

Public Goods, Social Norms and Naïve Beliefs

Edward Cartwright
Department of Economics,
Keynes College,
University of Kent,
Canterbury,
Kent. CT2 7NP. UK.
E.J.Cartwright@kent.ac.uk

Amrish Patel
Department of Economics,
Keynes College,
University of Kent,
Canterbury,
Kent. CT2 7NP. UK.
ap291@kent.ac.uk

4th December 2008

Keywords: social norm, signalling game, naive inference, public good

Abstract

An individual's contribution to a public good may be seen by others as a signal of attributes such as generosity or wealth. An individual may, therefore, choose their contribution so as to send an appropriate signal to others. In this paper we question how the inferences made by others will influence the amount contributed to the public good. Evidence suggests that individuals are naïve and biased towards taking things at "face value". We contrast, therefore, contributions made to a public good if others are expected to make rational inferences versus contributions if others are expected to make naïve inferences.

1 Social Exchange and Public Goods

A person's behavior is often influenced by an anticipation of social approval or disapproval from others. For example, the worker may fear stigma if he claims unemployment benefit (Lindbeck, Nyberg and Weibull 1999), the smoker may expect disapproval if he smokes at the dinner table (Nyborg and Rege 2003), the customer may hope to appear generous if he leaves a big tip (Azar 2004), the worker may fear contempt from colleagues if he works through a strike (Francis 1985) or the voter may be seen as a good citizen by turning out to vote (Knack 1992). One particular context where such 'social exchange' can prove both important and beneficial is collective action problems. Numerous authors have argued that, in collective action problems, voluntary cooperation can result from an expectation of social approval, or fear of social disapproval (e.g. Olson 1965, Andreoni 1990, Hollander 1990, Kandel and Lazear 1992, Fehr and Gächter 1999, Rege and Telle 2002 and Rege 2004). More generally, it has been argued that a person may invest resources in a public project in order to signal something good about themselves, such as, wealth, status or altruism (Glazer and Konrad 1996, Harbaugh 1998a, 1998b and Hardy and van Vugt 2006).

A general framework in which to model social exchange is as a signalling game (Bernheim 1994, Glazer and Konrad 1996 and Bernheim and Serverinov 2003, Benabou and Tirole 2006, Andreoni and Bernheim 2008, Tadelis 2008, Grossman 2008). The basic setting is that somebody, call him A, performs some *action*, such as to contribute or not to a public project, and this is observed by someone else, call him B. Once B has seen the action taken by A he may want to 'reward' or 'punish' A for being, say, generous or lazy. Agent *type*, whether A actually is generous or lazy, is not observed by B and so B has to *infer* A's type from his action. Crucially, for our purposes, this means A should take into account how B will interpret his actions. For example, if A contributes to a public good, will B infer that A is generous, or infer that he is not generous but trying to appear generous? Clearly, this will effect A's incentive to contribute to the public good.

The modelling of social exchange as a signalling game provides a rich

framework within to work. To the best of our knowledge, however, the previous literature has focused exclusively on Perfect Bayesian Equilibria as a characterization of behavior. This requires, to return to the earlier setting, that B will correctly interpret the meaning of A's action and A will choose an action to optimize his payoff on the assumption that B will correctly interpret his action. This puts relatively high expectations on the reasoning abilities of both A and B. Is it reasonable to assume such ability? The motivation for the current paper is a belief that in many settings people may have a bias towards 'naïve inferences' and fail to take into account the effect that incentives to seek social approval will have on actions. For example, if A contributes to a public good then, while it may be 'rational' for B to infer that A is not generous but merely trying to signal generosity, we suggest that a 'naïve' B may take things at 'face value' and infer a positive contribution as being a signal of generosity.

What we call 'naïve' inferences mean that a person can make systematic mistakes in interpreting the motivations or types of others. The naïve agent will, for example, constantly overestimate the generosity of a person who contributes to a public good. One might expect such a bias to disappear over time. There are, however, good reasons to suppose that it may not. First, it may be costly in terms of time or effort to form correct inferences over why someone has acted the way that they have, but the benefits from a correct inference may be relatively small. A simple cost-benefit analysis suggests therefore that 'taking things at face value' may be optimal. Second, trust is a cornerstone of all societies and relationships within them (Frank 1988, Coleman 1990 and Knack and Keefer 1997). If, therefore, somebody 'says that they are generous' there is a bias towards believing them. It may be more costly to not trust somebody, even if this means that type may be inferred correctly, than to trust somebody, even if wrong inferences are made (Frank 1988). Experimental studies also support the notion that people can systematically make errors in interpreting the actions of others. This could be because of an overconfidence in private information (Rabin and Schrag 1999, Charness and Levin 2005), or a fundamental attributions error that over weights the importance of observed behavior relative to situational

causes (Jones and Harris 1967), or simply an inability to reason through the thought processes of others (Thaler 1988, Eyster and Rabin 2005).

To better understand the potential consequences of naïve inferences we model a setting with three people, call them A, B and C. Person A chooses how much to invest in a public good. Having seen this investment B and C make an inference on A's type and give 'esteem' based on perceived level of generosity. The more generous the agent is perceived to be the more esteem they give. We shall assume that B makes 'Bayes rational' inferences while C makes 'naïve' inferences. The relative weight that A puts on the esteem of B and C can be varied. If A only cares about the esteem of the rational person B then we have the standard setting in which a Perfect Bayesian Equilibrium is obtained (Bernheim 1994). If, however, A puts positive weight on the esteem of the naïve person C then we depart from the standard setting in the sense that C will make naïve inferences and furthermore A will base his action on the basis that C will make naïve inferences.

Our interest is in how the naïvety of C's inferences will influence the behavior of A. More generally, this is a question of how naïvety in inferences, or an expectation of naïvety in inferences, will influence behavior. Does for example, a person invest more or less in a public project depending on how he expects others to interpret his investment. Given that people's investments in public projects, or conformity to norms more widely, appear to be motivated in large part by expected social approval or disapproval, one would expect that differences in how actions are interpreted can have important and interesting consequences for behavior that need to be taken into account. In this paper we shall show that naïvety in how inferences are made can have important and sometimes unexpected consequences on behavior.

To illustrate, in equilibrium the rational person B will often correctly infer the type of A. In particular, he understands the incentives of A to appear generous and so factors this into his interpretation of A's investment. Person A can, however, easily 'fool' the naïve agent C into thinking that he is generous through a high investment. This might suggest that there are incentives for A to invest more the more weight is put on the esteem of the

naïve agent C. We shall see, however, that *investment is often higher the more weight is put on the esteem of the rational person*. This is because to ‘earn’ esteem from the rational person B it is necessary for A to invest a sufficiently high amount to signal his true generosity. The rational inferences or ‘cynicism’ of B therefore lead A to increase his investment in order to get the esteem ‘that his true generosity deserves’. For what it is worth, we were personally very surprised, working through the model, to see how the presence of naïve inferences often *lowered* the investment of A.

We do also find, however, that the presence of naïve inferences can sometimes increase the investment of A in a way one may not expect. In particular, if an upper bound is imposed on the level of investment then a pooling equilibrium may result in which a person of a certain generosity or above invests the maximum amount. This means that B would not be able to infer the true generosity of A when A invests the maximum. It also means that B gives a relatively high esteem to anyone investing the maximum and much less esteem to anyone investing less than the maximum. We can look at the minimum level of generosity such that A would invest the maximum. There are good reasons to expect that the more weight A puts on the esteem of the rational person B the lower this minimum level would be. We show, however, that *increasing the weight on the esteem of the naïve person can sometimes lower the level of generosity for which A would invest the maximum amount*. The intuition for this result is that a less generous A will increase his investment, the more weight he puts on the esteem of the naïve person C, in order to ‘fool’ C into thinking he is more generous than he is. But, having increased his investment, he is that much closer to investing the maximum and so may, in order to get more esteem from rational person B, go ‘all the way’ and invest up to the maximum. This serves to illustrate how naïve and rational inferences can ‘complement’ each other to significantly change A’s investment.

It is important to recognize that the concept of Perfect Bayesian Equilibrium has been relaxed in many contexts to equilibria in which actions are optimal given inferences but inferences may be biased in some way relative to actions. It has been suggested, for instance, that individuals think ‘coarsely’

or by ‘analogy’, forming beliefs about groups of situations rather than for each specific situation (Jehiel 2005, Jehiel and Koessler 2008, Ettinger and Jehiel 2008, Mullainathan et al. 2008). Even if inferences are made for individual situations, it may be that they are correct on the equilibrium path of play, but incorrect off the equilibrium (Fudenberg and Levine 1993). But, they need not be correct on the equilibrium path if individuals either ignore how other people’s information affects their actions (Eyster and Rabin 2005) or ignore the informational content contained within an action (Esponda 2008). Finally, rather than just ignoring actions and information, agents may make take them into account but still make the wrong inference by taking actions at face value (Bodner and Prelec 2003, Eyster and Rabin 2008). It is this last possibility that is closest to what we consider in this paper.

We proceed as follows: Section 2 sets out the model, Section 3 details possible equilibria and Section 4 discusses the consequences of naïve beliefs before we conclude in Section 5. Proofs are contained in an Appendix.

2 The model

The model we use is inspired by Bernheim (1994). There are three agents A, B and C of which A is of primary interest. It will become clear as we proceed that each agent could be interpreted as representing many identical agents. Agent A is characterized by a type $t \in [0, 1]$ which we call his *level of generosity* and chooses how much to *invest in a public project* from set $X := [0, 1]$. If he is of type t and chooses to invest x then he receives *intrinsic utility* $-(x - t)^2$. The closer, therefore, is investment to level of generosity then the higher is intrinsic utility. Note that intrinsic utility does not depend on others but merely on the type of the agent. Note also that the level of generosity and of investment could be interpreted in many different ways and so framing the discussion in terms of public projects is one illustrative interpretation.

Once A has chosen an investment B and C accord A *esteem*. Esteem may equate with direct utility if, for example, agent A enjoys being seen as

generous. Or it may be indirect via some instrumental motive, for example, agent A may receive favours from B and C if they perceive him as generous. Both agents B and C base esteem on how generous they believe A to be. Specifically, if A is perceived to have type b by agent B or C , then B or C respectively, would accord A esteem $-(1-b)^2$. So, type $t = 1$ is the ideal type and the further is believed to be agent A 's type from this ideal then the lower the esteem. Basically, the more generous agent A is perceived to be, the more esteem he is given.

It is assumed that agent A 's type is private information. Some inference, therefore, must be made by agents B and C about A 's type based on his investment. An *inference function* $\phi(b, x)$ details a probability distribution over types for each possible investment x where

$$\int_T \phi(b, x) db = 1 \quad \text{for all } x \in X.$$

Informally, $\phi(b, x)$ is the probability that agent A is inferred to be of type b if he invests x . The utility of agent A is then given by

$$U(t, x, \phi_B, \phi_C) = -(x-t)^2 - E\theta \int_T (1-b)^2 \phi_B(b, x) db - E(1-\theta) \int_T (1-b)^2 \phi_C(b, x) db, \quad (1)$$

for some real numbers $E, \theta \geq 0$ where ϕ_B and ϕ_C are the inference functions of B and C . We think of E as the overall weight of esteem relative to intrinsic utility, while θ is the relative weight given to the esteem of agent B .

Throughout, we assume that agent B has a *Bayes rational inference function* ϕ_B to be defined below. Agent C , by contrast, is assumed to have a *naïve inference function* in the sense that

$$\phi_C(b, x) = \begin{cases} 1 & \text{if } b = x \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

If therefore, agent A chooses investment x agent C 'naïvely' assumes that A 's level of generosity is x . We shall discuss in more detail below why agent

C is naïve but clearly he fails to take into account that A may sacrifice intrinsic utility in search of esteem. Given that agent C has naïve inferences the utility function of A can be re-written

$$U(t, x, \phi_B, \phi_C) = -(x - t)^2 - E(1 - \theta)(1 - x)^2 - E\theta \int_T (1 - b)^2 \phi_B(b, x) db. \quad (3)$$

The utility of agents B and C is of no interest to us here and so we do not model why they care about A 's type. But, in interpretation we can think of agent B as being someone who either has an incentive to infer correctly the type of A or has experience in inferring type. This agent, is therefore, not easily fooled into thinking A is more generous than he actually is. Agent C by contrast takes things at 'face value' and so is easily 'fooled' into thinking that A is more than generous than he is.

3 Signalling equilibria

As in Bernheim (1994) we shall look for signalling equilibria that satisfy the D1 Criterion.¹ A signalling equilibrium basically requires actions to be optimal given inferences and inferences to be Bayes rational given actions. Clearly, we shall only require this of agents A and B and not C . Equilibrium will, therefore, be characterized by an investment function μ , that maps T into X and details the investment of agent A for each possible type, and an inference function ϕ_B .² We say that investment function μ and inference function ϕ_B are a *signalling equilibrium* if and only if:

1. $U(t, \mu(t), \phi_B, \phi_C) \geq U(t, x, \phi_B, \phi_C)$ for all $t \in X$ and $x \in X$.
2. ϕ_B is consistent with μ , a uniform prior on the type of agent A and the D1 Criterion.

The requirements of condition 2 will become clearer as we proceed. At this stage we discuss agent B 's inferences if he perceives that agent A would

¹See Fudenberg and Tirole (1991) for an overview of signalling equilibrium, Cho and Kreps (1987) for the introduction of the D1 Criterion and Cartwright (2009) for alternatives to the D1 Criterion.

²We assume pure strategies throughout.

invest the maximum amount 1 when of type $t \in [t_l, 1]$ for some t_l . The uniform prior requirement results in inferences

$$\phi_B(b, 1) = \begin{cases} \frac{1}{1-t_l} & \text{if } b \in [t_l, 1] \\ 0 & \text{otherwise} \end{cases}.$$

This means that agent A would receive esteem H from choosing to invest 1 where

$$H := -\frac{1}{1-t_l} \int_{t_l}^1 (1-b)^2 db = -\frac{(1-t_l)^2}{3}. \quad (4)$$

Agent C 's inferences need not be consistent with A 's investment function. Agent A does, however, behave optimally relative to C 's inferences. Also, agent B 's inferences are consistent with A 's investment function and, therefore, take into account the effect that the naïve inferences of C will have on A 's investment. Thus agents A and B behave as 'rational Bayesian agents' while agent C does not.

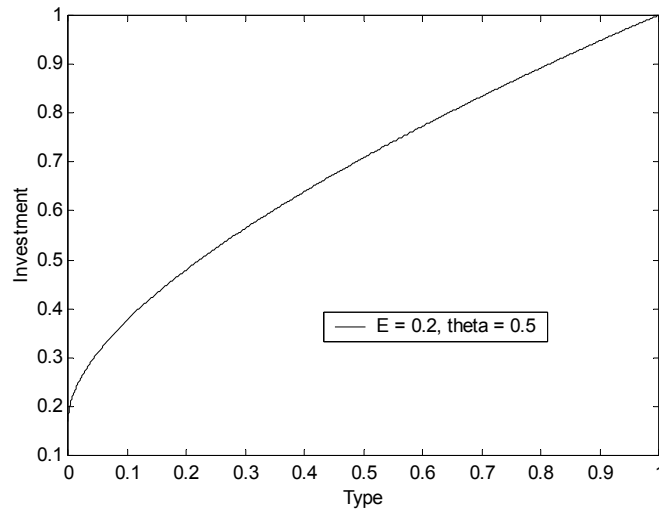
Several related solution concepts have been proposed in the literature that allow inferences of one or more agent to be inconsistent in some way including Self-Confirming Equilibrium (Fudenberg and Levine 1993), Cursed Equilibrium (Eyster and Rabin 2005), Analogy-Based Expectation Equilibrium (Jehiel 2005) and Best Response Trailing Naïve Inference (Eyster and Rabin 2008). Of these, our equilibrium is closest to Eyster and Rabin (2008). In their terminology, our agent C is inferentially naïve. The case of "face-value interpretation" in Bodner and Prelec (2003) is also related as they consider naïve inference in a self-signalling model.

3.1 Separating Equilibrium

One possible type of signalling equilibrium is a *separating equilibrium*. This equilibrium has the property that the investment of agent A reveals his level of generosity to B . That is, agent A will invest a different amount if a different type and so once agent B sees the investment of A he will know the type of A for sure. Figure 1 illustrates by plotting the equilibrium investment function when $E = 0.2$ and $\theta = 0.5$. For example agent A would

invest 0.6 if his level of generosity is 0.4 and agent B would infer that agent A has level of generosity 0.4 if he invests 0.6. Note that agent C would infer that A has level of generosity 0.6 if he A invests 0.6.

Figure 1: The investment function when $E = 0.4$ and $\theta = 0.5$.



In order to derive the equilibrium investment function we introduce an index variable v and solve for type t and investment x as a function of v . For any value of v we then find the level of investment $x(v)$ of a type $t(v)$ agent and so can derive the investment function. The following result summarizes possible separating equilibria.

Proposition 1: Let $\pi := E\theta(1 + E(1 - \theta))$. If $\pi \leq 0.25$ then there exists a unique signalling equilibrium and it is a separating equilibrium. If $\pi = 0.25$ then the investment function is described by

$$\begin{aligned} x(v) &= 1 - E\theta(4 + v)e^{-\frac{v}{2}} \\ t(v) &= 1 - \left(1 + \frac{v}{2}\right)e^{-\frac{v}{2}}. \end{aligned}$$

If $\pi < 0.25$ then the investment function is described by

$$x(v) = 1 - \frac{E\theta [(r_2 + \pi)e^{r_1v} - (r_1 + \pi)e^{r_2v}]}{\pi(r_2 - r_1)}$$

$$t(v) = 1 + \frac{r_1(r_2 + \pi)e^{r_1v} - r_2(r_1 + \pi)e^{r_2v}}{\pi(r_2 - r_1)}$$

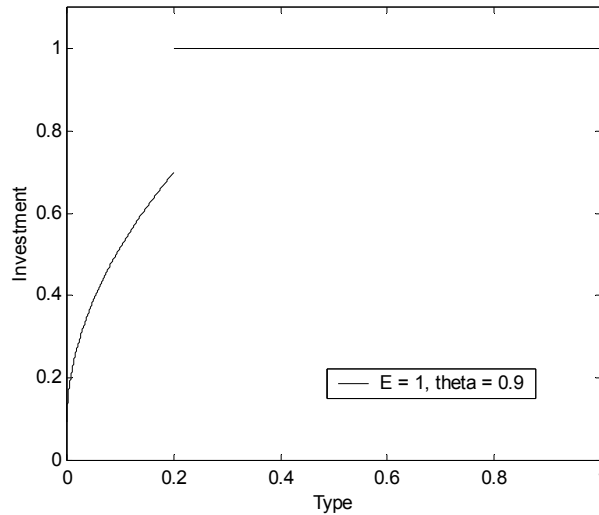
where

$$r_1 = -\frac{1 + (1 - 4\pi)^{0.5}}{2} \text{ and } r_2 = -\frac{1 - (1 - 4\pi)^{0.5}}{2}.$$

3.2 Pooling equilibrium

The second type of signalling equilibrium is a *pooling equilibrium*. In this case there exists a set of types $[t_l, 1]$ such that agent A would invest the maximum amount 1 if of type $t \in [t_l, 1]$. Note that this means agent B cannot infer the type of A if A does invest 1 (and hence the equilibrium is a pooling equilibrium). Figure 2 illustrates by plotting the equilibrium investment function when $E = 1$ and $\theta = 0.9$. We see that agent A would invest 1 if of type $t_l \simeq 0.2$ or above. Note that for t less than t_l the investment function is similar to that observed in a separating equilibrium.

Figure 2: The investment function when $E = 1$ and $\theta = 0.9$.



The value of t_l is found by solving for which type, agent A is indifferent between ‘revealing’ his type to agent B and thus receiving esteem $-(1-t_l)^2$ versus investing 1 and receiving esteem H (where H is calculated using (4)). Note that some investments should not be observed in equilibrium such as $x \in [0.75, 1)$ in this example. We still, however, have to specify agent B ’s inferences in this case. As in Bernheim (1994) we invoke the D1 Criterion meaning that agent B would infer that agent A has type t_l if his investment is between 0.75 and 1. Anyone investing 0.8, for instance, is inferred by agent B to have type 0.2. The following result summarizes possible pooling equilibria.

Proposition 2: Let $\pi := E\theta(1 + E(1 - \theta))$. If $\pi > 0.25$ then there exists a unique signalling equilibrium (that satisfies the D1 Criterion) and it is a pooling equilibrium. There exists type $t_l \in [0, 1]$ such that agent A invests 1 if of type $t \in [t_l, 1]$ and invests according to

$$\begin{aligned} x(v) &= 1 + \left(\frac{(2\pi - 1)}{m} \sin \frac{mv}{2} - \cos \frac{mv}{2} \right) \frac{E\theta e^{-\frac{1}{2}v}}{\pi} \\ t(v) &= 1 - \left(\cos \frac{mv}{2} - \frac{(2\pi - 1 - m^2)}{2m\pi} \sin \frac{mv}{2} \right) e^{-\frac{1}{2}v} \end{aligned}$$

if $t < t_l$ where $m = (4\pi - 1)^{0.5}$. If $\pi > \frac{3}{2}$ then $t_l = 0$, otherwise t_l solves

$$-(x_l - t_l)^2 - E(1 - \theta)(1 - x_l)^2 - E\theta(1 - t_l)^2 = -(1 - t_l)^2 - E\theta \frac{(1 - t_l)^2}{3} \quad (5)$$

where $x_l = x(t^{-1}(t_l))$.

3.3 Summary

Propositions 1 and 2 completely characterize the signalling equilibrium of the model. Before moving onto our main results and the contrast between naïve and rational inferences we can make some general observations. First, we see that if $E > 0$ then agent A will invest more than his level of generosity (as seen in Figures 1 and 2). Thus, a desire for esteem increases the investment

of agent A . If we view increased investment as socially optimal then it becomes apparent why social norms can be beneficial. It also becomes clear that agent C will necessarily overestimate the generosity of A and hence is naïve.

While it is not our primary interest here, we can note that a *ceteris paribus* increase in E increases the level of investment. Thus, the more agent A cares about the esteem of B and C relative to intrinsic utility then the more he invests. The effect, however, depends on the relative weight given to the esteem of the rational agent. For example, if $\theta = 1$ then there exists a pooling equilibrium when $E > 0.25$ and A will invest the maximum irrespective of his type when $E \geq 1.5$. Increasing E thus creates an incentive for agent A to invest the maximum. This is primarily because a less generous agent A would need to invest the maximum to ‘hide’ his lack of generosity from agent B . By contrast, if $\theta = 0$ there must be a separating equilibrium irrespective of how large is E . Thus, agent A would increase his investment but not up to the maximum. This is because the less generous agent A can hide his lack of generosity from agent C by merely investing more.

4 The consequences of naïve inferences

In order to see the consequences of naïve inferences we shall, in the following, fix a value of E , the relative weight given to esteem, and vary θ , the relative weight given to the esteem of the rational agent. The following is a trivial but useful observation that outlines what potentially could happen following a change in θ .

Remark 1: A ceteris paribus decrease in θ must result in a type t agent either:

- (1) increasing his investment from something less than 1 to 1,
- (2) decreasing his investment from 1 to something less than 1,
- (3) increasing his investment while still investing less than 1,
- (4) decreasing his investment while initially investing less than 1,
- (5) keeping his investment unchanged.

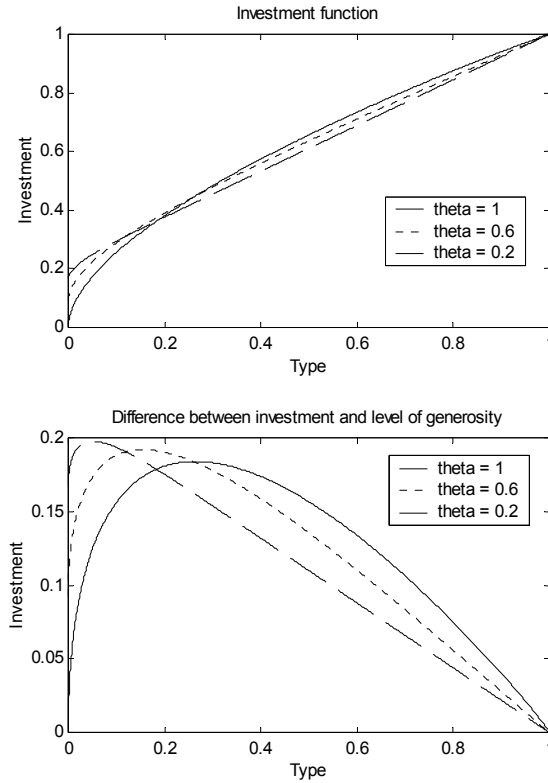
Our prior expectation when we began to work with the model was that some of the 5 possibilities would simply not occur. In particular, we thought that (1) and (4) seemed counter-intuitive. We shall demonstrate, however, that all 5 possibilities can occur (even in this relatively simple model). Thus, a change in θ could potentially increase, decrease, raise to the maximum or reduce from the maximum the investment of agent A . In some respects this may seem a negative result, and it certainly suggests that general results about the consequences of naïve inferences will be difficult to obtain. In explaining why anything can happen, we hope, however, to provide a better understanding of the consequences of naïve inferences. We shall do this in three parts: In Section 4.1 we consider an agent A who invests less than 1 both before and after any change in θ (and thus focus on cases (3) and (4)); in Section 4.2 we consider which type of agent A would invest 1 (and thus focus on cases (1) and (2)). In Section 4.3 we shall draw the discussion together.

4.1 Naïve inferences and investments below the maximum

Our focus in this section will be on those settings where agent A invests less than 1. We have already seen that if agent A invests less than 1 then the rational agent B will infer his type. This means that we are considering settings where *agent A will always get the same level of esteem from agent B*. By contrast, the esteem that agent A receives from agent C is clearly increasing in the investment that agent A makes. The investment of agent A will thus be such as to signal his type to agent B , and ‘fool’ the naïve agent C into inferring that he is more generous than he actually is. This could be seen to create an asymmetry whereby agent A will invest more than his type so that agent B does not give him *less* esteem than ‘he deserves’ and agent C gives him *more* esteem than ‘he deserves’.

Figure 3 illustrates the consequences of this by plotting the investment function μ when $E = 0.25$ for three different values of θ . To see the effects more clearly it is useful to also plot $\mu(t) - t$, the gap between investment and level of generosity.

Figure 3: The investment function when $E = 0.25$.



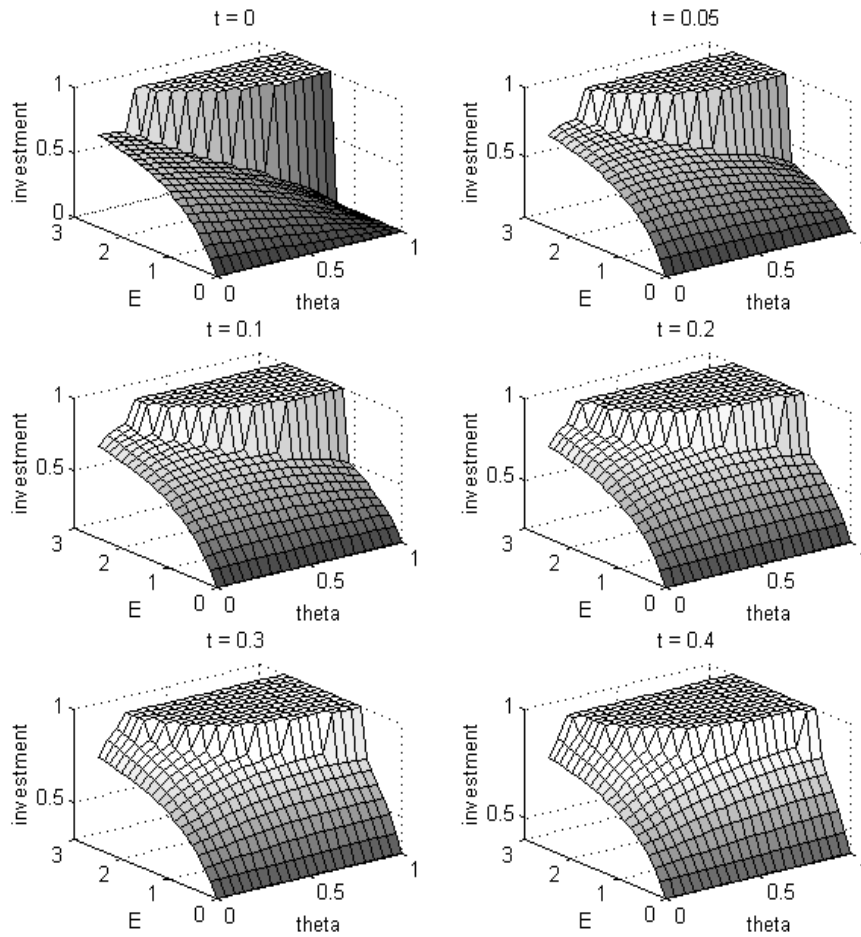
We see that as θ increases agent A relatively *decreases* his investment if *less generous* and relatively *increases* his investment if *more generous*. If agent A has an *intermediate level of generosity*, such as $t = 0.2$, then he invests relatively more for *intermediate values* of θ . This proves to generally be the case:

Proposition 3: Ignoring combinations of E and t such that agent A would invest 1, the investment of agent A is a, ceteris paribus,

- (i) decreasing function of θ for t sufficiently small,*
- (ii) increasing function of θ for t sufficiently large,*
- (iii) inverse ‘U shaped’ function of θ for intermediate values of t .*

This Proposition was obtained numerically by calculating the investment of agent A for a incremental values of θ, t and E . Figure 4 illustrates by plotting the investment of agent A for some values of t, E and θ . This figure also serves to clarify that sufficiently small means t very close to 0 while sufficiently large means any t above 0.4.

Figure 4: The investment of agent A for different values of t as a function of θ and E .



To understand the comparative statics given in Proposition 3 two observations are crucial: First, because of the *concavity of the esteem function*: the *smaller* is the investment of agent *A* the more sensitive is the esteem he receives from *agent C* to any change in his investment.³ Similarly, because of the *concavity of the investment function*: the *larger* is the investment of agent *A* the more sensitive is the esteem received from *agent B* to any change in investment.

Keeping these two observations in mind suppose that agent *A* has a level of generosity *near to 0* and so is investing relatively little. Small changes to his investment will not significantly alter the esteem he receives from the rational agent. By contrast, small changes to his investment could significantly increase the esteem that he receives from the naïve agent. Thus, the higher is the weight on the esteem of the naïve agent the more incentive for agent *A* to increase his investment.⁴ Essentially, the rational agent will give him low esteem whatever while the naïve agent is easily 'fooled' into giving more esteem, so the higher is the weight on the esteem of the naïve agent the more incentive to investment.

Now, suppose that agent *A* has a generosity *near to 1* and so is investing close to 1. Small changes to his investment will not significantly alter the esteem that he receives from the naïve agent. They could, however, significantly alter the esteem he receives from the rational agent. Thus, the higher the weighting on the esteem of the rational agent the more incentive for agent *A* to increase his investment. Essentially, the naïve agent will give him high esteem whatever while the rational agent *B* may infer he is less generous than he is, so the higher is the weight on the rational agent the more incentive to invest. This argument is, however, slightly incomplete; if the less generous agent *A* will increase his investment, as θ decreases, then

³Recall, that if agent *A* invests x then the naïve agent gives him esteem $-(1-x)^2$ and so the change in esteem from an increase in investment is $2(1-x)$.

⁴To be a bit more precise, it is simple to show (see the Appendix) that agent *A* will invest

$$\frac{E(1-\theta)}{1+E(1-\theta)}$$

if of type $t = 0$. Clearly, the more weight is put on the esteem of the naïve agent then the more agent *A* would invest.

might not the more generous agent A have to also increase his investment in order to differentiate himself and signal his generosity to agent B ?⁵ We see, in Proposition 3, that this turns out to not be the case, for the most generous type of agent A . An explanation for this is that, irrespective of the investment of a less generous agent A , there is a ‘fixed point’ where agent A invests 1 if type 1. This means that by investing arbitrarily close to 1 the generous agent A can signal his generosity.⁶

Finally, suppose that agent A has an *intermediate level of generosity*. A simple explanation for why investment is inverse U shaped would be that when θ is, say, 0.5 agent A values the esteem of *both* agent B and C and so has a ‘complementary incentives’ to invest. For a slightly more convincing explanation we need to return to the idea of a ‘ratchet effect’ briefly mentioned in the previous paragraph: Given that a less generous agent A would increase his investment as θ falls there is a need for the slightly more generous agent A to also increase his investment as θ falls to signal his level of generosity to agent B . That is, he needs to invest more to distinguish himself from the less generous agent A , who will now be investing more.⁷ But, we do need to ask whether agent A is all that concerned about signalling his type to agent B if θ has fallen. It is this that gives the inverse U shape. As θ falls but is still relatively large, agent A does value the esteem of agent B and so would increase his investment to counteract the increased investment of the less generous type. But as θ falls towards zero there becomes less incentive to signal type to agent B and so agent A decreases his investment towards his level of generosity (despite the less generous agent A still continuing to increase his investment).

⁵This, incidentally, is why we initially considered Property (4) of Remark 1 counter-intuitive.

⁶If agent A has level of generosity $1 - \varepsilon$, for some small ε , and as is not investing 1 (which is the focus of this Section) then we must have a separating equilibrium. This means that agent A invests 1 if and only if his level of generosity is 1. So, the higher is θ the closer will agent A invest to 1.

⁷Note that for intermediate levels of generosity the ‘fixed point’ of ‘agent A invests 1 if and only if his level of generosity is 1’ no longer has much effect.

4.2 Naïve inferences and investing the maximum

We now turn to question what types of agent A will invest the maximum. Our focus, therefore, is on t_l . The esteem that agent A receives from agent C changes continuously with investment. This, and the nature of the esteem function, means that there is never any incentive for agent A to invest 1 rather than something slightly less than 1 (unless his level of generosity is 1). By contrast, there can be a relatively large discontinuity in the esteem that agent B will give to agent A if he invests 1 rather than something arbitrarily close to 1. This would suggest that there is a greater incentive for agent A to invest 1 the higher is θ . More precisely, it suggests that the value of t_l should decrease the higher is θ . This is often the case, but it need not be:

Proposition 4: The minimum type t_l who invests 1 is a, ceteris paribus,
(i) decreasing function of θ for E sufficiently small or large;
(ii) a ‘U shaped’ function of θ for intermediate values of E .

This Proposition was also obtained by numerically solving for t_l for incremental values of θ and E . Figure 5 illustrates.

Figure 5: The value of t_l as a function of E and θ .

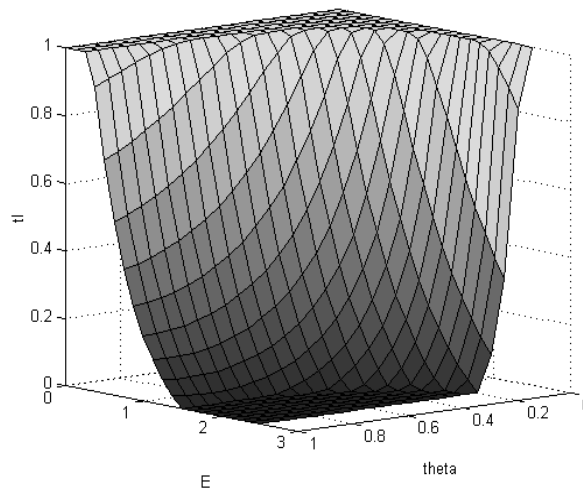
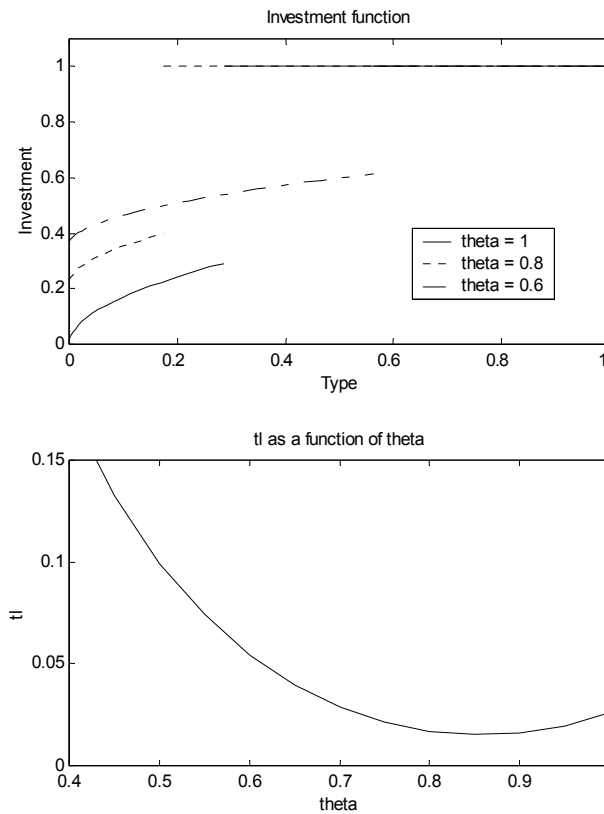


Figure 5 illustrates very clearly that t_l is typically decreasing in θ . Part (2) of Proposition 4 does, however, imply that t_l can decrease as θ decreases for some combinations of E and θ . Those with good eyesight may just be able to pick this out on Figure 5 for values of E between approximately 1.2 and 1.5. To provide a slightly more convincing illustration consider Figure 6 which plots a sample of investment functions and the value of t_l when $E = 1.4$. We calculate, for instance, that $t_l = 0.0287$ when $\theta = 1$ while $t_l = 0.0174$ when $\theta = 0.8$. Thus, t_l has decreased as θ has decreased.

Figure 6: Investment functions when $E = 1.4$.



It is not a surprise, given the intuition outlined above, that t_l is typically a decreasing function of θ . More interesting is to try and understand why

a decrease in θ can cause a decrease in t_l . A crucial point to recognize is that we observe this positive relationship precisely *when t_l is in the range of t for which we observe an inverse U shaped or negative relation between investment and θ* . That is, we observe this relationship when it is *only the very least generous agent A who will not invest 1*.

Given Proposition 3, we know that a *less generous* agent A will be inclined to *increase* his investment following a fall in θ (particularly if θ is initially high). What we need to understand is why he would want to increase his investment all the way to 1. Recall that the *benefit* of investing 1 is to receive *extra esteem, primarily from the rational agent B* . The *cost* of investing 1 is the *loss in intrinsic utility* from investing more than level of generosity. A decrease in θ , therefore, has two conflicting effects on the incentives of the less generous agent to invest 1: (a) On the *positive side*, the increase in investment that agent A would want to make makes it *less costly*, in terms of intrinsic utility, to ‘go all the way’ and invest 1. Basically, the increase in investment that agent A would make to receive esteem from agent C makes it cheaper for him to increase his investment to 1 and receive extra esteem from agent B .⁸ (b) On the *negative side*, a fall in θ clearly diminishes the value of the extra esteem that would be received from agent B .

With these preliminaries, suppose that t_l is small and θ is high. Agent A will invest 1 if of type t_l and less than 1 if of type $t_l - \varepsilon$, for some small ε . Clearly, continuity implies that it would require relatively little change in incentives for a type $t_l - \varepsilon$ agent A to want to invest 1. If, therefore, we reduce θ and effect (a), discussed above, outweighs effect (b), the type $t_l - \varepsilon$ agent would invest 1, and t_l will have decreased. Proposition 4 and Figure 6 show that this can be the case.⁹ This illustrates how agents B and C can provide complimentary incentives for agent A to invest. Specifically, if θ is high then the less generous agent A will not invest much, while if θ

⁸This is clearly apparent in the top panel of Figure 6.

⁹Note that if t_l is larger then effect (a) will also be negative (or at least less important) and so a reduction in θ will lessen the incentive for a type t_l or $t_l - \varepsilon$ agent to invest 1. This reinforces the intuition with which we began this Section in explaining why a reduction in θ will typically increase t_l .

is low the less generous agent A would invest more but still less than 1. For intermediate values of θ , however, the less generous agent A may invest 1.

4.3 Summary and discussion

We began this Section with Remark 1 setting out the five potential reactions to a decrease in θ . A hope and expectation when we started to work with the model was that some of these five possibilities could be ruled out. For example, we intuitively expected that an increase in θ would decrease the set of types for which agent A would invest 1. The reality, as we have now shown is a lot more subtle than this (even for the relatively simple model we have used). A decrease in θ may result in agent A investing less, more, the maximum or something less than the maximum. Only in the case where he is investing the maximum amount both before and after, does a decrease in θ have no effect. The consequences of naïve beliefs are thus complex and not easily characterized. We have, however, learnt something about the consequences of naïve inferences.

One thing we do see (in effects (2) and (4)) is that an expectation that others will make *rational inferences can provide strong incentives to invest more in the public good*. This will be the case for the *generous agent*. When inferences are rational the generous agent needs to invest more than his level of generosity in order to signal his generosity and distinguish himself from those trying to appear generous. Naïvety in inferences provides less incentive for the generous agent to invest in the public good because he will be inferred as generous merely by investing at his level of generosity. A simple corollary of Propositions 3 and 4 is that a sufficiently generous agent will invest more the higher is the weight on the inferences of the rational agent. An interesting question is whether this holds more generally than in the model considered in this paper. A conjecture is that the generous agent always invests more when $\theta = 1$ than when $\theta = 0$. We would, however, be surprised if one always obtains a monotonic relation between θ and investment, rather than a U or inverse U shaped relation.¹⁰

¹⁰This is because of the ratchet effect discussed in Section 4.1. We know that a less

We also see (in effects (1) and (3)), that an expectation that others will make *naïve inferences can provide strong incentives to invest more in the public good*. This will be the case for the *less generous agent*. When inferences are naïve the less generous agent will invest more than his level of generosity in order to ‘fool’ a naïve agent in to thinking he is generous. Rationality in inferences provides less incentive for the less generous agent to invest because it is simply too costly for him to try and appear generous when the observer knows that he is trying to appear generous. It is a very simple matter to prove, very generally, that the investment of the less generous agent will increase the more weight there is on the inferences of the naïve agent *provided* the agent is investing less than 1. The discussion of the last Section shows that we need the proviso that investment be less than 1 because the less generous agent *A*, who expects a combination of both naïvety and some rationality in inferences, may invest the maximum amount.¹¹ This means that there is no simple monotonic relation between θ and the investment of a less generous agent.

One clear consequence of the above is that how inferences are drawn will likely change the *gap in investment between the less and more generous*. In what direction will depend on the form of equilibrium. In the case of separating equilibria we would expect *greater inequality in investment* when inferences are *rational rather than naïve*. This is because the less generous agent will invest relatively little and the more generous agent will invest relatively more.¹² We know, however, see Proposition 2, that rational inferences are more likely to lead to a pooling equilibrium. One could argue that there is relatively *less inequality* in a *pooling equilibrium* given that agent *A* will invest the same amount for many types. To summarize, a decrease in θ

generous agent *A* will increase his investment as θ decreases. If this leads to a ‘sufficiently strong ratchet effect’ in which the generous agent has to invest more to distinguish himself from the less generous agent the consequences for investment are unclear.

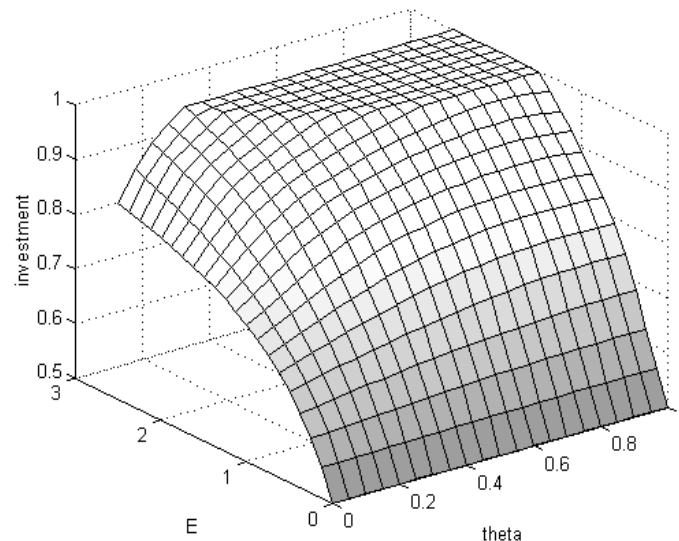
¹¹A related and relatively unexplored issue is whether a limit on contributions increases total contributions. The decrease in contributions that would result from constraining the most generous being offset by the increased investment of the less generous. Harbaugh (1998a, 1998b) does discuss how category reporting can increase contributions.

¹²Switching income level for level of generosity there is experimental evidence that ‘lower’ types do contribute more (Buckley and Croson 2006).

will likely lead to less inequality if there is initially a separating equilibrium, but more inequality if there is initially a pooling equilibrium.

A final question we could ask is whether the *expected investment* of agent A is increasing or decreasing the more weight on naïve inferences. If agent A is going to invest relatively more if less generous and relatively less if more generous the overall effect is ambiguous. We can, however, find the expected investment, assuming a uniform distribution over types, and we plot the results in Figure 7.

Figure 7: The expected investment of agent A as a function of E and θ .



Clearly we see that investment is higher the more weight is put on inferences of the rational agent. The simple reason for this is that the ‘ t sufficiently large’ in Proposition 3 is quite small (around 0.4), and so agent A will likely increase his investment the higher is θ . For this reason we would not draw too strong a conclusion about how such a relationship would generalize. In a setting, for example, where agent A has a higher probability of being a less generous type it is clear that we could obtain an increase in expected investment with a lower θ . Figure 7 does, however, reiterate the, perhaps

surprisingly, strong way in which rational inferences can significantly increase investment. Most notably, and very generally, it is only when θ is sufficiently high that we would ever obtain an expected investment of 1.

5 Conclusion

The motivation for this paper was that naïvety in inferences, or an expectation of naïvety in inferences, may alter the incentives of agents who seek social approval or disapproval. We hope to have demonstrated that naïvety in inferences can indeed have significant and sometimes difficult to predict consequences for behavior. We have framed the analysis in terms of investment in a public good. This is partly because we feel this is a setting where some naïvety in inferences is to be expected. It is also because we feel that acknowledging naïvety in inferences may lead to differing interpretations or expectations of how collective action problems can be resolved, or not. The model could, however, be applied in different contexts.

The model we have used is clearly stylized. Our intention was not to provide a realistic model of inference formation and action in collective action problems but merely to highlight the potential importance of considering naïve inferences. It would, however, be interesting to relax some of the modelling assumptions and consider more generally the consequences of naïve inferences on behavior. For example, we consider the two extremes of the agent who has naïve inferences and the agent who has rational inferences. In reality we may expect various mixtures of the two. Experimental evidence can hope to shed light on how people do in fact make inferences of others' actions. With this in hand we may hope to provide general answers to questions such as: Does an expectation of naïve inferences increase or decrease investment in public goods? Does a maximum limit on investment increase investment in a public good? How do inferences effect the distribution of who invests in the public good? Should a policy maker who is attempting to elicit contributions to a public good or find out attitudes towards a public good take into account, or try to influence, how people infer the actions or opinions of others?

6 Proofs of Propositions 1 and 2

If agent B perceives agent A to be of type b then agent A 's utility is $u(t, x, b) = -(x - t)^2 - E\theta(1 - b)^2 - E(1 - \theta)(1 - x)^2$. Indifference curves in the (x, b) plane for an agent of type t are thus given by, $(x - t)^2 + E\theta(1 - b)^2 + E(1 - \theta)(1 - x)^2 = D$, where D is an arbitrary constant. We can calculate the slope of an indifference curve of a type t agent through the point (b, x) as,

$$\frac{db}{dx} = -\frac{\partial u/\partial x}{\partial u/\partial b} = \frac{(1 + E(1 - \theta))x - t - E(1 - \theta)}{E\theta(1 - b)}.$$

In equilibrium there must (i) be a tangency between inference function ϕ_B and the indifference curve and (ii) the inferences of agent B must be correct implying that $\phi_B(x) = b = t$. Thus,

$$\phi_B'(x) = \frac{(1 + E(1 - \theta))x - \phi_B(x) - E(1 - \theta)}{E\theta(1 - \phi_B(x))}. \quad (6)$$

The differential equation (6) can be rewritten as system

$$\begin{bmatrix} dt/dv \\ dx/dv \end{bmatrix} = \begin{bmatrix} -1 & 1 + E(1 - \theta) \\ -E\theta & 0 \end{bmatrix} \begin{bmatrix} t - 1 \\ x - 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} t - 1 \\ x - 1 \end{bmatrix},$$

where v is some index. Rearranging the bottom equation gives

$$t = 1 - \frac{x'}{E\theta} \quad (7)$$

which can be inserted into the top equation to give the second-order differential equation

$$x'' + x' + E\theta(1 + E(1 - \theta))x = E\theta(1 + E(1 - \theta)). \quad (8)$$

The solution to this differential equation is easily found and so values of x and t can be traced out to show the investment x of a type t agent. From this can be derived appropriate inferences ϕ_B and investment function μ .

Note, however, that we cannot know at this stage whether ϕ_B and μ are consistent with equilibrium as, in particular, we may obtain an investment $x > 1$ which is not possible (see below).

The characteristic equation of (8) is $r^2 + r + E\theta(1 + E(1 - \theta))$. This equation has two distinct real roots if $E\theta(1 + E(1 - \theta)) < \frac{1}{4}$, repeated roots if $E\theta(1 + E(1 - \theta)) = \frac{1}{4}$ and two distinct complex roots if $E\theta(1 + E(1 - \theta)) > \frac{1}{4}$. All are clearly possible and so we need to distinguish these three cases.

Case (1): $E\theta(1 + E(1 - \theta)) = \frac{1}{4}$. The solution to equation (8) is

$$x = 1 + C_1 e^{-\frac{v}{2}} + C_2 v e^{-\frac{v}{2}} \quad (9)$$

for some constants C_1 and C_2 . To derive appropriate initial conditions consider an agent of type $t = 0$. If a type 0 agent is correctly perceived to be of type 0 then his payoff is

$$U(x, 0, 0) = -x^2 - E\theta - E(1 - \theta)(1 - x)^2.$$

Setting $\frac{du}{dx} = 0$ suggests that $t = 0$ as $x = \frac{E(1-\theta)}{1+E(1-\theta)}$. Appropriate initial conditions are thus $t = 0$ and $x = \frac{E(1-\theta)}{1+E(1-\theta)}$ as $v = 0$. Using equations (9) and (7) in turn gives

$$C_1 = \frac{-1}{1 + E(1 - \theta)}; \quad C_2 = E\theta - \frac{1}{2(1 + E(1 - \theta))}. \quad (10)$$

Using (7), (9), (10) and that $E\theta(1 + E(1 - \theta)) = 0.25$ one can prove the first part of Proposition 1.

Case (2): If $E\theta(1 + E(1 - \theta)) < 0.25$. The solution to equation (8) is

$$x = 1 + C_3 e^{r_1 v} + C_4 e^{r_2 v} \quad (11)$$

where

$$r_1 = -\frac{1 + (1 - 4E\theta(1 + E(1 - \theta)))^{0.5}}{2}; \quad r_2 = -\frac{1 - (1 - 4E\theta(1 + E(1 - \theta)))^{0.5}}{2}$$

and C_1 and C_2 are constants. Appropriate initial conditions remain $t = 0$ and $x = \frac{E(1-\theta)}{1+E(1-\theta)}$ as $v = 0$. So, from (11) we obtain $C_4 = \frac{-1}{1+E(1-\theta)} - C_3$ and using (7) we get

$$E\theta = r_1 C_3 + r_2 C_4 = (r_1 - r_2) C_3 - \frac{r_2}{1 + E(1 - \theta)}.$$

Thus,

$$C_3 = -\frac{r_2 + E\theta(1 + E(1 - \theta))}{(r_2 - r_1)(1 + E(1 - \theta))} \text{ and } C_4 = \frac{r_1 + E\theta(1 + E(1 - \theta))}{(r_2 - r_1)(1 + E(1 - \theta))} - C_3.$$

From this we obtain the second part of Proposition 1. To complete the proof of Proposition 1 we note that if $E\theta(1 + E(1 - \theta)) \leq 0.25$ then $x, t \in [0, 1]$ for all $v > 0$ and $\lim_{v \rightarrow \infty} t, x = 1$. Hence a separating equilibrium does exist.

Case (3): $E\theta(1 + E(1 - \theta)) > 0.25$. The characteristic equation of (8) has two distinct complex roots and so the solution of (8) has form,

$$x = 1 + e^{-\frac{1}{2}v} \left(C_5 \cos \frac{mv}{2} + C_6 \sin \frac{mv}{2} \right) \quad (12)$$

where $m = (4E\theta(1 + E(1 - \theta)) - 1)^{0.5}$ and C_5 and C_6 are constants. Appropriate initial conditions remain $t = 0$ and $x = \frac{E(1-\theta)}{1+E(1-\theta)}$ as $v = 0$. Using equations (12) and (7) in turn gives

$$C_5 = \frac{-1}{1 + E(1 - \theta)}; \quad C_6 = \frac{2E\theta + C_5}{m}.$$

This describes an investment function but results in an $x > 1$ which is not possible. We can, however, use condition (5) to find a type such that agent 1 is indifferent between investing 1 and investing as in a separating equilibrium. Note that agent 1 would want to invest 1 even if of type $t = 0$ when

$$-1 - \frac{E\theta}{3} > -\left(\frac{E(1-\theta)}{1+E(1-\theta)} \right)^2 - E\theta - E(1-\theta) \left(1 - \frac{E(1-\theta)}{1+E(1-\theta)} \right)^2.$$

From this we find that $t_l = 0$ when $2E\theta/3 > 1/(1 + E(1 - \theta))$. This com-

pletes the proof of Proposition 2. ■

References

- [1] Andreoni, J. (1990) "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal* 100: 464-477.
- [2] Andreoni, J. and Bernheim, D. B. (2008) "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," working paper.
- [3] Azar, O. H. (2004) "What Sustains Social Norms and How they Evolve? The Case of Tipping," *Journal of Economic Behavior and Organization* 54: 49-64.
- [4] Benabou, R. and Tirole, J. (2006) "Incentives and Prosocial Behavior," *American Economic Review* 96(5): 1652-1678.
- [5] Bernheim, D. B. (1994) "A Theory of Conformity," *Journal of Political Economy* 102(5): 841-877.
- [6] Bernheim, D. B. and Serverinov, S. (2003) "Bequests as Signals: An Explanation for the Puzzle of Equal Division," *Journal of Political Economy* 111(4): 733-764.
- [7] Bodner, R. and Prelec, D. (2003) "Self-Signaling and Diagnostic Utility in Everyday Decision Making," in *The Psychology of Economic Decisions*, Vol 1: Rationality and Well-Being. Brocas, I and Carrillo, J. D. (eds), Oxford University Press, Oxford.
- [8] Buckley, E. and R. Croson (2006) "Income and Wealth Heterogeneity in the Voluntary Provision of Linear Public Goods," *Journal of Public Economics* 90: 935-955.
- [9] Cartwright, E. (2009) "Conformity and Out of Equilibrium Beliefs," *Journal of Economic Behavior and Organization*, to appear.

- [10] Charness, G. and Levin, D. (2005) "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity and Affect," *American Economic Review* 95(4): 1300-1309.
- [11] Cho, I. K. and Kreps, D. M. (1987) "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics* 102: 179-221.
- [12] Coleman, J. (1990) *Foundations of Social Theory*. Harvard University Press, Cambridge, MA.
- [13] Esponda, I. (2008) "Behavioral Equilibrium in Economies with Adverse Selection," *American Economic Review*, 98(4): 1269-1291.
- [14] Ettinger, D. and Jehiel, P. (2008) "A Theory of Deception," working paper.
- [15] Eyster, E. and Rabin, M. (2005) "Cursed Equilibrium," *Econometrica* 73(5): 1623-1672.
- [16] Eyster, E. and Rabin, M. (2008) "Naïve Herding," working paper.
- [17] Fehr, E. and S. Gächter (1999) "Collective Action as a Social Exchange," *Journal of Economic Behavior and Organization* 39: 341-369.
- [18] Francis, H. (1985) "The Law, Oral Tradition and the Mining Community" *Journal of Law and Society* 12: 2267-2271.
- [19] Frank, R. (1988) *Passions within Reason: The Strategic Role of the Emotions*. W.W. Norton.
- [20] Fudenberg, D. and Tirole, J. (1991) *Game Theory*. MIT Press, Cambridge, MA.
- [21] Fudenberg, D. and Levine, D. K. (1993) "Self-Confirming Equilibrium," *Econometrica* 61(3): 523-545.
- [22] Glazer, A. and Konrad, K. A. (1996). "A Signaling Explanation for Charity," *American Economic Review* 86(4): 1019-1028.

- [23] Grossman, Z. (2008) "Social-Image vs Self-Image in Motivating Giving," working paper.
- [24] Harbaugh, W. T. (1998a) "The Prestige Motive for Making Charitable Transfers," *American Economic Review* 88(2): 277-282.
- [25] Harbaugh, W. T. (1998b) "What Do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow," *Journal of Public Economics* 67: 269-284.
- [26] Hardy, C. L. and van Vugt, M. (2006) "Nice Guys Finish First: The Competitive Altruism Hypothesis," *Personality and Social Psychology Bulletin* 32(10): 1402-1413.
- [27] Hollander, H. (1990) "A Social Exchange Approach to Voluntary Contribution," *American Economic Review* 80: 1157-1167.
- [28] Jehiel, P. (2005) "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory* 123: 81-104.
- [29] Jehiel, P. and Koessler, F. (2008) "Revisiting Games of Incomplete Information with Analogy-Based Expectations," *Games and Economic Behavior* 62: 533-557.
- [30] Jones, E. E. and Harris, V. A. (1967) "The Attribution of Attitudes," *Journal of Experimental Social Psychology* 3: 1-24.
- [31] Kandel, E. and E. Lazear (1992) "Peer Pressure and Partnerships," *Journal of Political Economy* 100: 801-812.
- [32] Knack, S. (1992) "Civic Norms, Social Sanctions and Voter Turnout," *Rationality and Society* 4: 133-156.
- [33] Knack, S. and Keefer, P. (1997) "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation," *Quarterly Journal of Economics* 112: 1251-1288.

- [34] Lindbeck, A., S. Nyberg and J. Weibull (1999) "Social Norms and Economic Incentives in the Welfare State," *The Quarterly Journal of Economics* 114: 1-35.
- [35] Mullainathan, S., Schwartzstein, J. and Shleifer, A. (2008) "Coarse Thinking and Persuasion," *Quarterly Journal of Economics* 123(2): 577-619.
- [36] Nyborg, K. and Rege, M. (2003) "On Social Norms: The Evolution of Considerate Smoking Behavior," *Journal of Economic Behavior and Organization* 52: 323-340.
- [37] Olson, M. (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge MA.
- [38] Rabin, M. and Schrag, J. L. (1999) "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics* 114(1): 37-82.
- [39] Rege, M. (2004) "Social Norms and Private Provision of Public Goods," *Journal of Public Economic Theory* 6(1): 65-77.
- [40] Rege, M. and K. Telle (2002) "An Experimental Investigation of Social Norms," *Journal of Public Economics* 88: 1625-1644.
- [41] Tadelis, S. (2008) "The Power of Shame and the Rationality of Trust" working paper.
- [42] Thaler, R. H. (1988) "Anomalies: The Winner's Curse," *The Journal of Economic Perspectives* 2: 191-202.