^i. First observe that if $x_g \geq x^*$, then $w^i(x_g, gx_g) - w^i(x^*, gx^*) \geq 0$. For

$$
\begin{aligned}
& w^i(x_g, gx_g) - w^i(x^*, gx^*) \\
= \; & w^i(x_g, gx_g) - w^i(b(gx^*), gx^*) + w^i(b(gx^*), gx^*) - w^i(x^*, gx^*) \\
= \; & \int_{b(gx^*)}^{x_g} \left\{ \frac{d}{dx} w^i(x, b^{-1}(x)) \right\} dx + \left[w^i(b(gx^*), gx^*) - w^i(x^*, gx^*) \right].
\end{aligned}
$$

(b^{-1} is well defined since b is strictly increasing by $w^i_{12} > 0$.) Now the second term in brackets is non-negative by $b(\cdot)$ a best response (and positive whenever $x^* \neq x_g$. On the other hand, since $w^i_1(x, b^{-1}(x)) = 0$, $\frac{d}{dx} w^i(x, b^{-1}(x)) = w^i_2(x, b^{-1}(x))$ and this is positive, so the integral is non-negative. Thus the whole term is non-negative. (This whole step holds for any w^i including $w^1 \equiv w$.)

Next for $x_g < x^*$ we have

$$
\begin{aligned}
& w^i(x_g, gx_g) - w^i(x^*, gx^*) \\
& = w(x_g, gx_g) - w^i(x^*, gx^*) \\
& \text{(since } x_g = b(gx_g)) \\
& = w(x_g, gx_g) - w(b(gx^*), gx^*) + i \left[w(b(gx^*), gx^*) - w(x^*, gx^*) \right]
\end{aligned}
$$

and as $i \to \infty$ the sign of this depends on the sign of the term in square brackets, which is positive since $b(gx^*)$ is the unique best response to gx^*. Thus for i high enough the terms of (11) are always non-negative, and are positive whenever $x_g \neq x^*$.

□

Appendix 3: Experiment [19]

To explore the effect of anonymity in public goods games, we ran a laboratory experiment. Here we briefly describe the setup and report results (fuller details are in a companion pa-

[19]Note to reviewer/editor: This may be included as a section in the main text, put online, or removed entirely.

per [Author's other paper]). Groups of five subjects played two rounds of a linear public goods game. Three subjects took part in the first rounds; after this contribution levels were reported and the other two subjects could choose to exclude one of the first three. In the Revealed treatment, individual players (for example, the lowest contributor) could be targeted for exclusion. In the Anonymous treatment, contribution levels were reported without player labels, so individual players could not be targeted. Lastly, all non-excluded subjects took part in the second round public goods game. Thus, the first round was the anonymous or revealed signaling institution, while the second round represented the basic CAP. The whole game was repeated 15 times with stranger matching.

[Figure 3 about here]

In line with our theory, anonymity lowered contributions in the first round, but increased them in the second round, so that players earned more overall (Figure *3*). Second round contributions also declined much less over time in the anonymous treatment. Other results confirmed key predictions from our theory. In the anonymous treatment, first-round behavior was a better predictor of second-round behaviour (Figure *4*). As a result, subjects' expectations of second round contributions were more accurate. Many players were conditional contributors who gave more if they expected others to do so too; for this reason, second round contributions were correlated with other players' first round contributions in the anonymous treatment, but not in the revealed treatment.

[Figure 4 about here]

Tables and Figures

Figure 1: Anonymous and revealed welfare by α_G.

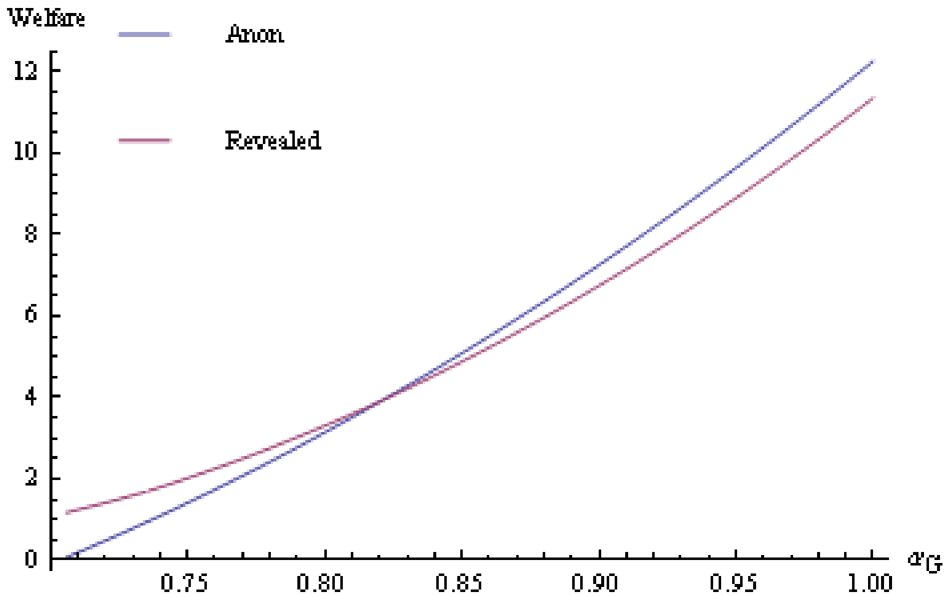

Figure 2: Anonymous and revealed welfare by $\underline{g}$.

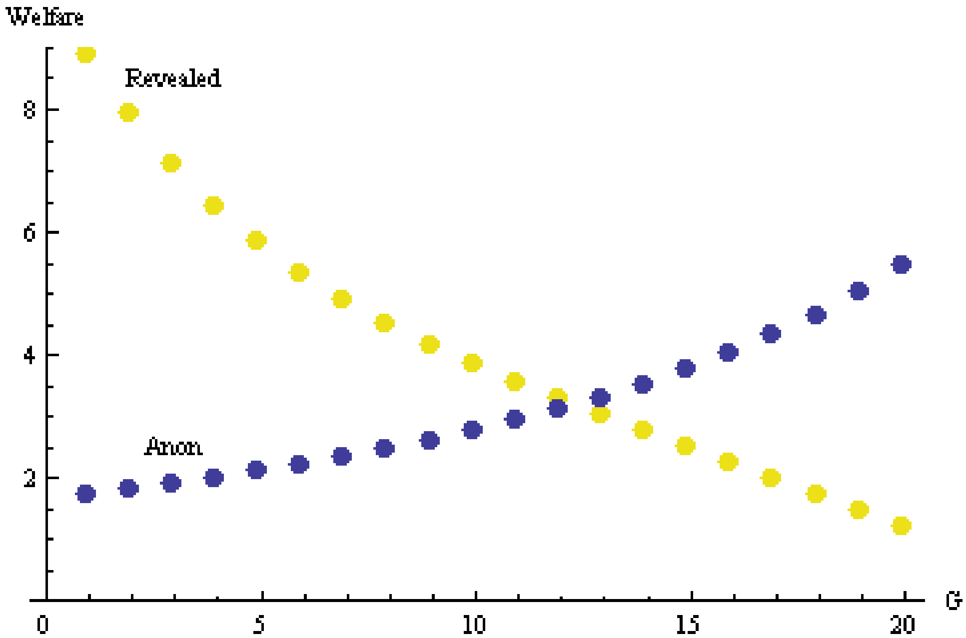

Figure 3: Mean Contributions by Repetition by Treatment

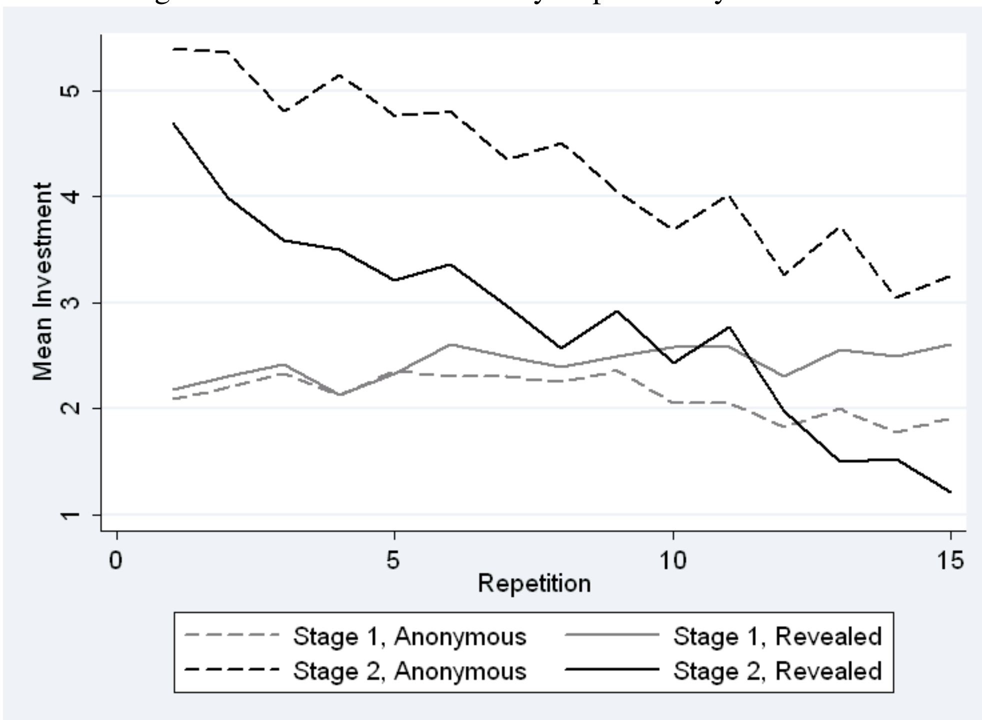

Figure 4: Density of Stage 2 investments by Stage 1 investment.