Public Good Contributions, Signalling, and the Evolution of Trust

Joel M. Guttman*

This paper develops a theory of the evolution of preferences for honesty, trust, and the voluntary provision of public goods in a society composed exclusively of rational, Bayesian optimisers. Unlike conventional evolutionary models, player types are not defined by their *strategies*, but rather (as in standard economic theory) by their *preferences.* Thus agents do not play fixed, 'wired-in' strategies, but rather choose strategies that maximise their expected utilities. In each stage of his or her 'career', an agent decides (a) whether to honour trust in a bilateral market transaction, and (b) whether to contribute to the provision of a non-excludable public good. We study the evolution of a community consisting of 'opportunists', who simply maximise material payoffs, and 'honest types', who prefer to honour trust and contribute to the provision of public goods. While individual interactions are one-shot, agents know the history of play of all others in the community. In equilibrium, opportunists contribute (with positive probability) to the provision of the public good in order to maintain reputations for being honest types, so as to be trusted in their market transactions over the course of their careers. A novel result of the model is that the presence of a public good enhances the evolutionary stability of the honest type.
JEL codes: D64, H41, Z13

*Department of Economics, Bar-Ilan University, 52900 Ramat-Gan Israel. E-mail: guttman@mail.biu.ac.il

> [M]any ostensible works of disinterested public spirit are no doubt initiated and carried on with a view primarily to the enhanced repute, or even to the pecuniary gain, of their promoters.
> Veblen (1899, p. 340)

1 Introduction

The voluntary provision of non-excludable public goods is one of the most prominent puzzles of modern economics. A long series of public good experiments indicate that real-world agents typically contribute to the provision of non-excludable public goods at levels exceeding those predicted by conventional economic models. Similarly, experimental work on the ultimatum game shows that individuals behave non-selfishly in situations where co-operation has mutual benefits.[1]

In recent years, theorists have tried to interpret these observations by introducing the assumption that some or all economic agents have preferences for reciprocity, leading them to contribute to the provision of public goods if they expect others to do the same.[2] Explaining behaviour by introducing preferences that are consistent with such behaviour, however, is a solution of the last resort. Economists would like to be able to *explain* the existence of such 'honest types' rather than simply introduce them by *assumption*. As Smith (2003, p. 467) stated in his Nobel Lecture, 'Technically, the issue is how most productively to model agent 'types' by extending game theory so that types are an integral part of its predictive content, rather than merely imported as an *ex post* technical explanation of experimental results.'

This paper proposes an evolutionary model that *predicts* the presence of honest types, in order to develop a more satisfactory explanation of why both honest and opportunistic individuals voluntarily provide themselves with public goods at levels far exceeding those predicted by the conventional, static Cournot-Nash contributions mechanism.

An alternative way of viewing the model is as a game-theoretic explanation of the emergence of self-enforcing norms of generosity and fairness. Society, by means of institutions like schools and the family, attempts to imbue children with the idea that individuals should help each other in times

[1] See Ledyard (1995) for a survey of these experimental results. See Güth and Tietz (1990) for a survey on the ultimatum game.

[2] See, e.g., Bolton and Ockenfels (2000).

of natural disasters, and, more generally, should not take a 'free ride' on the voluntary contributions of others in the provision of public goods. Similarly, we are taught that it is not fair to take advantage of others in bilateral transactions. It is not considered fair, for example, for a seller to receive payment for a good and then not supply the good.[3] The question remains, however, whether these norms have an effect on everyday social interactions, or are mere moral exhortations with little practical relevance.

In the theory developed here, one type of agent receives utility from abiding by these (closely related) norms of fairness. It seems reasonable to assume that some children more readily 'absorb' these norms than others. Given that the same institutions try to imbue agents with norms both in the area of voluntary collective action and in the area of honesty in private transactions, we would expect that agents who abide by the norm of contributing to the provision of public goods are relatively likely to be individuals who are honest in their private transactions. And indeed, there is evidence that volunteering in the provision of public goods is widely regarded as an indication that the volunteer is relatively likely to be helpful and honest in private interactions.[4] The question that immediately arises is, will this 'honest' type survive in a competitive environment, in which opportunists, who obtain no utility from abiding by such social norms, are present as well?

The starting point of the theory is the pioneering contribution of Kreps et al. (1982). Kreps et al. proved that if one of two rational players assigns a small probability to the proposition that his or her opponent is an irrational 'tit-for-tat' (TFT) type, or alternatively, if both players assign a small probability to the proposition that their opponent has a taste for reciprocity (preferring joint co-operation to exploiting his or her opponent), then, in a finitely repeated Prisoner's Dilemma (PD) game, co-operation is an equilibrium outcome for at least some of the stages of the game. The assumed 'seed'

[3]For models in which parents, or society as a whole, invest resources to influence the tastes of the young, see Akerlof (1983), Becker (1991, 1993), Guttman, et al. (1992), Stark (1995), Bisin and Verdier (1998), and Guttman (2001a, 2001b).

[4]Katz and Rosenberg (2005) cite manuals of the 'how to get a good job' genre as providing such evidence. For example, Schaefer (2000) states, 'Volunteering experience is very important to anyone looking for a job... It shows that you are someone who cares about your community and that you are willing to spend your own time to help others.' Orndorff (2000) similarly writes, 'Recruiters like to see you get involved in your community.' Katz and Rosenberg (2005) state that 'A quick survey of the career and resume books in a local bookstore revealed that almost all mention volunteering as a positive strategy.'

of uncertainty leads the player (incorrectly) suspected of being an irrational TFT type or being a reciprocator type to justify this doubt, by cultivating a reputation for being the suspected type. This carefully preserved reputation induces the opponent to cooperate, at least in the initial stages of the game.

Kreps et al. simply introduced the players' prior beliefs (i.e., the uncertainty of one or both players regarding the opponent's type) by assumption. This assumption seems arbitrary, in that it requires that one or both players have slightly mistaken priors about their opponents. Why would two rational and selfish agents suspect each other of being something other than rational and selfish?[5] A simple answer to this question would be that the two players are drawn from a larger population in which players of the suspected reciprocator type exist. Rational players, who know the population proportions of the various types, use these proportions to develop their prior beliefs regarding their opponents' type. But the question then arises, why would we expect reciprocator types to exist if their lifetime payoffs are lower than those of opportunists? A popular saying has it that 'nice guys finish last.' This paper shows that this saying is not always correct—particularly where public good contributions can serve as signals to separate the 'nice guys' from those who are not so nice.

The present paper develops an 'indirect' evolutionary model which provides the conditions under which *part* of a population of fully rational individuals will, in evolutionary equilibrium, consist of 'reciprocators' in the sense of Kreps et al. (1982) in their 'Model 2' (which assumes two-sided uncertainty regarding the opponent's preferences). I develop a theory in which honest types survive in a competitive, evolutionary environment where such players must compete with 'opportunistic types' who contribute to the provision of the public good only when it pays in material terms. The presence of honest types in the population induces opportunists, as well, to contribute, in order to preserve reputations for being honest types. These reputations are required in order to have trading partners in bilateral market transactions.

These bilateral transactions are introduced in order to model agents' private incentive to contribute to the provision of public goods. As Olson [(1983), p. 15] observed, public goods are usually provided in a larger social context in which agents repeatedly transact with each other in the supply of

[5]Note that rationality and selfishness are not the same. In Model 1 of Kreps, et al. (1982), one agent's uncertainty regards his or her counterpart's *rationality,* while in Model 2, the uncertainty relates to the other agent's *selfishness.*

private goods. By free-riding, agents may lose reputations for trustworthiness that are valuable in their private-good transactions.[6]

The novelty of the present model is not the idea that opportunistic agents contribute to the provision of public goods in order to maintain reputations for trustworthiness. As the opening quotation from Veblen (1899) illustrates, this is a very old idea. The novelty of the model is the demonstration that a nucleus of *truly* honest types, whose presence is required in order for the reputational mechanism to operate, is evolutionarily stable specifically when opportunities exist for the voluntary provision of public goods.

Standard evolutionary models are generally regarded as differing fundamentally from conventional economic theory, in that individuals are not usually assumed to be rational. Instead, individuals are assumed to play pre-programmed strategies such as simply cooperating or defecting all the time, or (more realistically) rewarding co-operation and punishing defection by some strategy such as TFT. The early results of Axelrod (1984), which popularised the idea that TFT is a winning strategy in an evolutionary environment, thus explaining co-operation, have been shown to be rather fragile. For example, a pure-defect 'mutant' can easily invade an all-TFT population if the PD interactions are finitely repeated. If, on the other hand, PD interactions are infinitely repeated, then there is no pure strategy that is evolutionarily stable [Boyd and Lorberbaum (1987)].

This paper proposes to use a somewhat less conventional type of evolutionary model in order to endogenise the prior beliefs of rational players regarding their opponents' type. The model employs what has been called the *indirect* evolutionary approach, in which all players are assumed to be rational, and the evolutionary mechanism determines the population mixture of players with differing preferences.[7] In other words, agent types are defined not by the agents' *strategies* but by their *preferences.* Agents choose strategies to maximise expected payoff. However, expected payoff, for non-opportunistic types, may not be expected *material* payoff. Thus, unlike conventional evolutionary models in the tradition of Axelrod (1984), the model does not depart from the traditional assumption of maximising behaviour, but only introduces player types with 'non-standard' preferences, and studies their ability to survive when 'standard' or opportunistic types are present as

[6] A related strand of literature studies gift-giving as signalling [Camerer (1988); see also Kranton (1996), and Carmichael and MacLeod (1997)].

[7] Early contributions to this literature include Frank (1987, 1988), Güth and Yaari (1992), and Güth (1995).

well. As in conventional evolutionary models, an agent type's survivability is determined by the agent's expected material payoff or 'fitness.'

Readers who are used to regarding the evolutionary and the conventional economic paradigms as diametrically opposed may wonder what is gained by this marriage of the two approaches. The answer is simply that by combining these approaches, we can explain why players *rationally* expect their opponents to be cooperative or honest, rather than simply introducing these expectations by assumption, as in standard reputational models. Thus we obtain an empirically testable theory of rational co-operation, rather than one that can predict any outcome by the introduction of appropriate prior beliefs. The evolutionary model, moreover, becomes more intellectually satisfying to economists, since we do not jettison economists' traditional (and often successful) assumption of rationality.

A basic result of the literature [e.g., Güth (1995)] is that unless agents observe some signal of their opponents' type, preferences that do not reflect the 'objective' or materialistic payoffs of the agent—for example, the honest type's preferences of the present model—eventually will be driven to extinction. Honest types can survive only if agents exhibit a signal correlated, however weakly, with their type.[8]

The question then arises [see Samuelson (2001)]: why cannot opportunistic types also signal that they are 'prosocial' types, thus making the signal uninformative? The present model provides an answer to this question. If a public good is provided voluntarily in a society that also plays a repeated market trust game, contributions to the provision of the public good can signal that the contributor is an honest type. When the proportion of honest types in the population is sufficiently small, opportunists will signal (contribute to the provision of the public good) in equilibrium less frequently than honest types contribute, despite the resulting damage to their reputations.

Thus the coupling of the market trust game with the public good contribution game generates a 'complementarity' between the two games. On the one hand, the public good game provides the signalling mechanism that enables the honest types to survive, despite their material disadvantage in the repeated market trust game. On the other hand, as noted earlier, the opportunistic agents' incentive to maintain their reputations in the repeated market trust game is what makes it optimal for them to contribute to the

[8] Models that incorporate such signals include Frank (1987), Güth (1995), and Guttman (2003). Frank (1988) presents evidence that such signals are, in fact, emitted by humans.

provision of the public good.

The remainder of this paper is organised as follows. Section 2 introduces the two player types, 'opportunist' and 'honest,' and sets out the other assumptions of the model. Section 3 shows how, for a given prior belief regarding the opponent's type, one obtains a Perfect Bayesian Equilibrium (PBE) in which rational players may voluntarily provide themselves with a non-excludable public good. Section 4 endogenises these prior beliefs in an indirect evolutionary framework, by comparing the equilibrium expected payoffs of the honest types and opportunists. It is shown that the existence of a public good that needs to be supplied voluntarily by the community, enhances the evolutionary stability of the honest type. Section 5 discusses evidence from a variety of sources that can be used to assess the validity of the model. Section 6 offers concluding observations.

2 Assumptions

2.1 Bilateral Trust Games

In each period of their 'careers,' agents are randomly matched to play two-stage 'trust games,' which can be viewed as market transactions, in which the costs of using the legal system to enforce the terms of the contract are prohibitive.[9] In this game, the first mover (the buyer) decides whether to buy a good from the second mover (the seller). The decision of the buyer to buy will be called 'trusting' the seller. If the buyer decides not to trust (abstains from buying), the game ends, and the payoffs of both players are zero. If the buyer trusts, then the seller can honour this trust by producing a good of an agreed-upon level of quality, entailing an expenditure of effort which carries a cost of $c \in (0, 1)$.

If the seller honours the trust given him or her, both players benefit from this trade. The buyer's net payoff (consumer's surplus) is unity, and the seller's net payoff is $1 - c > 0$. Alternatively, the seller can fail to honour the trust by expending a smaller amount of effort. For simplicity of exposition, the cost to the seller of this smaller amount of effort is zero (only the high effort level is irksome), but the product produced will then be defective with

[9] For empirical studies demonstrating the importance of trust and reputation, as opposed to legal enforcement, in real-world market transactions, see Macauley (1963), Ellickson (1991), and Bernstein (1992).

probability 1. If a defective good is produced, the buyer's net payoff is $-a < 0$. In this case, the seller still receives the payment transferred by the buyer, which he or she values at 1.

Both players have an 'outside option' which gives a payoff of zero. The buyer utilises this outside option when he or she decides not to trust. The seller exercises his or her outside option by not offering a good for sale.

There are two types of players in the model: 'opportunistic' and 'honest'. The proportion of honest types in the community is $p \in [0, 1]$. Opportunistic types simply maximise their material payoffs, while honest types have preferences that induce them to behave otherwise. In particular, an honest seller receives a payoff of $-\varepsilon < 0$ if he or she fails to honour trust, and exerts the low level of effort. Thus an honest seller, if trusted, will always honour trust.

In contrast, an opportunistic seller's optimal move in a one-shot game of this type, if trusted, is to exert the lower level of effort, thus receiving a payoff of 1 rather than $1 - c$. It is therefore optimal for the buyer not to trust a seller known to be an opportunist. Thus, in the subgame perfect equilibrium of a one-shot game with complete information, no trust is granted, and the payoffs of both players are zero. Figure 1 shows the one-shot version of the market transaction game (only material payoffs are shown).

When applying the theory of repeated games, the analyst must decide whether to model the game as finitely or infinitely repeated. The present model assumes that the game has a finite, commonly known endpoint T. Clearly, the assumption of a commonly known final stage is an abstraction from the real world, in which the maximum length of agents' (finite) 'careers' is uncertain. Nevertheless, after some stage, the probability that the agent will not survive another stage rises quite sharply. Thus it seems reasonable to assume that opportunistic agents, at some stage in their careers, will not have sufficient reputational incentives to honour their commitments. Moreover, when this stage is reached, it will be fairly well known to the other members of a small, tightly knit community. This stage, for the purposes of the present analysis, will be considered the end of the agent's career.[10]

For simplicity of exposition, I assume that time is continuous, and is indexed by $t \in [0, T]$, where T is the commonly known end of the agent's

[10]The assumption of a finitely repeated game with a commonly known endpoint greatly simplifies the analysis, but the results would be very similar if we were to analyze an infinitely repeated game, with the proviso that a third type, who always defects, is also introduced into the model. See Guttman and Surana (2004) for such a model.

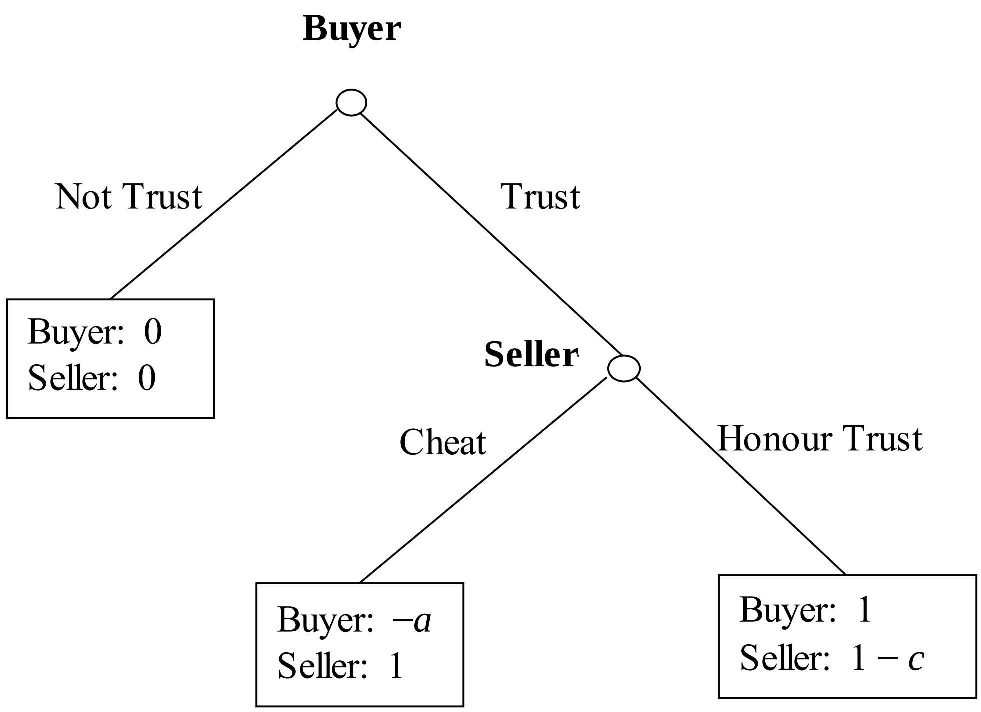

Figure 1: Market Trust Game

career. Agents have a discount rate $r > 0$, which may reflect time preference, or uncertainty that the agent will remain in the community into the next time period.[11]

Each player's type is his or her private information. Agents know only the proportion of honest types, p.

At each time t, each player is randomly matched with two other agents. With one of these agents, the player plays the role of a buyer in a trust game as outlined above, and with the other agent, he or she plays the role of a seller. In the next time period, the player is randomly rematched. It is assumed that the number of agents in the community is large enough that the probability of being matched twice with the same seller is negligible.

Information of a player's history of play spreads, but not instantaneously, throughout the community. If an agent sells a low-quality product, this becomes known, after a time interval $\delta > 0$,[12] to all members of the community, for the rest of the agent's career in the community.

In a given *generation* of players, the proportion of honest types p is fixed. Agents of a given generation have children, and the proportions of the types change in the younger generation as a function of the relative *undiscounted expected material payoffs* of the various types. I leave unspecified the exact mechanism by which these types change, but the underlying idea is that the higher the agent's material payoff, the more children he or she is likely to raise, or the more likely that the children will adopt the preferences of the parent,[13] and these children will be raised with equal probability over the agent's career.

[11] If information on the agent's history of play does not flow from one community to the next, agents who cheat can make a 'fresh start' in building their reputations by moving to another community. To simplify the model, we assume that the cost of moving is high enough to make such opportunistic moves suboptimal. However, there may be exogenous, unforeseen opportunities that outweigh these costs and therefore exogenously 'transfer' agents from one community to the next. The agent's discount rate may therefore reflect his or her uncertainty that such a move may take place in the future.

[12] It is assumed, for simplicity, that the entire community learns simultaneously that the agent cheated, after the time interval δ. See Guttman and Yacouel (2004) for a model in which information of cheating spreads gradually in the market.

[13] See Boyd and Richerson (1985) who develop the idea of 'cultural evolution,' which assumes that children's preferences are adapted to those of the more materially successful agents of the previous generation. Guttman (2001b) provides a model in which parents invest in influencing the preferences of their children so as to maximise the children's lifetime income.

2.2 Voluntary Provision of a Public Good

In addition to playing this trust game, agents decide whether to contribute to the provision of a public good, the benefits of which are enjoyed by all members of the community. Contributions are discrete: either the agent contributes or he or she does not ('free rides'). An important assumption of the model is that public good contributions are made prior to agents' decisions of whether to honour trust or cheat in their private-good transactions. To simplify the model, I assume, in fact, that each agent either contributes or free-rides only once, just before deciding whether to honour trust or cheat at $t = 0$. For simplicity, it is assumed that the information of whether the agent contributes or free-rides immediately becomes known to all members of the community.[14]

I assume that the quantity provided of the public good is a linear function of the number of agents contributing. Each agent's cost of contributing, net of his or her private benefit from the increased quantity of the public good, is denoted $v > 0$.

For the purposes of the model, it is not necessary that the expenditure v actually be a contribution to the provision of a public good. Any transfer of resources (time or money) to help another individual, that is highly visible to the other members of the community and is not likely to call forth an equally valuable, reciprocal act of kindness from the same individual to the current contributor, would serve the same purpose. The 'common denominator' of all such actions is that they are positively correlated with the same ethical qualities that lead agents to keep their word in bilateral transactions.

Since opportunists are assumed to maximise their material payoffs, they will not contribute to the provision of the public good, unless such contributions yield private benefits in their private-good transactions. Honest types, on the other hand, are assumed to receive utility from contributing, or a 'warm glow' as it is often called, which has the value $b > v$.[15] To repeat

[14]Since contributing to the provision of a public good is often a highly visible action, it is not implausible to assume that information of an agent's contributing or free-riding spreads quickly in the community. If, however, we were to introduce a time lag as we do in the case of private good cheating, our results would not be very different.

[15]Thus their preferences imply that honest agents will always contribute to the provision of the public good, just as they always honour trust (if trusted) in their market transactions. *Prima facie*, it may appear that little is therefore gained by assuming that honest types are rational. This appearance, however, is misleading. Recall that as buyers, honest types behave in precisely the same, sophisticated manner as opportunists. The

the argument stated in Section 1, it is true that not all agents who receive utility from honest dealing in their private market transactions also receive utility from contributing to the provision of public goods. But it is widely presumed (as is documented by Katz and Rosenberg, 2005) that there is a positive correlation between the utilities that agents receive from these two actions. I only assume that this common presumption is not totally incorrect. In a more elaborate model, we could let b be a random value, with a minimum value of zero, to allow for heterogeneity between honest types with respect to their values for b. To keep the present model as simple as possible, I assume a common b for all honest agents.

3 Analysis of the Game: Exogenous Population Mixture

In this section, the proportion of honest types p is treated as an exogenous parameter. In Section 4, this variable will be endogenised by embedding the model in an indirect evolutionary process. I begin by analysing the game without public good contributions. In Section 3.2, these contributions will be introduced into the game.

Throughout, the solution concept will be Perfect Bayesian Equilibrium (PBE).

3.1 The Game without Public Goods

Consider an opportunistic agent who is deciding whether to cheat (not honour trust given him or her, by exerting the low level of effort) for the first time in his or her career. We first observe:

Proposition 1. *If a seller ever sells a defective product at time t, he or she will not be trusted, in equilibrium, from time $t + \delta$ to the end of his or her career in the community.*

Proof. If an agent sells a defective product, this implies that he or she is an opportunist. It then becomes common knowledge that the agent will cheat at the end of his or her career, from time $T - \delta$ to time T. The agent will therefore not be trusted from time $T - \delta$. But this implies that the agent

expected-payoff maximising decisions of all buyers, honest and opportunistic, are crucial to the operation of the model.

has no incentive to honour trust from time $T-2\delta$, and therefore will not be trusted. Thus, by backwards induction, the agent will not be trusted after it is known that he or she sold a defective product, i.e., from time $t+\delta$. ¥

Proposition 1 implies that, in PBE, an opportunist will either cheat (exert the low level of effort) at $t=0$ and then not be trusted from time δ to the end of his or her career, or will honour trust up to some point in time, then cheat, and not be trusted (with a lag of δ) afterward. Let $\bar{t} \in [0,T]$ be the time that the agent starts cheating. Then the opportunist's discounted payoff as a seller, if trusted, will be

$$\pi_o = (1-c)\int_0^{\bar{t}} e^{-rt}dt + \int_{\bar{t}}^{\bar{t}+\delta} e^{-rt}dt. \tag{1}$$

The first term of this expression is the agent's discounted payoff from honouring trust in market transactions up to stage $\bar{t}$. The second term is the agent's discounted payoff from cheating over the time interval $[\bar{t}, \bar{t}+\delta]$, when his or her instantaneous payoff is 1. Differentiating π_o w.r.t. $\bar{t}$, we find that π_o increases as $\bar{t}$ increases, if r is sufficiently small. We thus have[16]

Proposition 2. *The opportunistic agent, if trusted, will optimally cheat only from time $T-\delta$ onward if $r < -\ln(c)/\delta$. Otherwise, the opportunistic agent will optimally cheat at $t=0$.*

Proof. See Appendix. ¥

Proposition 2 implies that the opportunist's maximal discounted payoff is

$$\overline{\pi}_o = \begin{cases} \dfrac{1-e^{-r\delta}}{r} & \text{if} \quad r \geq -\dfrac{\ln(c)}{\delta} \\ \left(\dfrac{1-c}{r}\right)[1-e^{-r(T-\delta)}] + \dfrac{1}{r}[e^{-r(T-\delta)} - e^{-rT}] & \text{if} \quad r < -\dfrac{\ln(c)}{\delta} \end{cases}$$

We have been assuming that the seller is trusted. To determine the conditions under which this assumption will hold true, note first that only honest sellers will honour trust at time T. Thus the buyer's expected payoff of trusting at time T is

$$E\pi_b = 1 \cdot p - a(1-p)$$

[16]We assume, in the case of indifference between cheating and honoring trust, that the opportunist cheats. Recall that $c \in (0,1)$. Therefore $-\ln(c)/\delta > 0$.

which will be positive if and only if

$$p > \frac{a}{1+a}. \qquad (2)$$

Denote the r.h.s. of (2) as p_{min}. Thus we find that *sellers will be trusted at time T if and only if $p > p_{\mathsf{min}}$*. For the same reason, sellers will be trusted at all times $t < T$ if $p > p_{\mathsf{min}}$. If this condition holds, and if in addition $r < -\ln(c)/\delta$, then (by Proposition 2) opportunistic sellers will optimally honour trust at all times $t < T - \delta$. Thus we conclude that if $r < -\ln(c)/\delta$ and $p > p_{\mathsf{min}}$, buyers will trust sellers, in PBE, throughout the sellers' careers, and opportunistic sellers will honour trust up to, but not including, time $T - \delta$.

Now suppose that $r < (1/c) - 1$ but $p \leq p_{\mathsf{min}}$.[17] Proposition 2 states that even opportunistic sellers (and not only honest sellers) will honour trust up to, but not including, stage $T - \delta$, if sellers are trusted, provided that $r < -\ln(c)/\delta$. Assume momentarily that sellers were therefore trusted up to, but not including, time $T - \delta$. An opportunistic seller's best response to this strategy of the buyers would be to cheat at time $T - 2\delta$, since he or she will not be trusted at $T - \delta$, regardless of whether he or she honours trust or cheats just before T. But then a buyer's best response would be to trust at all times up to (but not including) time $T - 2\delta$. Given this, however, sellers will cheat at $T - 3\delta$, and so forth. Thus we conclude that there is *no PBE in which sellers are trusted at all times prior to, but not including, time $T - \delta$.* Using a similar backwards induction argument, we find that (in PBE) either sellers are trusted throughout their careers, including time T, or they will not be trusted at all.

We thus arrive at

Proposition 3. *In equilibrium, sellers will be trusted with probability 1 throughout their careers if and only if $p > p_{\mathsf{min}}$.*

Combining Propositions 1, 2 and 3, we obtain

Proposition 4. *The unique pure strategy PBE of the repeated market trust game (without a public good) is:*

- *If $p \leq p_{\mathsf{min}}$, sellers will not be trusted. Therefore the payoffs of all agents will be zero.*

[17] For expositional simplicity, we assume that in the borderline case $p = p_{\mathsf{min}}$, sellers are not trusted.

- *If $p > p_{\mathsf{min}}$, sellers will be trusted throughout their careers. Honest sellers will honour trust throughout their careers, while opportunistic sellers will honour trust up to, but not including, time $T - \delta$, if $r < -\ln(c)/\delta$. If $r \geq -\ln(c)/\delta$, opportunistic sellers will cheat at $t = 0$.*

3.2 The Game with a Public Good

Let us now introduce a public good, which is produced under the conditions specified in Section 2.2. Recall that the cost of contributing, net of the agent's private return from the added provision of the good generated by his or her contribution, is v. Opportunists do not contribute, unless contributing gives sufficient returns in their market transactions. Honest types have a greater incentive to contribute, since they receive utility from contributing, $b > v$. Therefore, honest types will always contribute.

To analyse the game with a public good, first note that if $\bar{\pi}_o \leq v$, opportunists do not have a sufficient incenive to contribute, since even if contributing is necessary to conduct market transactions, the cost of contributing is greater than the benefit. Therefore, if $\bar{\pi}_o \leq v$, only honest types will contribute.

Now suppose that $\bar{\pi}_o > v$. If this inequality holds, there are two cases. In Case 1, $p > p_{\mathsf{min}}$, so that an agent whose type is unknown will be trusted (Proposition 3). If an agent does *not* contribute, however, he or she will be assumed to be an opportunist, and therefore will not be trusted, by the backwards induction argument used in the proof of Proposition 1. Contributing, then, is a prerequisite to conducting market transactions. Therefore an opportunist will contribute if and only if his or her payoff in the market trust game, $\bar{\pi}_o$, is greater than the cost of contributing, v.

In Case 2, $p \leq p_{\mathsf{min}}$. In this case, if $\bar{\pi}_o > v$, there is no equilibrium in pure strategies. To see this, suppose that opportunists always contribute to the provision of the public good. Then the act of contributing will not signal that the agent is an honest type; no information is conveyed by the fact that the agent contributes. Given that $p \leq p_{\mathsf{min}}$, no agents will be trusted, even if they contribute. But if agents are not trusted even if they contribute, opportunists have no incentive to contribute, so this cannot be an equilibrium. Now suppose that opportunists always free-ride. Then contributing will indeed serve as a signal that the agent is honest, so that contributors will be trusted, while free-riders will not be trusted. But this implies that opportunists will optimally contribute, given that $\bar{\pi}_o > v$. So this alternative situation, in which opportunists always free-ride, is also not an equilibrium.

Therefore the equilibrium, in Case 2 with $\bar{\pi}_o > v$, must be characterised by mixed strategies. In the mixed strategy equilibrium, the expected discounted payoffs of the two pure strategies (contribute and free-ride) must be equal. Since free-riders reveal that they are opportunists, their lifetime payoffs will be zero. Therefore, in equilibrium, the discounted lifetime expected payoffs of the opportunists who contribute must also be zero. We thus have, in equilibrium,

$$\Pr(\textit{trust}\,|\textit{contribute})\ \bar{\pi}_o - v = 0,$$

where $\Pr(\textit{trust}\,|\textit{contribute})$ is the probability that an agent will be trusted, if he or she contributes. Therefore

$$\Pr(\textit{trust}\,|\textit{contribute}) = \frac{v}{\bar{\pi}_o}.$$

In the mixed strategy equilibrium, buyers must now be indifferent between trusting and not trusting, in order for it to be optimal to mix between these two pure strategies. Let $\Pr(\textit{honest}\,|\textit{contribute})$ denote the probability that the seller is honest, if he or she contributes to the provision of the public good. We have

$$E\pi_b = 1 \cdot \Pr(\textit{honest}\,|\textit{contribute}) - a[1 - \Pr(\textit{honest}\,|\textit{contribute})] = 0, \tag{3}$$

since the buyer's payoff of not trusting is zero. This equation must hold in order for buyers to be indifferent between trusting and not trusting at time T of the seller's career, when an opportunistic seller will cheat if $r < (1/c) - 1$ (by Proposition 2), or at $t = 0$, when an opportunistic seller will cheat if $r \geq (1/c) - 1$. At this stage, only if an agent is believed to be honest with probability at least p_{min}, will it be optimal to trust. As we found above, if $r < (1/c) - 1$, sellers must be trusted at time T in order for there to be trust at any stage of their careers. Rearranging (3), we have, in equilibrium,

$$\Pr(\textit{honest}\,|\textit{contribute}) = \frac{a}{1+a}. \tag{4}$$

Let $\Pr(\textit{contribute}\,|\textit{honest})$ be the probability that an honest agent will contribute, which is simply unity, and $\Pr(\textit{contribute}\,|\textit{opportunist})$ be the probability that an opportunist will contribute. Then the probability that a randomly drawn agent will contribute is

$$\begin{aligned}\Pr(\textit{contribute}) &= p\Pr(\textit{contribute}\,|\textit{honest}) \\ &\quad +(1-p)\Pr(\textit{contribute}\,|\textit{opportunist}) \\ &= p+(1-p)\Pr(\textit{contribute}\,|\textit{opportunist})\ .\end{aligned}$$

By Bayes' Theorem,

$$\begin{aligned}\Pr(\textit{honest}\,|\textit{contribute}) &= \frac{p\Pr(\textit{contribute}\,|\textit{honest})}{\Pr(\textit{contribute})} \\ &= \frac{p}{p+(1-p)\Pr(\textit{contribute}\,|\textit{opportunist})} \qquad (5)\end{aligned}$$

Combining (4) and (5), we have, in equilibrium,[18]

$$\Pr(\textit{contribute}\,|\textit{opportunist}) = \frac{p}{a(1-p)}. \qquad (6)$$

We thus obtain

Proposition 5. *In equilibrium,*

- *If $\bar{\pi}_o \leq v$, only honest types will contribute to the provision of the public good, and only honest types will be trusted.*
- *If $\bar{\pi}_o > v$, then*
  1. *If $p > p_{\mathsf{min}}$, then all agents will contribute to the provision of the public good, and all will be trusted with probability 1.*
  2. *If $p \leq p_{\mathsf{min}}$, then honest types will contribute with probability 1, while opportunists will contribute with probability $p/[a(1-p)]$. Agents who contribute will be trusted with probability $v/\bar{\pi}_o$, while agents who free-ride will not be trusted.*

[18] $\Pr(\textit{contribute}\,|\textit{opportunist})$ cannot exceed unity, since when $p = p_{\mathsf{min}} = a/(1+a)$, which is the highest p consistent with this case, $\Pr(\textit{contribute}\,|\textit{opportunist}) = 1$.

4 Endogenising the Proportion of Honest Types

I now endogenise the proportion of honest types by assuming an evolutionary mechanism which selects for the type that is relatively successful (has a higher lifetime material payoff). In particular, I assume that:

$$\frac{\Delta p}{\Delta g} = \phi(E\pi_h - E\pi_o), \tag{7}$$

where $\Delta p/\Delta g$ is the change in p from one generation to the next, $E\pi_h$ is the sum of the (undiscounted) material payoffs of the honest types, $E\pi_o$ is the sum of the material payoffs of the opportunists. Note that the *undiscounted* sum of the type's payoffs, rather than the (discounted) present value, is the relevant measure of the agent's fitness, since it is assumed that agents reproduce (or instill their preferences in their children) with equal probability throughout their careers. Thus the reproductive success (in either a biological or cultural sense) of a type depends simply on the number, unweighted by any discounting, of the number of children raised (or culturally influenced) by adult agents of that type. This assumption is standard in the socio-biological literature. It is further assumed that $\phi'(\cdot) > 0$ and $\phi(0) = 0$.

As buyers, all agents face the same population mixture and behave identically. Thus it is only as sellers that their payoffs can differ, due to their differing strategies depending on their type. The evolutionary analysis is based on the equilibrium strategies of the various types, based on Proposition 5.

Let us define a *critical point* as a point p on the segment $[0, 1]$ at which the lifetime, undiscounted expected payoffs of the honest types and opportunists are equal. Let us further define an *evolutionary equilibrium region* (EER) as a subsegment of contiguous critical points on the segment $[0, 1]$, such that for points in the neighborhood of this region, the difference in lifetime expected undiscounted payoffs will induce p to move into that region, by (7). Finally, define an *evolutionary equilibrium point* (EEP) as a point $p \in [0, 1]$ such that, for any other point p' in the neighborhood of p, the evolutionary selection dynamic (7) will lead the proportion of honest types to move in the direction of p over time.[19]

[19] See Martinez Coll and Hirshleifer (1991) for a discussion of these concepts.

4.1 Evolution of p in the Absence of a Public Good

As a benchmark for comparison, consider the case in which there is no public good. In this case, if $p > p_{\mathsf{min}}$, all agents will be trusted (Proposition 4). If, in addition, $r < -\ln(c)/\delta$, opportunists will honour trust up to (but not including) time $T-\delta$, and will cheat thereafter. Honest types will honour trust throughout their careers. Thus the opportunists' undiscounted lifetime payoffs will be $(T-\delta)(1-c)+1\cdot\delta$, while the honest types' lifetime payoffs will be $T(1-c)$. Therefore, the undiscounted lifetime payoffs of the honest types will be lower than those of the opportunists by $c\delta$. Thus, by (7), p will *decrease* from one generation to the next. If $p \leq p_{\mathsf{min}}$, no agents will be trusted, so that the lifetime payoffs of all agents will be zero. Thus, when $p \leq p_{\mathsf{min}}$, there will be no tendency for p to change from one generation to the next. We thus conclude that when $r < -\ln(c)/\delta$, there is an EER of $[0, p_{\mathsf{min}}]$.

If, on the other hand, $r \geq -\ln(c)/\delta$ and $p > p_{\mathsf{min}}$, opportunists will cheat over the time interval $[0, \delta]$, and will not be trusted thereafter. Thus the opportunists' undiscounted lifetime payoffs will be $1\cdot\delta$. The honest types will honour trust, and thus have lifetime payoffs of $T(1-c)$. We thus conclude that, if $p > p_{\mathsf{min}}$, p will increase (decrease) from one generation to the next if T is greater (less than) $\delta/(1-c)$. When $p \leq p_{\mathsf{min}}$, no agents will be trusted, so that the lifetime payoffs of all agents will be zero. Therefore, if $T > \delta/(1-c)$, there will be an EEP at $p=1$, while if $T < \delta/(1-c)$, there will be an EER of $[0, p_{\mathsf{min}}]$.

We may summarise these results in the following proposition.

Proposition 6. *In the absence of a public good, if $r < -\ln(c)/\delta$, there will be an EER at $[0, p_{\mathsf{min}}]$. In this region, agents will not be trusted. If $r \geq -\ln(c)/\delta$ and $T < \delta/(1-c)$, there will again be an EER at $[0, p_{\mathsf{min}}]$. But if $r \geq -\ln(c)/\delta$ and $T > \delta/(1-c)$, there will be an EEP at $p=1$.*

Since opportunists maximise their expected payoffs, it may be surprising that honest types can receive higher average payoffs than opportunists, as is the case if $r \geq -\ln(c)/\delta$ and $T > \delta/(1-c)$, over the interval $p \in (p_{\mathsf{min}}, 1]$. The reason for this result is that the payoffs relevant to the evolutionary stability of a given type are *undiscounted* expected payoffs, while opportunists maximise their *discounted* expected payoffs. Thus (discounted) expected-payoff-maximising agents cheat (and therefore reveal their type) too often in terms of their fitness—their lifetime undiscounted expected payoffs.

4.2 Evolution of p with a Public Good

In the analysis that follows, I focus on the non-trivial and, presumably, more empirically relevant case, $\bar{\pi}_o > v$. If this inequality does not hold, the only contributors will be the honest types, and the evolutionary results will be roughly similar to the ones obtained above, for the case in which there is no public good. In particular, if $r < -\ln(c)/\delta$, there will be an EER at $[0, p_{\mathsf{min}}]$, while if $r \geq -\ln(c)/\delta$, there will either be an EER at $[0, p_{\mathsf{min}}]$, if $T < (\delta + v)/(1 - c)$, or an EEP at $p = 1$, if $T > (\delta + v)/(1 - c)$.

If $\bar{\pi}_o > v$ and in addition $p > p_{\mathsf{min}}$, Proposition 5 states that all agents will contribute to the provision of the public good. If $r < -\ln(c)/\delta$, opportunists will then have higher lifetime material payoffs than honest types, given that, up to (but not including) time $T - \delta$, the two types' equilibrium behaviour is identical. Beginning at time $T - \delta$, the opportunists cheat in their market transactions, saving the cost $c\delta$, and thus giving the opportunists higher lifetime material payoffs. If $r \geq -\ln(c)/\delta$, opportunists will cheat at $t = 0$, while honest types honour trust throughout their careers. Thus the sum of the honest types' material payoffs is simply $(1 - c)T - v$, while the opportunists' material payoff is $1 \cdot \delta - v$. We therefore have, in this case, $E\pi_h > E\pi_o$ if $T > \delta/(1 - c)$. If, on the other hand, $T < \delta/(1 - c)$, the opportunists will have higher lifetime material payoffs. Thus, over the interval $p \in (p_{\mathsf{min}}, 1]$, our results are precisely the same as the results we obtained without a public good.

When $p \leq p_{\mathsf{min}}$, Proposition 5 states that honest types will contribute to the provision of the good with probability 1, while opportunists will contribute with probability $p/[a(1-p)]$. All agents who contribute will be trusted with probability $v/\bar{\pi}_o$, while those who do not contribute will not be trusted. Therefore the honest type's lifetime material payoff is

$$E\pi_h = (v/\bar{\pi}_o)(1 - c)T - v, \tag{8}$$

while the opportunist's lifetime material payoff is

$$E\pi_o = \begin{cases} [(v/\bar{\pi}_o)(T(1 - c) + c\delta) - v]\{p/[a(1 - p)]\} & \text{if} \quad r < -\ln(c)/\delta \\ [(v/\bar{\pi}_o)\delta - v]\{p/[a(1 - p)]\} & \text{if} \quad r \geq -\ln(c)/\delta \end{cases} \tag{9}$$

Define

$$\theta \equiv T(1 - c) - \bar{\pi}_o.$$

Multiplying (8) by $\bar{\pi}_o/v$,we have

$$\theta = \frac{E\pi_h \bar{\pi}_o}{v}.$$

Thus θ is positive if and only if $E\pi_h$ is positive.[20] Moreover, it is easily shown that $E\pi_o$ is positive.[21] Thus we require $E\pi_h > 0$ (and therefore $\theta > 0$) as a necessary condition for $E\pi_h - E\pi_o > 0$. Subtracting (9) from (8), equating to zero, and solving for p, we obtain [assuming $\theta > 0$]

$$E\pi_h - E\pi_o \gtreqless 0 \text{ as } \begin{cases} p \lesseqgtr \dfrac{a}{1 + a + \dfrac{c\delta}{\theta}} & \text{if } \quad r < -\ln(c)/\delta \\ p \lesseqgtr \dfrac{a}{a + \dfrac{\delta - \bar{\pi}_o}{\theta}} & \text{if } \quad r \geq -\ln(c)/\delta \end{cases} \tag{10}$$

Consider first the case $r < -\ln(c)/\delta$. Since $p_{\mathsf{min}} \equiv a/(1+a)$,

$$\frac{a}{1 + a + \dfrac{c\delta}{\theta}} < p_{\mathsf{min}} \text{ if } \theta > 0. \tag{11}$$

If $\theta > 0$, (7), (10) and (11) together imply that we have an EEP at $p = a/[1 + a + (c\delta/\theta)]$. We thus obtain

Proposition 7. *Suppose* $\bar{\pi}_o > v$ and $\theta > 0$. *If* $r < -\ln(c)/\delta$, *there will be a unique EEP at* $p = a/[1 + a + (c\delta/\theta)]$. *If* $\theta < 0$, *honest agents will not survive, in evolutionary equilibrium.*

Proof. See Appendix. $\square$

The intuition underlying the evolutionary stability of the honest type when $\theta > 0$ is that when p, the buyer's prior probability that the seller is honest, is relatively small, the signal 'emitted' by a contributor to the public good must be relatively 'powerful' in order to boost the buyer's posterior probability that the seller is honest up to p_{min}, giving the buyer a sufficient incentive to trust. Since honest types always contribute, this signal can only be sufficiently informative or powerful if only a small proportion of opportunists contribute. Given that only a small proportion of opportunists contribute,

[20] See Section 4.3 for a demonstration that θ will be positive for reasonable parameter values.

[21] See the proofs of Propositions 7 and 8.

on the average honest types will have many more market transactions and thus higher lifetime payoffs.

Now consider the case $r \geq -\ln(c)/\delta$. There are now four subcases, specified in Table 1.

Table 1. Evolutionary Equilibria when $r \geq -\ln(c)/\delta$

	$\dfrac{a}{a+\dfrac{\delta-\bar{\pi}_o}{\theta}} \geq p_{\mathsf{min}}$	$\dfrac{a}{a+\dfrac{\delta-\bar{\pi}_o}{\theta}} < p_{\mathsf{min}}$
$T > \delta/(1-c)$	EEP at $p = 1$	EEPs at $p = \dfrac{a}{a+\dfrac{\delta-\bar{\pi}_o}{\theta}}$ and $p = 1$
$T < \delta/(1-c)$	EEP at $p = p_{\mathsf{min}}$	EEP at $p = \dfrac{a}{a+\dfrac{\delta-\bar{\pi}_o}{\theta}}$

The explanation of Table 1 is as follows. The rows correspond to the two cases analysed above, which determine the change in p from one generation to the next when initially $p > p_{\mathsf{min}}$. If $T > \delta/(1-c)$, we have $E\pi_h > E\pi_o$, and conversely when $T < \delta/(1-c)$. The columns determine the evolution of p when initially $p < p_{\mathsf{min}}$, using (7) and (10). Thus, for example, in the upper left-hand cell, we have $E\pi_h > E\pi_o$ both when $p \leq p_{\mathsf{min}}$ and when $p > p_{\mathsf{min}}$. Therefore, there is a unique EEP at $p = 1$. The information in the remaining cells is determined similarly. Summarising,

Proposition 8. *Suppose* $\bar{\pi}_o > v$ and $\theta > 0$. *When* $r \geq -\ln(c)/\delta$, *the evolutionary equilibria are as specified in Table 1.*

Proof. See Appendix. ¥

4.3 Discussion

We conclude that in all cases, in the *presence* of a public good, the EEP will be sufficiently large to support trust. When the evolutionary equilibrium p is no greater than p_{min}, the probability of trust is $v/\bar{\pi}_o$ and is therefore increasing with the cost of contributing to the public good, v. Moreover, in this equilibrium, opportunists contribute to the provision of the public good with probability $p/[a(1-p)]$.

In contrast, in the *absence* of a public good, in most cases there is only an EER at $[0, p_{\mathsf{min}}]$. In this region, no agents are trusted. Only if $r \geq -\ln(c)/\delta$

and $T > \delta/(1-c)$, there will be an EEP at $p = 1$. Even in this case, there is an 'inert' region, $[0, p_{\mathsf{min}}]$, in which the payoffs of both types are equal at zero, and there is no tendency for p to change from one generation to the next. Only if p were initially not in this region, or were to escape somehow through (unmodelled) mutations, would trust emerge—and this is true, to repeat, only in the 'best case.'

The contrast between our results, with and without the public good, becomes sharper when we take account of the fact that the case $r < -\ln(c)/\delta$ is much more empirically plausible than the case $r \geq -\ln(c)/\delta$. To see this, let us define the unit of time as a year. Figure 2 shows the locus dividing between the parameter spaces consistent with these two cases, for $r = 0.05$. (For smaller discount rates, the curve would be steeper, to the right of the curve shown.) I show δ varying from 0 to 2 (an unreasonably high value, implying that it takes 2 years for information of an agent's cheating to spread to the entire community). The area to the left of the curve is the parameter space consistent with $r < -\ln(c)/\delta$, while the area to the right of the curve is the parameter space consistent with $r > -\ln(c)/\delta$. As the figure makes clear, c, the cost of producing the high-quality private good, would have to be very close to its maximum value, unity, in order for $r \geq -\ln(c)/\delta$ to hold, particularly for reasonable values of δ.

Given that $r < -\ln(c)/\delta$ is the more empirically plausible case, we find that without the public good, the unique evolutionary equilibrium region is $[0, p_{\mathsf{min}}]$ (Proposition 6), which cannot support trust, while in the presence of the public good, there is a unique EEP at $a/[1 + a + (c\delta/\theta)]$ (Proposition 7), which supports trust in the equilibrium of the repeated market trust game.

The effect of the public good in enhancing the evolutionary stability of the honest type is due to its effect on the evolutionary dynamics in the region $[0, p_{\mathsf{min}}]$. Without the public good, there is no trust in this region, and therefore p does not change from one generation to the next. With the public good, a mixed strategy signalling equilibrium emerges (provided that $\theta > 0$) in which honest types contribute with higher probability than opportunistic types and thus have the benefit of more frequent market transactions. The payoff from these transactions outweighs the higher cost (on the average) of producing the high-quality private good, when p is sufficiently small and thus the equilibrium probability that the opportunists contribute $p/[a(1-p)]$ is sufficiently low.

Recall that evolutionary fitness depends on agents' *undiscounted* payoffs, while the opportunist maximises his or her *discounted* payoff. This explains

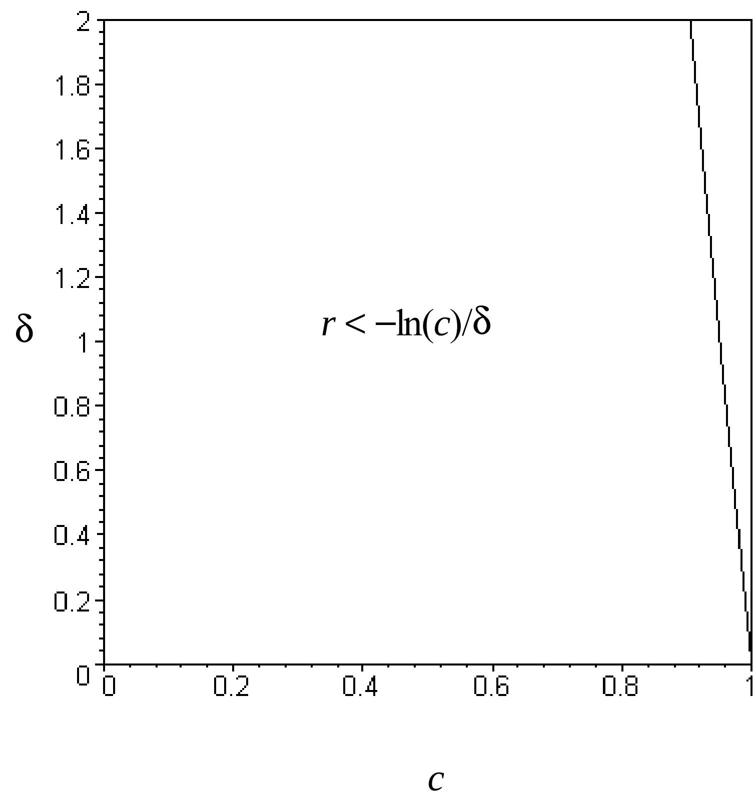

Figure 2: Parameter Spaces for $r < \ln(c)/\delta$ and $r > \ln(c)/\delta$; $r = 0.05$

the importance of the assumption, which appears empirically plausible, that contributions to the provision of the public good do not take place continuously over time, but only periodically. (I use the simplifying assumption that contributions take place only at $t = 0$, but similar results would be obtained as long as contributions do not take place continuously over time.) This feature of the model implies that opportunists, who maximise their *discounted expected payoffs*, contribute 'too little' to the provision of the public good in terms of maximising their evolutionary *fitness*, enabling the survival of the honest type. This bias in the opportunists' decisions whether or not to contribute (which, again, is rational in terms of their objective of maximising discounted expected payoffs) is critical only in the region $[0, p_{\mathsf{min}}]$, in which the mixed strategy equilibrium implies that opportunists are at the margin of indifference between contributing and free-riding.

In order to assess the realism of the condition $\theta > 0$ [thus making the honest type evolutionarily stable, when $r < -\ln(c)/\delta$], Figure 3 shows the minimum T, denoted T_{min}, required to make θ positive, for various values of r. The figure is drawn for $c = 0.5$ and for two values of δ, $1/12$ (one month) and 1 (one year). As the figure shows, a career length of only about 13.5 years is required for the honest type to be evolutionarily stable even when δ is rather implausibly large (one year), for a discount rate of only 1 percent. As r increases or as δ decreases, the minimum career length becomes even smaller. As c increases, T_{min} also increases, but even if $c = 0.9$ and $r = 0.01$, T_{min} is only approximately 37.5 years.

5 Some Empirical Evidence

Let us return momentarily to Veblen's acute observation, quoted at the beginning of this paper. As Veblen (1899, pp. 340–41) noted,

> In the case of some considerable groups of organisations or establishments... the invidious motive is apparently the dominant motive both with the initiators of the work and with their supporters. This last remark would hold true especially with respect to such works as lend distinction to their doer through large and conspicuous expenditure; as, for example, the foundation of a university or of a public library or museum...

Veblen (1899) was careful to add, however, that the very fact that such contributions enhance the reputation of the contributor implies that there

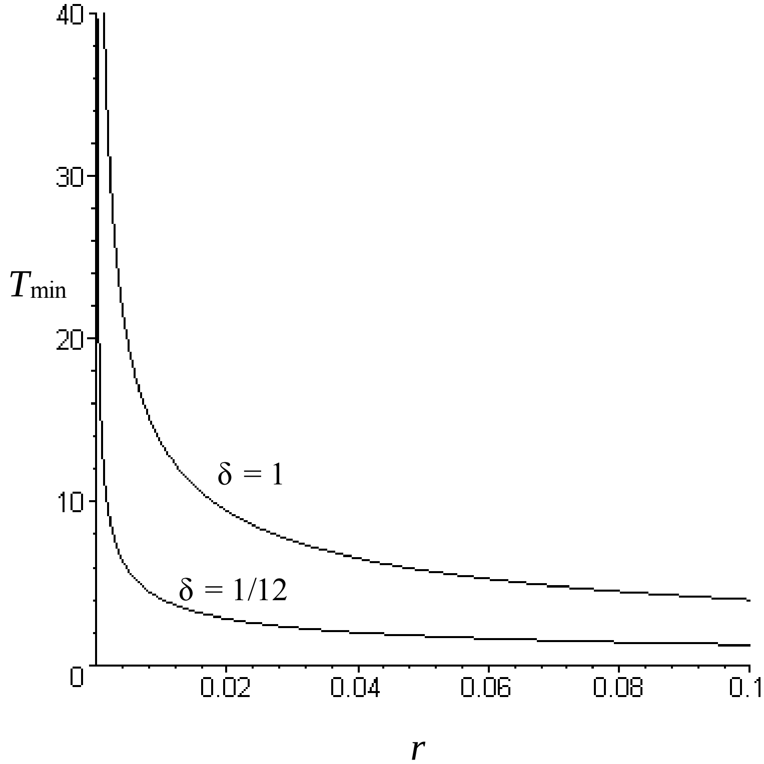

Figure 3: Minimum T for $\theta > 0$

is a presumption that some individuals (honest types in the present model) undertake such activities for non-selfish reasons:

> The fact itself that distinction or a decent good fame is sought by this method is evidence of a prevalent sense of the legitimacy, and of the presumptive effectual presence, of a non-emulative, non-invidious interest, as a constituent factor in the habits of thought of modern communities.

A recent field experiment (Soetevent, 2004) shows that when Dutch churches collect money allowing members' contributions to be visible to their neighbors on the church bench, contributions increase in comparison to situations in which such contributions cannot be seen. Similarly, Hoffman, et al. (1994, 1996) find that increased anonymity decreases—but does not eliminate—generosity in the dictator game.

The signalling motive for contributions to the provision of public goods may explain the anomalously generous behaviour of victims of natural disasters. Hirshleifer (1987, p. 135) cites evidence provided by Konreuther and Dacy (1969) of market behaviour after the Alaskan earthquake of 1964:

> In circumstances where prices might have been expected to rise sharply, they were unchanged or actually fell, remaining at abnormal levels for a period of weeks or months. The pattern held even for commodities in especially urgent demand, e.g. milk and canned juices in a period when public water supplies were out of commission. Nor was this a matter of informal rationing by the merchant, at a disequilibrium price; rather, the evidence suggests that the customers rationed themselves, taking no more than some minimal or 'equitable' quantity from the shelves. Investigating further, Dacy and Konreuther found similar phenomena in other disasters: refugees from floods or tornadoes, for example, have received shelter in private homes, sometimes for periods of months, at little or no charge (e.g. the Dutch floods of 1953).

Hirshleifer (1987, p. 138) suggests that such behaviour can be understood as private contributions to the provision of a public good, 'maintaining that *alliance* we call society.' But such contributions can themselves be viewed as signals that the contributor is an honest, civic-minded individual, signals that will strengthen the agent's reputation in normal times.

Individuals who regularly endure particularly harsh climates, for example the inhabitants of the Scandinavian countries in Europe and the Midwestern states in the United States, have regular opportunities to exhibit the 'public spirit' shown by disaster victims. Snowstorms, a regular part of winters in such locations, provide ample opportunities to contribute to the 'alliance we call society.' Helping another individual whose car is stuck in a snowdrift and shovelling snow from public sidewalks adjacent to one's home are examples of such behaviour. Given the presence of regular opportunities to voluntarily provide public goods in such locations, our model would predict that the evolutionary stability of the honest type would be enhanced. This prediction is supported by the unusually high levels of trust observed in the Scandinavian countries and in the Midwestern states (Alesina and La Ferrara, 2002; Uslaner, 2002; Guttman and Surana, 2004).

6 Concluding Remarks

This paper has developed a theory of the voluntary provision of public goods that combines the insights of two, distinct traditions: (a) the conventional game-theoretic literature of repeated games with incomplete information, in which players develop reputations which stem initially from (assumed) uncertainty of one player regarding the other player's type, and (b) the indirect evolutionary literature, in which player types are defined not by wired-in *strategies*, but rather by their *preferences,* and agents choose strategies to maximise their expected payoffs, as in standard economic theory.

The present analysis illustrates the complementarity of these two theoretical traditions. The evolutionary stability of agents with non-opportunistic preferences is easier to understand when the interaction of agents is modeled as a repeated game in which both private-good and public-good decisions are made. Conversely, the repeated-game reputational mechanism does not need to rely on *ad hoc* assumptions regarding players' prior beliefs when these beliefs are endogenised by making them correspond to the population proportions of the various types, generated by an evolutionary process.

The theory utilises the fact that the voluntary provision of public goods usually takes place in a broader social context, in which players also buy and sell private goods. Contributing to the provision of public goods, in such contexts, serves as a signal of the trustworthiness of the contributor, a signal that is important in obtaining trust in private good interactions. Thus, in the model, opportunistic agents contribute to the provision of public goods in order to preserve reputations as honest agents. But these reputations could not be established if there were no *true* honest types in the population, whose existence creates uncertainty by agents as to their partner's type. The model explains why these true honest types can persist in the population, despite their vulnerability to exploitation by opportunists.

Appendix

Proof of Proposition 2

Calculating the integrals in (1) of the text, we have

$$\begin{aligned}\pi_o &= \frac{(1-c)(1-e^{-r\bar{t}})}{r} + \frac{e^{-r\bar{t}} - e^{-r(\bar{t}+\delta)}}{r} \\ &= \frac{1-c}{r} + \frac{ce^{-r\bar{t}}}{r} - \frac{e^{-r(\bar{t}+\delta)}}{r}.\end{aligned}$$

Differentiating w.r.t. $\bar{t}$,

$$\frac{\partial \pi_o}{\partial \bar{t}} = e^{-r(\bar{t}+\delta)} - ce^{-r\bar{t}}.$$

Dividing through by $e^{-r\bar{t}}$, we obtain

$$\frac{\partial \pi_o}{\partial \bar{t}} \gtreqless 0 \text{ as } e^{-r\delta} \gtreqless c.$$

Taking logs of both sides of the right-hand expression, we obtain

$$\frac{\partial \pi_o}{\partial \bar{t}} \gtreqless 0 \text{ as } r \lesseqgtr -\frac{\ln c}{\delta}.$$

Proof of Proposition 7

From (8) and (9) in the text, we have (in the case $r < -\ln(c)/\delta$)

$$\begin{aligned}E\pi_h - E\pi_o &= \left[\left(\frac{v}{\bar{\pi}_o}\right)(1-c)T - v\right] \\ &\quad - \left[\left(\frac{v}{\bar{\pi}_o}\right)(T(1-c)+c\delta) - v\right]\left(\frac{p}{a(1-p)}\right).\end{aligned}$$

As noted in the text, $E\pi_h$ is positive if and only if $\theta > 0$. Note further than $T(1-c)+c\delta$, the undiscounted sum of the payoffs of the opportunistic type if trusted, must be greater than $\bar{\pi}_o$, the same type's discounted payoff if trusted. Therefore the second bracketed term must be positive. It follows that a necessary condition for $E\pi_h - E\pi_o$ to be positive is that $\theta > 0$. If indeed $\theta > 0$, then when $p = 0$, $E\pi_h - E\pi_o > 0$.

Since the second bracketed term is positive, it is clear that $\partial(E\pi_h - E\pi_o)/\partial p < 0$. It follows that if $\theta > 0$, $E\pi_h - E\pi_o$ will go from positive to negative values as p increases from 0 to 1. Therefore the critical point where $E\pi_h - E\pi_o = 0$ is an EEP.

Setting $E\pi_h - E\pi_o = 0$, solving for p, and dividing through by θ, we obtain the evolutionary equilibrium p given in the proposition.

Proof of Proposition 8

From (8) and (9) in the text, we have (in the case $r \geq -\ln(c)/\delta$)

$$E\pi_h - E\pi_o = \left[\left(\frac{v}{\bar{\pi}_o}\right)(1-c)T - v\right] - \left[\left(\frac{v}{\bar{\pi}_o}\right)\delta - v\right]\left(\frac{p}{a(1-p)}\right).$$

As noted in the text, the first bracketed term will be positive if and only if $\theta > 0$. Therefore, if $p = 0$, $E\pi_h - E\pi_o$ will be positive if $\theta > 0$. Moreover, δ, which is the undiscounted payoff of the opportunistic type if trusted, must be greater than $\bar{\pi}_o$, the same type's discounted payoff if trusted. Therefore the second bracketed term is positive. Thus, by the same argument used above, $E\pi_h - E\pi_o$ will go from positive to zero values as p rises from 0 to 1.

The critical p at which $E\pi_h - E\pi_o = 0$,

$$\hat{p} \equiv \frac{a}{a + \dfrac{\delta - \bar{\pi}_o}{\theta}},$$

is derived by setting the above expression equal to zero and solving for p. If $\hat{p} > p_{\mathsf{min}}$, we have $E\pi_h > E\pi_o$ for all $p \leq p_{\mathsf{min}}$. In this case, p will rise, from one generation to the next, until it becomes greater than p_{min}. At that point in time, p will continue to increase if $T > \delta/(1-c)$. Therefore, if $\hat{p} > p_{\mathsf{min}}$ and $T > \delta/(1-c)$, we obtain an EEP at $p = 1$. If, on the other hand, $\hat{p} > p_{\mathsf{min}}$ and $T < \delta/(1-c)$, p will increase when $p \leq p_{\mathsf{min}}$ but will decrease when $p > p_{\mathsf{min}}$. There will then be an EEP at p_{min}.

If $\hat{p} < p_{\mathsf{min}}$, we obtain an EEP at $\hat{p}$. If, in addition, $T < \delta/(1-c)$, then p will decrease from one generation to the next if p initially is greater than p_{min}. In this case, the EEP at $\hat{p}$ will be unique. If, on the other hand, $T > \delta/(1-c)$, then p will increase from one generation to the next if p initially is greater than p_{min}. In this case, there will be a second EEP at $p = 1$.

References

Akerlof, G.A. (1983). 'Loyalty filters', *American Economic Review,* vol. 73, pp. 54-63.

Alesina, A., and La Ferrara, E. (2002). 'Who trusts others?' *Journal of Public Economics,* vol. 85, pp. 207-34.

Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Becker, G.S. (1991). *A Treatise on the Family,* 2nd edition, Cambridge, MA: Harvard University Press.

Becker, G.S. (1993). 'Nobel lecture: the economic way of looking at behavior', *Journal of Political Economy,* vol. 101, pp. 385-409.

Bernstein, L. (1992). 'Opting out of the legal system: extralegal contractual relations in the diamond industry', *Journal of Legal Studies*, vol. 21, pp. 115-57.

Bisin, A. and Verdier, T. (1998). 'On the cultural transmission of preferences for social status', *Journal of Public Economics,* vol. 70, pp. 75-97.

Bolton, G.E., Ockenfels, A. (2000). 'ERC: a theory of equity, reciprocity, and competition', *American Economic Review,* vol. 90, pp.166-193.

Boyd, R., and Lorberbaum, J.P. (1987). 'No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game', *Nature,* vol. 327, pp. 58-59.

Boyd, R., and Richerson, P.J. (1985). *Culture and the Evolutionary Process,* Chicago: University of Chicago Press.

Camerer, C. (1988). 'Gifts as economic signals and social symbols', *American Journal of Sociology,* vol. 94, pp. S180-S214.

Carmichael, H.L., and MacLeod, W.B. (1997). 'Gift giving and the evolution of cooperation', *International Economic Review,* vol. 38, pp. 485-509.

Ellickson, R.C. (1991). *Order Without Law: How Neighbors Settle Disputes,* Cambridge, Mass. and London: Harvard University Press.

Frank, R.H. (1987). 'If *Homo Economicus* could choose his own utility function, would he want one with a conscience?' *American Economic Review,* vol. 77, pp. 593-604.

Frank, R.H. (1988). *Passions Within Reason: The Strategic Role of the Emotions.* New York: Norton.

Güth, W. (1995). 'Incomplete information about reciprocal incentives: an evolutionary approach to explaining cooperative behavior', *International Journal of Game Theory,* vol. 24, pp. 323-44.

Güth, W., and Tietz, R. (1990). 'Ultimatum bargaining behavior – a survey and comparison of experimental results', *Journal of Economic Psychology,* vol. 11, pp. 417-49.

Güth, W., and Yaari, M. (1992). 'An evolutionary approach to explaining reciprocal behavior in a simple strategic game', in (U.Witt, ed.), *Explaining Process and Change: Approaches to Evolutionary Economics*, Ann Arbor: University of Michigan Press.

Guttman, J.M. (2001a). 'Self-enforcing reciprocity norms and intergenerational transfers: theory and evidence', *Journal of Public Economics,* vol. 81, pp. 117-51.

Guttman, J.M. (2001b). 'Families, markets, and self-enforcing reciprocity norms', *Annales d'Economie et de Statistique,* vol. 63-64, pp. 89-110.

Guttman, J.M. (2003). 'Repeated interaction and the evolution of preferences for reciprocity', *Economic Journal,* vol. 113, pp. 631-56.

Guttman, J.M., Nitzan, S., and Spiegel, U. (1992). 'Rent seeking and social investment in taste change', *Economics and Politics*, vol. 4, pp. 31-42.

Guttman, J.M. and Surana, S. (2004). 'Public good contributions as signals: reputation, trust, and the evolution of altruistic preferences', Unpublished working paper, Bar-llan University.

Guttman, J.M. and Yacouel, N. (2004). 'Economies of scale in reputation and the theory of the firm', Unpublished working paper, Bar-Ilan University.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). 'Preferences, property rights, and anonymity in bargaining games', *Games and Economic Behavior*, vol. 7, pp. 346-80.

Hoffman, E., McCabe, K., and Smith, V.L. (1996). 'Social distance and other-regarding behavior in dictator games', *American Economic Review*, vol. 86, pp. 653-60.

Hirshleifer, J. (1987). *Economic Behaviour in Adversity*, Chicago: University of Chicago Press, ch. 4.

Katz, E., and Rosenberg, J. (2005). 'An economic interpretation of institutional volunteering', *European Journal of Political Economy*, vol. 21, pp. 429-43.

Konreuther, H. and Dacy D. (1969). 'The peculiar economics of disaster,' in D. Dacy and H. Konreuther, *The Economics of Natural Disasters: Implications for Federal Policy,* New York: Free Press.

Kranton, R.E. (1996). 'The formation of cooperative relationships', *Journal of Law, Economics, and Organization,* vol. 12, pp. 214-33.

Kreps, D.M., Milgrom, P., Roberts, J., and Wilson, R. (1982). 'Rational cooperation in the finitely repeated prisoners' dilemma', *Journal of Economic Theory,* vol. 27, pp. 245-52.

Ledyard, J.O. (1995). 'Public goods: a survey of experimental research' in (J.H. Kagel and A.E. Roth, eds.), *Handbook of Experimental Economics*, Princeton: Princeton University Press

Macauley, S. (1963). 'Non-contractual relations in business: a preliminary study', *American Sociological Review*, vol. 94, pp. 1010-41.

Martinez Coll, J.C. and Hirshleifer, J. (1991). 'The limits of reciprocity', *Rationality and Society,* vol. 3, pp. 35-64.

Olson, M., Jr. (1983). ‘The political economy of comparative growth rates’, in (D.C. Mueller, ed.), *The Political Economy of Growth*, New Haven: Yale University Press.

Orndorff, R. (2000). *The Unofficial Guide to the Perfect Job*, Foster City, Calif.: ARCO, IDG Books.

Samuelson, L. (2001). ‘Introduction to the evolution of preferences’, *Journal of Economic Theory* vol. 97, pp. 225-30.

Schaefer, N. (2000). *Good Job: A Young Person’s Guide to Finding, Landing, and Loving a Job*, Toronto: Stoddart.

Smith, V.L. (2003). ‘Constructivist and ecological rationality in economics’, *American Economic Review,* vol. 93, pp. 465-508.

Soetevent, A.R. (2004). ‘Anonymity in giving in a natural context — a field experiment in thirty churches’, *Journal of Public Economics*, forthcoming.

Stark, O. (1995). *Altruism and Beyond: An Economic Analysis of Transfers within Families and Groups,* Cambridge: Cambridge University Press.

Uslaner, E.M. (2002). *The Moral Foundations of Trust*, Cambridge: Cambridge University Press.

Veblen, T. (1899). *The Theory of the Leisure Class,* New York: Macmillan.