Public Goods, Social Norms and Naive 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

May 24, 2008

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 naive 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 naive inferences.

Keywords: signalling, naive beliefs, public goods.
JEL: D8, H41.

*We would like to thank, without implicating, Alan Carruth, Jagjit Chadha and Mathan Satchi for useful discussions. We have also benefited from the comments of seminar participants at the Economics Internal Seminar at Kent. Usual disclaimers apply.

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, a worker may fear stigma if he claims unemployment benefit (Lindbeck, Nyberg and Weibull 1999), a smoker may expect disapproval if he smokes at the dinner table (Nyborg and Rege 2003), a customer may hope to appear generous if he leaves a big tip (Azar 2004), a worker may fear contempt from colleagues if he works through a strike (Francis 1985) or a 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 Gachter 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 a signalling game (Bernheim 1994, Glazer and Konrad 1996 and Bernheim and Serverinov 2003). 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 example, B may give esteem to a generous person but show disapproval of a lazy person. Note that agent type - whether A actually is generous or lazy - is not observed by B. The distinction between type and action means that B has to infer A's type from his action. It also means that A has to try and predict how B will infer his actions. For example, if A contributes to a public good then B may infer that A is not generous but does want to try and appear generous. Consequently, if A really is generous then he may want to increase his contribution to a sufficiently high level that B will infer that he must be genuinely generous.

The modelling of social exchange as a signalling game provides a rich framework within which to work. To the best of our knowledge, however, the previous literature has focussed 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 'naive 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 'naive' B may take things at 'face value' and infer a positive contribution as being a signal of generosity.

What we call 'naive' inferences mean that a person can make systematic mistakes in interpreting the motivations or types of others. The naive 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 not to 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).

Numerous authors have relaxed the concept of Perfect Bayesian Equilibrium to an equilibrium in which actions are optimal given inferences but inferences may be biased in some way relative to actions (e.g. Fudenberg and Levine 1993, Eyster and Rabin 2005, Eliaz and Speigler 2006 and Yildiz 2007). That is the approach taken in this paper. Specifically, 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 'naive' 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 naive person C then we depart from the standard setting in the sense that C will make naive inferences and furthermore A will base his action on the basis that C will make naive inferences.

Our interest is how the naivety of C's inferences will influence the behavior of A. More generally, this is a question of how naivety in inferences, or an expectation of naivety 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 naivety in how inferences are made can have important and sometimes perverse 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 naive agent C into thinking that he is generous through a high investment. This would suggest that there are incentives for A to invest more the more weight is put on the esteem of the naive agent C. We demonstrate, however, that this need not be the case. That is, the investment of A may decrease the more weight is put on the esteem of the naive agent. The intuition for this result is that in order to earn esteem from the rational person B it is necessary for A to invest more to signal his true generosity. The rational inferences or 'cynicism' of B can therefore lead A to increase his investment in order to get the esteem from B 'that his true generosity deserves'.

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 if 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 that A would need in order to choose to invest the maximum. Intuitively, the more weight A puts on the esteem of the rational person B then the lower we might expect this minimum to be. Typically this is the case. We demonstrate, however, that it need not always be the case when we introduce a naive agent. The intuition for this result is that a less generous person A will increase his investment, the more weight he puts on the esteem of the naive person C, in order to 'fool' C into thinking he is more generous than he is. But, having increased his investment, person A 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.

We proceed as follows: Section 2 sets out the model, Section 3 details possible equilibria and Section 4 discusses the consequences of naive 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 he chooses how much to *invest in a public project* from the set $X := [0, 1]$. If he is of type t and chooses to invest x then he receives *intrinsic utility* $-(t-x)^2$. Note that intrinsic utility does not depend on others but merely on the type of the agent. The closer is investment to the level of generosity then the higher is intrinsic utility. 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,

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

for some real numbers $\lambda, \theta \geq 0$. We think of $E = \lambda + \theta$ as the overall weight of esteem, relative to intrinsic utility, while λ and θ are the relative weights given to the esteem of agents B and C.

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 *naive inference function* in the sense that,

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

If therefore, agent A chooses investment x agent C 'naively' assumes that A's level of generosity is x. Agent C thus fails to take into account the fact that A may sacrifice intrinsic utility in search of esteem. We shall discuss why agent C is naive shortly. Given that agent C has naive inferences the utility function of A can be re-written as,

$$U(t,x,\phi_B,\phi_C) = -(x-t)^2 - \theta(1-x)^2 - \lambda \int_T (1-b)^2 \phi_B(b,x)\, db. \tag{2}$$

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.

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 Bayes rational given actions. Clearly, we shall only require this of agents A and B and not C. Equilibrium

[1] See Fudenberg and Tirole (1991) for an overview of signalling equilibrium and Cho and Kreps (1987) for the introduction of the D1 Criterion.

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 choose action 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,

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

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 naive 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 equilibrium concepts have been proposed in the literature that allow inferences of one or more agent to be inconsistent in some way including Cursed Equilibrium (Eyster and Rabin 2005), Self-Confirming Equilibrium (Fudenberg and Levine 1993), equilibria with dynamic inconsistency (Eliaz and Speigler 2006) or 'wishful thinking' (Yildiz 2007).

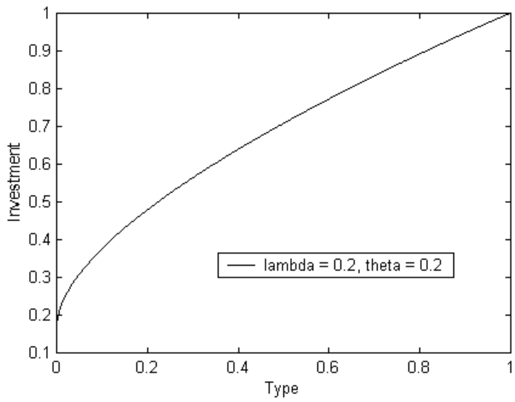
3 Types of equilibria

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

[2] We assume pure strategies throughout.

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 investment function when $\lambda = 0.2$ and $\theta = 0.2$. For example agent A would invest 0.6 if his level of generosity is 0.4 and agent B would infer that agent 1 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. We shall discuss how to interpret Figure 1 more in the following.

Figure 1: The investment function when $\lambda = 0.2$ and $\theta = 0.2$



To solve for the 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[3].

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

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

If $\lambda(1+\theta) < 0.25$ *then the investment function is described by*

$$\begin{aligned} x(v) &= 1 - \frac{(r_2 + \lambda(1+\theta))e^{r_1 v} - (r_1 + \lambda(1+\theta))e^{r_2 v}}{(r_2 - r_1)\,(1+\theta)} \\ t(v) &= 1 + \frac{r_1\,(r_2 + \lambda(1+\theta))e^{r_1 v} - r_2\,(r_1 + \lambda(1+\theta))e^{r_2 v}}{\lambda(r_2 - r_1)\,(1+\theta)} \end{aligned}$$

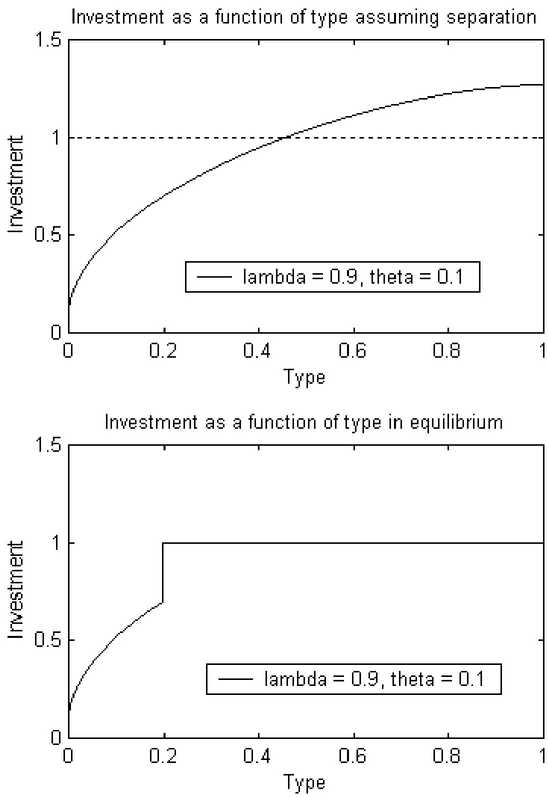
[3] The proof of Propositions 1 and 2 are found in the Appendix.

where

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

If $\lambda(1+\theta) > 0.25$ then there exists a pooling equilibrium. In a pooling equilibrium there exists a set of types $[t_l, 1]$ such that agent A would invest the maximum amount of 1 if of type $t \in [t_l, 1]$. One consequence of this is that agent B cannot infer the type of A if A does invest 1 and hence the equilibrium is a pooling equilibrium. Basically, if we derive the inference function on the assumption of a separating equilibrium we find that A would want to invest more than 1, which is not possible. Limiting investment to 1 there are many types for which agent A would invest 1. To illustrate we plot in Figure 2 the signalling equilibrium when $\lambda = 0.9$ and $\theta = 0.1$.

Figure 2: Deriving the investment function when $\lambda = 0.9$ and $\theta = 0.1$.



The top panel of Figure 2 illustrates that if we were to assume separation then agent A would want to invest more than 1 if of type 0.5 or above. Clearly this is not consistent with equilibrium. In fact, as the bottom figure shows, equilibrium requires agent A to invest 1 if of type $t_l \simeq 0.2$ or above. 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 (3)). 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, resulting in agent B inferring that agent A has type t_l if his investment is between 0.75 and 1.

Proposition 2 *If $\lambda(1+\theta) > 0.25$ then there exists a unique pooling equilibrium that satisfies the D1 Criterion. There exists type $t_l \in [0,1]$ such that agent 1 invests 1 if of type $t \in [t_l, 1]$ and invests according to*

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

if $t < t_l$ where $m = (4\lambda(1+\theta)-1)^{0.5}$. If $2\lambda/3 > 1/(1+\theta)$ then $t_l = 0$ otherwise t_l solves

$$-\left(x_l - t_l\right)^2 - \theta\left(1 - x_l\right)^2 - \lambda(1-t_l)^2 = -\left(1-t_l\right)^2 - \lambda\frac{(1-t_l)^2}{3} \qquad (4)$$

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

Propositions 1 and 2 completely characterize the signalling equilibrium of the model.

Before moving onto our main results and the contrast between naive and rational inferences we can make some general observations. First, we see that if $\lambda + \theta > 0$ 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 naive.

While it is not our primary interest here, we can note that a ceteris paribus increase in λ or θ increases the level of investment. Thus, the more agent A cares about the esteem of B and C then the more he invests. Increasing λ does, however, have a different effect than increasing θ. When $\lambda > 0.25$ then there must be a pooling equilibrium in the sense that agent A will invest the maximum if of many types. If $\lambda \geq 3/2$ then agent A will invest the maximum irrespective of his type. Increasing λ thus creates an incentive for agent A to invest the maximum. This is primarily because a less generous agent A would invest the maximum to 'hide' his lack of generosity. By contrast, if $\lambda = 0$ then there must be a separating equilibrium irrespective of how large is θ. This is because the less generous agent A can hide is lack of generosity to agent C by merely investing more and without need to invest the maximum.

4 The consequences of naive inferences

In the following we fix a value of $E = \lambda + \theta$, the total weight given to esteem, and vary λ and θ, the weights given to the esteem of the rational and naive agent. We shall provide two somewhat counter-intuitive results on the effect of increasing the relative weight given to the esteem of the naive agent on the investment of agent A. We begin by stating the results and explain why we view them as counter-intuitive before discussing the results in more detail. This discussion allows us to gain some insight on how the inferences of agents B and C influence the investment of A.

In any signalling equilibrium it must be the case that if agent A does not choose to invest 1 then the rational agent B will correctly infer his type. This creates a discontinuity in the esteem given by agent B (with more esteem given to those who invest 1) and suggests that A has little incentive to increase his investment unless he makes the 'jump' to investing 1. *If, therefore, the weight put on the esteem of the rational agent is relatively high we might expect to see agent A either investing an amount close to his level of generosity or investing 1.*

By contrast, agent A can easily 'fool' the naive agent C into thinking he is more generous than he is by investing more. An incremental increase in investment will thus increase the esteem received from C. Note, however, that there is a diminishing marginal gain in esteem from being perceived as more generous and indeed zero marginal gain to being perceived as type 1 rather than some type near to 1. *If, therefore the weight put on the esteem of the naive agent is relatively high we might expect to see agent A investing* *more than his level of generosity but not investing 1.*

Putting this intuition together suggests that agent A should be (1) relatively more inclined to invest 1 the more weight is put on the esteem of the rational agent but (2) if he does not invest 1 then he invests relatively more the more weight is put on the esteem of the naive agent. We already have some evidence that this is the case. For example, it is simple to derive (see the Appendix) that agent A invests $\mu(0) = \theta/(1+\theta)$ if of type 0 and so investment is increasing in θ but never equal to 1. Also, we know that a separating equilibrium exists when $\lambda(1+\theta) = 0.25$ and so it is only when the weight λ put on the esteem of the rational agent is sufficiently high that agent A will invest 1 when not of type 1. It turns out, however, that the conjectures (1) and (2) above are false. Specifically.

Result 1 *Fixing the total weight to esteem E, a ceteris paribus increase in θ (and corresponding decrease in λ) can result in agent A decreasing his investment even if he did not invest 1 initially.*

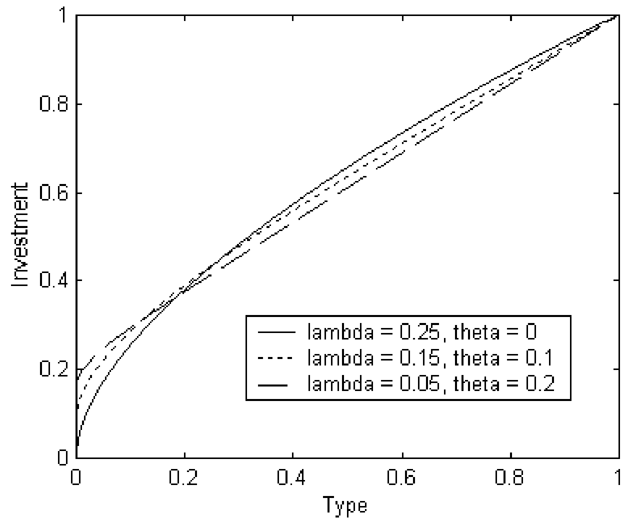
Result 2 *Fixing the total weight to esteem E, a ceteris paribus increase in θ (and corresponding decrease in λ) can decrease t_l and thus increase the set of types for which agent A chooses to invest 1.*

We discuss each result in turn.

4.1 Why naive inferences can reduce investment?

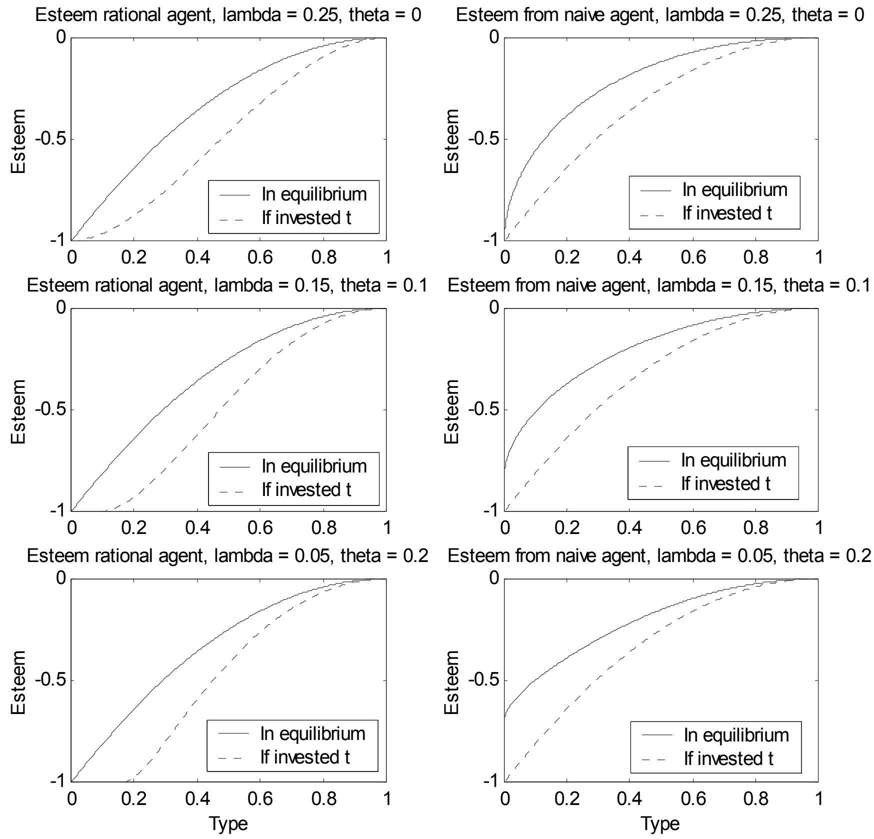
To illustrate result 1 it is easiest to set $\lambda + \theta \leq 0.25$ implying that a separating equilibrium always exists and so agent A would only invest 1 if of type 1. Figure 3 plots data for the case of $\lambda + \theta = 0.25$ and clearly illustrates Result 1. In particular, as θ increases and more weight is put on the esteem from the naive agent, agent A relatively decreases his investment if of type $t > 0.2$.

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



Investing more than his level of generosity is costly in terms of intrinsic utility for agent A. We see, however, that if $t \in (0, 1)$ then A is always willing to pay this cost and invest more than his level of generosity. There are two incentives for doing so: (1) By investing more he can 'fool' the naive agent C into inferring that he is more generous than he actually is. (2) By investing more he signals his type to agent B. Note that B cannot be 'fooled' in the same way as agent C and will always correctly infer the type of A. But, A must invest more than his level of generosity or be inferred by B as being of a lower level of generosity than he actually is. Note that this is the case even if $\theta = 0$. Figure 4 illustrates by contrasting the esteem that agent A gets in equilibrium versus the esteem that he would receive if he were to invest according to his level of generosity. In summary, agent A invests more than his level of generosity in order that agent C gives him more esteem than 'he deserves' and agent B gives him the esteem that 'he deserves'.

Figure 4: The esteem that agent A receives from agents B and C in equilibrium and the esteem he would receive if he invested his level of generosity when $E = 0.25$.

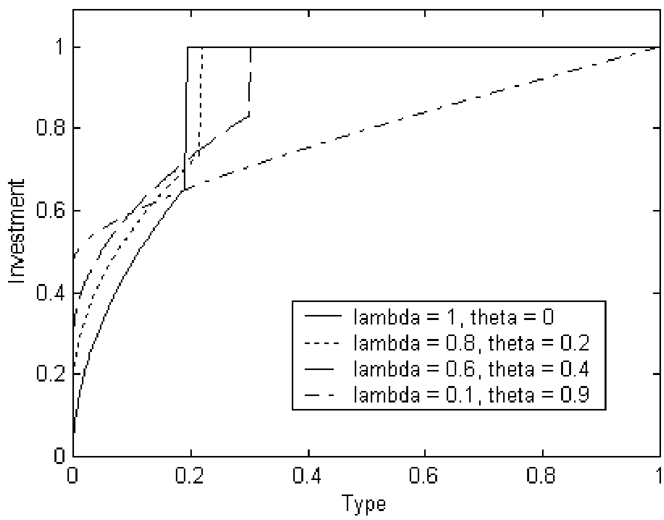


We can now see that result 1 holds because of the concavity of the esteem function. If agent A has a level of generosity near to 0 then the rational agent will always give him low esteem. So, it is relatively more important to 'fool' the naive agent into thinking that he is more generous than he actually is. Thus, the higher is the weighting of the esteem of the naive agent then the more A increases his investment. By contrast, if A is generous then naive agent C will always give him high esteem and so it is relatively more important for A to make sure that the rational agent B knows that he is generous. Thus, the higher is the weighting of the esteem of the rational agent the more agent A invests. Result 1 follows, therefore, from the need for agent A to increase his investment in order to 'convince' the rational agent B of his generosity. It should be clear that Result 1 can hold whenever the esteem function is concave.

4.2 Why naive inferences can increase investment?

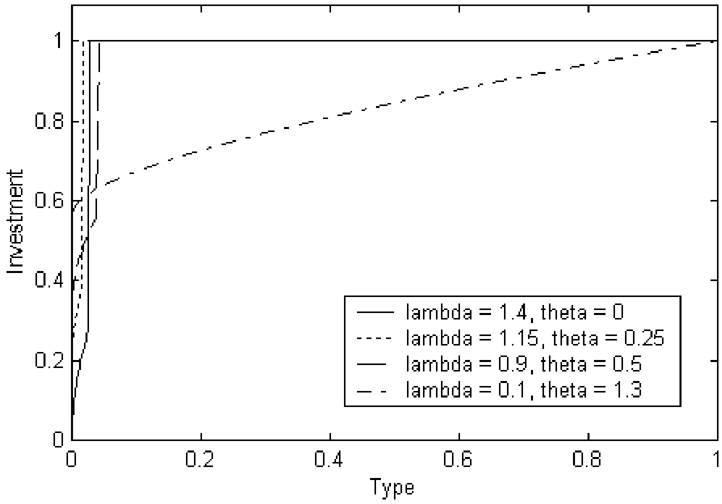
With regard to Result 2 the principle focus is on the smallest level of generosity consistent with agent A investing 1; that is, we focus on t_l. Typically, as we might expect, t_l decreases as more weight is put on the esteem of the naive agent. This is illustrated in Figure 5 which treats the case of $E = 1$.

Figure 5: Investment functions when $E = 1$.



We see, for example, that $t_l = 0.1937$ when $\lambda = 1$, increases to 0.2185 and 0.3033 when $\lambda = 0.8$ and 0.6 respectively. When $\lambda = 0.1$ there is a separating equilibrium and agent A would only invest 1 if of type 1. This picture fits the intuition sketched out previously where the less is the weighting on the esteem of the rational agent then the less appears to be the incentive for agent A to invest 1 so as to 'conceal' his type to agent B. Result 2 suggests, however, that this need not always be the case and this is apparent from Figure 6 which plots the investment function when $E = 1.4$.

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

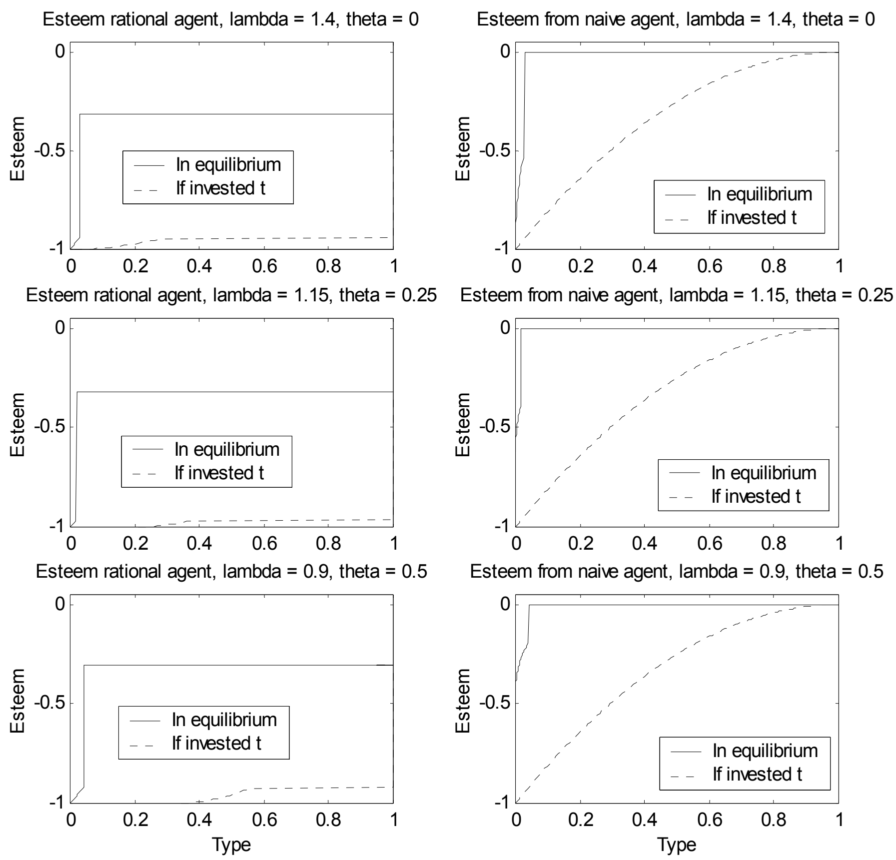


When $\lambda = 1.4$ we find $t_l = 0.0287$ but when $\lambda = 1.15$ we observe a lower t_l of 0.0174. Thus, a ceteris paribus increase in the weight of the esteem of the naive agent increases the set of types for which agent A would invest 1.

In order to explain this we first note that when λ is relatively high there is a strong incentive to invest 1 as evidenced by agent A choosing the norm for most possible types. One consequence of this incentive is that the esteem agent A receives from agent B is relatively low even if he does invest 1 as seen in Figure 7. We also see, as seen before, that agent A has a relatively strong incentive to invest more than his level of generosity if his type is near to 0 because in doing so he can 'fool' agent C into giving him significantly more esteem.

Now, we can explain Result 2. If $\theta = 0$ and $\lambda = 1.4$ agent A would not invest 1 if of type $t = 0.02$ because the increment in esteem he would receive from agent B if were to invest 1 is not quite sufficient to compensate for the loss in intrinsic utility. As θ increases, agent A does have an incentive to invest more if of type 0.02 in order that agent C would infer him to be more generous and thus accord him more esteem. But if agent A is going to increase his investment then it becomes less costly for him to 'go all the way', invest 1 and pick up the higher esteem that agent B gives to those who invest 1. In short, this is what we observe. Result 2, therefore, follows because of the combination of incentives for a not generous agent A to first increase investment in order to get more esteem from agent C and then increase investment to 1 in order to get the high esteem from agent B. It is clear that this result is also quite general and likely to occur whenever agent A is only indifferent between investing 1 or not when he has one of the least generous types.

Figure 7: The esteem that agent A receives from agents B and C in equilibrium and the esteem he would receive if he invested his level of generosity when $E = 1.4$.



4.3 What we learn from Results 1 and 2?

Result 1 shows that an expectation that others will make rational inferences can provide strong incentives to invest more in the public good. Basically, agent A has to invest more in order that he can signal his generosity and prove to the rational agent B that he really is generous and not just trying to appear generous. Typically, in, for example, the labour market , this process of signalling is seen to involve a costly dead-weight loss (Spence 1973). In this context, however, increased investment could be a good thing if it increases investment in the public good towards the Pareto optimal level. Some caution, however, is needed in this interpretation because an expectation that others will make rational inferences also leads to greater inequality in investment. In particular, the less generous will invest relatively more if they expect others to make naive inferences while the generous will invest relatively more if they expect others to make rational inferences. An expectation of rational inferences thus means a relative increase in the proportion of investment that comes from the generous.[4]

Result 2 demonstrates that an upper limit on investment coupled with an expectation of naivety in inferences can result in a less generous agent significantly increasing his investment. The fact that esteem can be easily obtained when inferences are partly naive provides the less generous agent with an incentive to increase his investment. The possibility of obtaining significantly more esteem from those with rational inferences if the investment is increased to the maximum provides a second incentive. Some expectation of naivety in inferences can thus induce a less generous agent to invest the maximum as opposed to relatively little. This suggests, amongst other things, that setting an upper limit on investment in a public project could increase total contributions.[5] The decrease in contributions that would result from constraining the most generous could be more that offset by the increased investment of the less generous, who want 'to hide their lack of generosity'.

At first glance results 1 and 2 may seem contradictory as one says naive inferences decrease investment and the other says that they increase investment. The results are, however, consistent. In particular, result 1 primarily concerns a more generous agent who has an incentive to increase his investment the more is the weight on the esteem of the rational agent because he wants to signal to those with rational inferences that he is generous. Result 2, by contrast, primarily concerns a less generous agent who has an incentive to increase his investment when there is more weight on the esteem of the naive agent and may consequently increase his investment to the maximum.

[4] Switching income level for level of generosity there is experimental evidence that 'lower' types do contribute more (Buckley and Croson 2006).

[5] Related is Harbaugh (1998a, 1998b) who discusses how category reporting can increase contributions.

5 Conclusion

The motivation for this paper was that naivety in inferences, or an expectation of naivety in inferences, may alter the incentives of agents who seek social approval or disapproval. We hope to have demonstrated that naivety 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 naivety in inferences is to be expected. It is also because we feel that acknowledging naivety 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 naive inferences. It would, however, be interesting to relax some of the modelling assumptions and consider more generally the consequences of naive inferences on behavior. For example, we consider two extremes of the agent who has naive 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 naive 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?

Appendix - Proof of Propositions 1 and 2

If agent B perceives agent A to be of type b then agent A's utility is $u(t,x,b) = -(x-t)^2 - \lambda(1-b)^2 - \theta(1-x)^2$. Indifference curves in the (x,b) plane for an agent of type t are thus given by, $(x-t)^2 + \lambda(1-b)^2 + \theta(1-x)^2 = D$, 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,

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

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,

$$\phi_B'(x) = \frac{(1+\theta)\,x - \phi_B(x) - \theta}{\lambda\,(1 - \phi_B(x))}. \tag{5}$$

The differential equation (5) can be rewritten as system

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

where v is some index. Rearranging the bottom equation gives

$$t = 1 - \frac{x'}{\lambda} \tag{6}$$

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

$$x'' + x' + \lambda(1+\theta)x = \lambda(1+\theta). \tag{7}$$

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

Case (1): $\lambda(1+\theta) = \frac{1}{4}$. The solution to equation (7) is

$$x = 1 + C_1 e^{-\frac{v}{2}} + C_2 v e^{-\frac{v}{2}} \tag{8}$$

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

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

Setting $\frac{du}{dx} = 0$ suggests that $t = 0$ as $x = \frac{\theta}{1+\theta}$. Appropriate initial conditions are thus $t = 0$ and $x = \frac{\theta}{1+\theta}$ as $v = 0$. Using equations (8) and (6) in turn gives

$$C_1 = \frac{-1}{1+\theta}; \qquad C_2 = \lambda - \frac{1}{2\,(1+\theta)}. \tag{9}$$

Using (6), (8), (9) and that $\lambda(1+\theta) = 0.25$ one can prove the first part of Proposition 1.

Case (2): If $\lambda(1+\theta) < 0.25$. The solution to equation (7) is

$$x = 1 + C_3 e^{r_1 v} + C_4 e^{r_2 v} \tag{10}$$

where

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

and C_3 and C_4 are constants. Appropriate initial conditions remain $t = 0$ and $x = \frac{\theta}{1+\theta}$ as $v = 0$. So, from (10) we obtain $C_4 = \frac{-1}{1+\theta} - C_3$ and using (6) we get

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

Thus,

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

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

Case (3): $\lambda(1+\theta) > 0.25$. The characteristic equation of (7) has two distinct complex roots and so the solution of (7) has form,

$$x = 1 + e^{-\frac{1}{2}v} \left(C_5 \cos \frac{mv}{2} + C_6 \sin \frac{mv}{2} \right) \tag{11}$$

where $m = (4\lambda\,(1+\theta) - 1)^{0.5}$ and C_5 and C_6 are constants. Appropriate initial conditions remain $t = 0$ and $x = \frac{\theta}{1+\theta}$ as $v = 0$. Using equations (11) and (6) in turn gives

$$C_5 = \frac{-1}{1+\theta}; \qquad C_6 = \frac{2\lambda + C_5}{m}.$$

This expression describes an investment function but results in an $x > 1$ which is not possible. We can, however, use condition (4) 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

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

From this we find that $t_l = 0$ when $2\lambda/3 > 1/(1+\theta)$. This completes the proof of Proposition 2. ■

References

[1] Andreoni, J. (1990). "Impure altruism and donations to public goods: a theory of warm-glow giving," *Economic Journal*, 100: 464-477.

[2] 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.

[3] Bernheim, D. B. (1994). "A theory of conformity," *Journal of Political Economy*, 102(5): 841-877.

[4] 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.

[5] Brennan, G. and Petit, P. (2004). *The Economy of Esteem: An essay on civil and political society*. Oxford University Press, Oxford.

[6] 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

[7] Carmichael, H. L. and MacLeod, W. B. (1997). "Gift giving and the evolution of cooperation," *International Economic Review*, 38(3): 485-509.

[8] 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): 1300-309.

[9] Cho, I. K. and Kreps, D. M. (1987). "Signaling games and stable equilibria," *Quarterly Journal of Economics*, 102: 179-221.

[10] Coleman, J. (1990). *Foundations of Social Theory*. Havard University Press, Cambridge, MA.

[11] Eliaz, K. and Spiegler, R. (2006). "Contracting with diversely naïve agents," *Review of Economic Studies*, 73: 689-714.

[12] Eyster, E. and Rabin, M. (2005). "Cursed Equilibrium," *Econometrica*, 73(5): 1623-1672.

[13] Fehr, E. and S. Gachter (1999). "Collective action as a social exchange," *Journal of Economic Behavior and Organization* 39: 341-369.

[14] Francis, H. (1985). "The law, oral tradition and the mining community," *Journal of Law and Society*, 12: 2267-2271.

[15] Frank, R. (1988). *Passions within Reason: The Strategic Role of the Emotions*. W.W. Norton.

[16] Fudenberg, D. and Tirole, J. (1991). *Game Theory*. MIT Press, Cambridge, MA.

[17] Fudenberg, D. and Levine, D. K. (1993). "Self-confirming equilibrium," *Econometrica*, 61(3): 523-545.

[18] Glazer, A. and Konrad, K. A. (1996). "A signaling explanation for charity," *American Economic Review*, 86(4): 1019-1028.

[19] Harbaugh, W. T. (1998a). "The prestige motive for making charitable transfers," *American Economic Review*, 88(2): 277-282.

[20] 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.

[21] 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.

[22] Hollander, H. (1990). "A social exchange approach to voluntary contribution", *American Economic Review*, 80: 1157:1167.

[23] Jones, E. E. and Harris, V. A. (1967). "The attribution of attitudes," *Journal of Experimental Social Psychology*, 3: 1-24.

[24] Kandel, E. and E. Lazear (1992). "Peer pressure and partnerships" *Journal of Political Economy*, 100: 801-812.

[25] Knack, S. (1992). "Civic norms, social sanctions and voter turnout", *Rationality and Society*, 4: 133-156.

[26] Knack, S. and Keefer, P. (1997). "Does social capital have an economic payoff? A cross-country investigation," *Quarterly Journal of Economics*, 112: 1251-1288.

[27] Lindbeck, A. (1997). "Incentives and Social Norms in Household Behavior," *American Economic Review*, 87: 370-377.

[28] Lindbeck, A., S. Nyberg and J. Weibull (1999). "Social norms and Economic Incentives in the Welfare State," *Quarterly Journal of Economics*, 114: 1-35.

[29] Nyborg, K. and Rege, M. (2003). "On social norms: the evolution of considerate smoking behavior," *Journal of Economic Behavior and Organization*, 52: 323-340.

[30] Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA.

[31] Rabin, M. and Schrag, J. L. (1999). "First impressions matter: a model of confirmatory bias," *Quarterly Journal of Economics*, 114(1): 37-82.

[32] Rege, M. (2004). "Social norms and private provision of public goods," *Journal of Public Economic Theory*, 6(1): 65-77.

[33] Rege, M. and K. Telle (2002). "An Experimental Investigation of Social Norms", *Journal of Public Economics*, 88: 1625-1644.

[34] Spence, M. (1973). "Job market signaling," *Quarterly Journal of Economics*, 87(3): 355-374.

[35] Thaler, R. H. (1988). "Anomalies: The Winner's Curse," *Journal of Economic Perspectives*, 2: 191-202.

[36] Yildiz, M. (2007). "Wishful thinking in strategic environments," *Review of Economic Studies*, 74: 319-344.