Continuous Preferences Can Cause Discontinuous Choices: an application to the impact of incentives on altruism

Paul Seabright
IDEI, Université de Toulouse-1 & CEPR[1]
Manufacture des Tabacs, 21 allée de Brienne,
31042 Toulouse Cedex, France
email seabrigh@cict.fr
telephone +33 561 12 86 17
www.idei.asso.fr

This version: January 20, 2004

Abstract

Implementing incentive systems can sometimes backfire in practice: experimental evidence and folklore both suggest that offers of explicit rewards can expose surprising discontinuities in behaviour. This paper models two such discontinuities that have been claimed by psychologists and experimental economists to constitute important exceptions to the standard economic theory of human motivation. The first ("type discontinuity") is the observation of a discontinuity in the distribution across population types of values of the willingness to accept payment in return for performing certain ("civic") actions, such as giving blood or performing public service. It is claimed that this distribution is bimodal, even discontinuous: many people have a zero WTA, many have a large positive WTA, but nobody has a small positive WTA. The second ("payment discontinuity", also known as "crowding-out") is that people who are willing to perform certain actions for free will refuse to perform them for a low price, even if they subsequently agree to perform them if the price is raised enough. Civic virtue may, on this view, be crowded out by the introduction of explicit incentives; people may stop doing things they were previously prepared to do without reward. The paper shows that both phenomena may be observed as a result of individuals' acting in a first period of public service in the knowledge that the terms of their action signal their type, and their type will affect a process of assortative matching in a second period. Type Discontinuity, but not Payment Discontinuity, is observed in a signaling game in which individuals announce the prices at which they will perform a civic action. Payment Discontinuity, but not Type Discontinuity, is observed in a screening game in which individuals have only a binary participation decision available to signal their type. The proportion of individuals participating when rewards are zero can be higher than when rewards are positive but small.

Keywords : crowding effect, intrinsic motivation, assortative matching, economic psychology, incentive schemes
JEL codes: A12, C70, D10, D60, H41, J22

[1] I am grateful to Emmanuelle Auriol, Ernst Fehr, Jennifer Gann, Patrick Rey and Jean Tirole for valuable discussions on these questions, as well as participants in the Workshop on Social Norms at the Wissenschaftskolleg Berlin during my visit there in December 2001.

1. Introduction

Are people who demand small bribes almost as honest as people who demand no bribes at all? Is selling something for a very low price almost equivalent to giving it as a gift? And can offering someone a small fee make them less willing to do something they would gladly have undertaken for free?

Implementing incentive systems can sometimes backfire in practice: experimental evidence and folklore both suggest that offers of explicit rewards can expose surprising discontinuities in behaviour. Economists brought up on incentive theory tend to believe that a small reward is, though small, a genuine reward, and should be expected to increase people's motivation. Yet they may be surprised to observe that small rewards can sometimes be scorned by those to whom they are offered; conversely, those who accept or solicit small rewards can be disdained by others. More dramatically, some kinds of action that would be considered intrinsically acceptable or even admirable are considered "tainted" or repugnant when undertaken for a reward, even a large one (an example is the donation of a kidney, which is considered admirable when undertaken without reward, but "morally offensive and ethically indefensible" when undertaken for sale, to cite one contribution to the debate on the passage of the US National Organ Transplant Act)[2].

Can we understand the nature of the discontinuity here? In fact there is not one discontinuity but two, and as will be seen these are quite distinct even though they are often

[2] See Gann (2001) who writes: "In September 1999 an individual offered his right kidney for sale on eBay, an internet based auction site. In America, where there are over 47,000 patients awaiting kidney transplants, and where the average wait for a kidney transplant

confused. One is a discontinuity of behaviour in the space of types, while the other is a discontinuity of behaviour in the space of incentives. The first (the Type Discontinuity) is the observation of a discontinuity in the distribution across the population of values of the willingness to accept payment in return for performing certain ("civic") actions, such as giving blood or performing public service. It is claimed that this distribution is bimodal, even discontinuous: we may see some individuals willing to perform these actions for free and some willing to do so only for a significant payment, but we will rarely see the intermediate case where some individuals perform them only for a small payment. Furthermore the reactions of others to the intermediate case tends to reinforce the discontinuity: people with a low but positive willingness-to-accept may be stigmatized as "cheap", but people willing to act for free are not considered even cheaper.

The second (the Payment Discontinuity) is also known more familiarly in the literature on economic psychology as "crowding-out". This is the finding that people who are willing to perform certain actions for free will refuse to perform them for a low price, even if they subsequently agree to perform them if the price is raised enough. Civic virtue may, on this view, be crowded out by the introduction of explicit incentives; people may stop doing things they were previously prepared to do without reward.

These two alleged discontinuities are quite distinct from each other. The Type Discontinuity is a claim about the distribution of values of the willingness to accept payment;

---

nearly doubled between 1988 and 1996, this excited considerable interest. The bidding had reached $5.8 million before being shut down by the administrators of eBay".

it is compatible with an entirely orthodox interpretation of WTA. The Payment Discontinuity, by contrast, appears to cast doubt on the very concept of WTA according to which higher associated payments always make a given option more attractive. In fact, as will be seen below, both can be reconciled with more orthodox accounts of motivation; the difference between them will lie in the nature of the actions available to agents and the interpretation that others place upon them.

In section 2 I briefly survey evidence for the Payment Discontinuity, and discuss explanations that have been advanced for it. I point out that such explanations, though often both intuitively and empirically plausible, rely upon an unexplained difference in the perception by individuals of the commercial and voluntaristic behaviour, which it should be a part of the theory to explain (in a sense they make the Type Discontinuity a part of the explanation for the Payment Discontinuity). In section 3, therefore, I propose a simple economic model to explain both discontinuities. I also show that the circumstances that favour the Type Discontinuity do not necessarily favour the Payment Discontinuity, which is observed under somewhat more restrictive circumstances. Section 4 summarises and concludes.

In contrast to the Payment Discontinuity, the Type Discontinuity as such has been subjected to much less careful investigation. The evidence for it is more casual and anecdotal (it is often regarded as so obvious as to be a plausible component of the explanation for the Payment Discontinuity). To make it plausible let me ask the reader to consider our highly different reactions to two pairs of cases:

- In the first variant of the first case, imagine going to a politician who has undertaken to raise in your country's parliament an issue affecting your civil rights. He explains that there are costs associated with the preparation of your dossier and you are required to pay an official fee, based on a published tariff, equal to a hundred dollars. However, he will devote himself enthusiastically to pursuing your cause. You return the next week and he apologises profusely but says there has been a rise in the published tariff of fees so that you must now pay two hundred dollars
- In the second variant of the first case, when you return the next week, he explains that he has realized that it will be lot more effort to prepare your case than he had realized, and that he therefore requests you to pay him, in addition to the published fee of a hundred dollars, a small bribe of one hundred dollars to make it worth his while.
- In the first variant of the second case, you need a transfusion of a rare blood type. You learn that a donor has been found who has donated his blood. You pay the transfusion service a hundred-dollar administration fee.
- In the second variant of the second case, the transfusion service charges you nothing but says that the donor has asked a price of a hundred dollars for selling you the blood.

By understanding why we react so differently to the behaviour of the parties in each of these pairs of cases, I suggest we can understand why individuals do not consider that

making gifts can be considered close to making sales for low prices, nor that taking small bribes is close to taking no bribes at all. And we would not expect to observe many people asking for very small bribes. If they ask at all, they will demand significant sums.

2. Markets and the "crowding-out" of reciprocity

So what evidence is there for the Payment Discontinuity? A number of writers, beginning most famously with Richard Titmuss (1970), have claimed that giving explicit (and usually but not necessarily monetary) compensation to individuals can undermine their sense of civic duty, specifically by diminishing their willingness to do things for the public good that they would have been entirely willing to do for free. Titmuss suggested that this might be particularly true of the market for blood. He claimed that paying donors negatively affected their willingness to donate blood, thereby leading to dominance among blood donors of those who needed to donate for financial reasons, whose blood was likely to be medically much less suitable. Titmuss himself did not provide convincing empirical evidence in support of his claim; other authors such as Solow (1971) and Arrow (1972) agreed with him that altruistic motivation might be important but assumed that price incentives could be regarded as additive, so that the supply curve for blood would be positively sloped in the standard way.

As far as I am aware there has been no more recent empirical work on the market for blood that might allow Titmuss's claim to be more rigorously tested (though Gann, 2001,

develops a theoretical model which takes his motivational assumptions for granted and shows their consequences for the quality of aggregate blood supply). But empirical work in other contexts by a number of researchers suggests the possibility of crowding-out of civic virtue should be taken very seriously (see Frey & Jegen, 2001, for an overview). One of the best-known early studies was by Deci (1971) who suggested that paying experimental subjects to solve puzzles during an experiment decreases their subsequent willingness to solve such puzzles for fun. Gneezy & Rustichini (2000a) show that when children doing volunteer work (going from house to house collecting donations for charity) are paid a small monetary reward, the intensity with which they work declines, though it recovers again with subsequent increases in the level of payment. They call the effect "pay enough or don't pay at all", and although this is clearly interpretable as a crowding-out effect, a similar phrase might be used to describe the Type Discontinuity. The same authors (Gneezy & Rustichini, 2000b) also report an experiment in which the introduction of a fine for parents who collected their children late from kindergarten increased the rate of late collection (a phenomenon they interpret as being due to the fact that the possibility of paying for late collection reduces the perceived element of social disapproval)

Bruno Frey and co-authors (Frey et.al., 1996; Frey & Oberholzer-Gee, 1997) have suggested that willingness of individuals to contribute to public goods may be undermined by explicit payment. In particular, they draw on survey evidence of people's willingness to accept privately noxious but socially necessary facilities (such as nuclear waste recycling plants). This evidence reveals that offering compensation does not increase the acceptability

of such projects, and indeed often elicits complaints about “bribery”; the authors interpret these findings as due to the “crowding-out” of public spirit by private incentives.

There remain some unresolved issues about how to interpret these findings (see Fehr & Falk, 2001). Some of these unresolved issues are directly empirical – such as whether people’s reported willingness to accept is a reliable indicator of their actual willingness to accept in practice (there are significant discrepancies between the two in the results reported by Frey et.al., 1996). Some have to do with the difficulty of designing experiments to control for other effects than those strictly describable as crowding-out (often negative reciprocity and loss aversion may be present in the same context, for instance – see Fehr & Falk, 2001, p.37). Some have to do with the difficulty of knowing what signals are being perceived by subjects in the experimental contexts (including signals about the social norms that are relevant to that context) and therefore to what non-experimental settings the findings could be considered relevant. Nevertheless, prima facie evidence for crowding-out has appeared often enough for it to be worth considering what motivational foundations could explain such a phenomenon.

Most theoretical explanations appeal to the presence of two distinct sources of motivation, sometimes known as extrinsic and intrinsic motivation; the first suggests that the actions concerned are performed in order to achieve some other end (such as payment), while the second suggests that the action yields satisfaction or pleasure in itself. For reasons that are then usually left unexplained, the nature of the extrinsic motivation interacts with the strength of the intrinsic motivation in some way, the two nevertheless remaining quite

distinct arguments of the utility function[3]. For instance, Frey & Oberholzer-Gee (1997) suppose that individuals gain utility from ordinary consumption (and thereby indirectly from money), but also from "behaving in an altruistic manner or living up to her civic duty". Offering payment for actions that are thought to be part of an individual's civic duty increases that person's consumption possibilities but reduces her utility from behaving altruistically. It may therefore reduce an individual's utility overall, and may therefore reduce the willingness to undertake such actions. A similar mechanism is invoked by Gann to explain reduced supply of blood when payments are made. "Giving blood" and "selling blood" are thus considered two distinct kinds of activity, the former yielding some intrinsic utility but the other not. Offering payment transforms the former activity into the latter, implying a utility loss.

There are two limitations of this theoretical approach, which are my reason for outlining a different framework. There is nothing at all implausible about the claim that individuals may be altruistically motivated[4], and whether this motivation is best captured by adding arguments to the utility function depends on the problem in hand (there are clearly some kinds of altruism that do not increase the person's well-being, in that they reflect duty rather than delight, and may move the individual to action even though she may heartily wish

[3] This is not true of Bénabou & Tirole (2003), who suggest an interesting mechanism whereby offers of explicit incentives by an informed principal signals something to an agent about her own type, and this type information interacts with intrinsic motivation. In the model of this paper, by contrast, the principal has no private information.

[4] There are also many examples of individuals creating public goods for free when these arise as by-products of activities that are privately valuable for them. See Bessen (2001) on the open-source software movement.

she did not have that particular duty[5]). However, there is something rather arbitrary about supposing that feelings of altruism attach to actions performed purely under certain descriptions, and that an action with identical consequences might not elicit altruism even if the person concerned knew the consequences were the same. It is not that such framing effects are necessarily implausible (framing is a pervasive experimental phenomenon[6]), but they are certainly arbitrary. There seems no more general account of why two actions may be described in ways that elicit such different reactions, nor why such descriptions could be stable under reflective consideration of the consequences[7]. I suggest that it should be part of the goal of the theory to explain *why* giving and selling are considered to be radically different activities, and that such a distinction should ideally not be presumed from the outset.

The almost comic arbitrariness of such descriptions is turned to good effect in Mark Twain's *Adventures of Tom Sawyer*:

> "[Tom] had discovered a great law of human action, without knowing it - namely, that in order to make a man or a boy covet a thing, it is only necessary to make the thing difficult to attain. If he had been a great and wise philosopher, like the writer of this book, he would now have comprehended that Work consists of whatever a body

[5] A similar point underlies Sen's famous distinction between "sympathy" and "commitment" (Sen, 1977).
[6] See most obviously Kahneman & Tversky (1979).
[7] Robert Nozick replied to an argument of Bernard Williams about doctoring being an activity that was intrinsically about curing patients by asking what distinguished it from "schmoctoring", which was just like doctoring except that *its* purpose was to make money for the practitioners.

> is obliged to do, and that Play consists of whatever a body is not obliged to do. And this would help him to understand why constructing artificial flowers or performing on a treadmill is work, why rolling tenpins or climbing Mont Blanc is only amusement. There are wealthy gentlemen in England who drive four-horse passenger coaches twenty or thirty miles on a daily line in the summer, because the privilege costs them considerable money; but if they were offered wages for the service, that would turn it into work, and then they would resign".

The second shortcoming of the theory is that it implies considerable short-sightedness among agents, since an agent who wished to continue to enjoy the warm glow attached to performing her civic duty could simply give the money she receives to a charity and think of the action as not only a performance of her civic duty but also a form of "raising money for good causes". There is a lot of anecdotal evidence that in some contexts people do indeed reason in this way (some academic journals pay referees a – fairly miserable – fee but then invite them to donate the fee to charity). Nevertheless, if they always reasoned thus there would be no crowding out effect, and it is hard to think that crowding out arises, if indeed it does, purely because of agents' stupidity.

I want instead to propose that the reason why individuals may reject monetary payments for actions they would perform for free is, at least partly, a desire to send a signal to other individuals about the type of person they are[8]. Now it might be thought that this

[8] This does not rule out the possibility that individuals may also wish to send signals to themselves (an important theory along these lines with application to crowding out has been developed by Benabou & Tirole, 2003). Seabright (2001) discusses this possibility in relation

merely replaces one type of arbitrariness by another, but as will be seen the model derives a qualitative discontinuity between the signals sent by individuals even though there is no discontinuity between their types. Some individuals perform certain civic actions for free, while others do so for payment, but the lowest payment demanded by anyone is significantly above zero.

The key to the result is the way in which individuals benefit from being recognised by others as of a particular type, namely the type that does civic actions for free. If this simply enabled individuals to gain greater monetary rewards in the future it is hard to see how it would be considered particularly meritorious. On the other hand, if it simply gave people greater psychic rewards we would be replacing the arbitrary psychic rewards of performing one's civic duty with the no less arbitrary psychic rewards of being recognised as performing one's civic duty.

Instead I propose that the main rewards that come from signalling one's performance of one's civic duty consist in the increased likelihood of subsequent interaction with other people who also perform their civic duty. Human social life is full of networking and interaction. Only very rarely do we interact with a whole mass of our fellow citizens; much

---

to the theory of consumer branding: "in principle one could imagine two main kinds of rationale. One, which is compatible with modern evolutionary psychology, is that individuals may have within themselves multiple centres of cognition and reasoning (see Dennett, 1995; Pinker, 1998) which find it valuable sometimes to communicate through the external world rather than internal neural channels, perhaps because internal communication suffers from a lack of credibility. I tell myself I am rich, good-looking and successful; to silence the skepticism of my inner voice I behave in ways that make it seem more likely to myself that I am indeed rich, good-looking and successful. Another explanation is that consumers may

more often we interact in families, workgroup, societies, associations and all the multifarious institutions of civil society. All of these institutions benefit from reciprocal behaviour, and the quality of life that you enjoy in such institutions is determined not only by what you bring to the interaction yourself but also by the kinds of people with whom you interact.

There has been much work recently examining the characteristics of institutions in which the benefits of association to an individual depend upon both that individual's characteristics and the characteristics of the other members (Shimer & Smith, 2000). Such phenomena have been applied to understanding growing inequality in household income (Deaton, 1995; Lerman, 1996), poverty traps in developing economies (Kremer, 1993), peer-group lending in poor countries (Ghatak, 1999), rising divorce rates (Weiss, 1993), transmission rates of HIV infection (Dow and Philipson, 1996), racial and class segregation in the schooling system (Benabou, 1994) and the changing employment structure of US firms (Kremer & Maskin, 1996; Acemoglu, 1998; Mailath et.al., 2000). A key feature of such institutions is that they give rise to what is called "assortative matching". Individuals scoring highly on some relevant (utility- or productivity-enhancing) feature tend to match with other individuals who also score highly on that feature, and low-scoring individuals match with other low-scoring individuals. The reason for this is that although everyone may wish to match with the high-scoring individuals, individuals who are themselves high-scoring have a greater ability to bid for such matches. One consequence of it is that low-scoring individuals suffer twice over; once from their own low score and once from the low score of the other individuals with whom they are obliged to interact.

---

find out about their own characteristics through consumption decisions: I do not know how

In this model, therefore, I shall propose that individuals differ in the extent to which they derive benefits from performing some civic action. Individuals who benefit greatly from doing so, and who can credibly signal that they do, will tend in subsequent social interactions to be matched with other individuals who also benefit from performing the civic action. Both are likely to enjoy enhanced welfare as a result.

The result is that those individuals whose benefits from performing the civic action are above some threshold level will do so for free, while all others will do so only for a fee. The fee demanded by those whose benefits are only just below the threshold level is substantially above zero, since by revealing that they are not in the "civic virtue" group they forgo they chance of associating in the future with highly civically virtuous individuals.

---

fit I am until I go to the gym, I do not know whether I like caviar till I try it, and so on".

3. A model of commercial and voluntaristic transactions

In this model there is a continuum of individuals who each live for two periods. There is no discounting. In the first period they engage in a public ("civic") activity, while in the second they match with other individuals and engage in a private activity. While the civic activity creates benefits for other members of society, these benefits play no intrinsic part in the analysis that follows and therefore do not need to be explicitly defined. However, individuals differ in the extent to which they are motivated by acting for the good of society. An individual i has a twice continuously differentiable utility function $U_i(m_i,c_i,b_i,\theta_i,\theta_j)$, where

m_i is the individual's holding of a money numeraire

c_i is the cost to the individual of performing the public activity

b_i is the benefit to the individual of performing the public activity

θ_i is the individual's type

θ_j is the type of the individual with whom she is matched.

We suppose that the utility can be written as follows:

$$(1) \qquad U_i(m_i,c_i,b_i,\theta_i,\theta_j) = m_i - c + b_i(\theta_i) + V_i(\theta_i,\theta_j)$$

with $\partial b_i(\theta_i)/\partial\theta_i > 0$, $\partial V_i(\theta_i,\theta_j)/\partial\theta_i > 0$, $\partial V_i(\theta_i,\theta_j)/\partial\theta_j > 0$ and $\partial^2 V_i(\theta_i,\theta_j)/\partial\theta_i\partial\theta_j > 0$.

This implies that the costs of performing the civic action are identical (and common knowledge) across individuals, while the benefits are an increasing function of the individual's type. The utility *V(.)* from the private activity is a function of the types of both the individual and the matched partner.

Actions take place in this model in one of two ways:

- the first is a signalling game, in which the public authority first announces a threshold price p^*, and individuals then announce a non-negative price p_i at which they will be willing to engage in the civic activity. Those who have announced prices below p^* participate in the public activity, receiving their announced price; all others receive a reservation utility normalised to zero. In the second period individuals are matched with each other. We assume that the matching process randomly matches those who have the same expected type conditional on their first period action[9].
- The second is a screening game, in which the public authority announces a price p' at which participation in the civic activity will be remunerated; all individuals who choose to participate receive this price. Then each individual announces a participation decision a_i after which participation takes place,

[9] We do not model the matching process explicitly but draw on the standard findings in the literature.

then individuals are matched with others according to their expected type conditional on their participation decision.

To summarize, in the signalling game, an action profile is a function $p_i = h(\theta_i, p^*)$ where $p_i \in [0,\infty)$. In the screening game, an action profile is a function $a_i = g(\theta_i, p')$ where $a_i \in \{0,1\}$.

In both games we look for a Nash Equilibrium, subject to the constraint that in the second period individuals are matched with those that have the same expected type conditional on their first period action[10]. We suppose that there is a distribution $F(\theta)$ of types θ along an interval $[\theta^L, \theta^H]$, where $\theta^L < c < \theta^H$. We assume that θ^H is sufficiently high that some individuals will wish to announce zero prices, while θ^L is sufficiently low that some individuals will wish to announce positive prices. For simplicity, we also assume $b_i(\theta_i) = \theta_i$, so that types can be considered as indexed by the benefits they gain from the public activity. Together these assumptions imply that there are some individuals who would prefer not to engage in the public activity without payment while there are others who benefit from doing so even without payment.

It will be useful to define $v^H(\theta_i)$ as the expected utility of an individual of type θ_i in the second period if she pools with all weakly higher types $\theta_j \geq \theta_i$, that is, if she is matched at random with one of the set of all individuals with types (weakly) higher than her own.

[10] This is not strictly a Perfect Bayesian Equilibrium since no actions are chosen in the second period, but the conditions on the expectations are the same as in a PBE.

Similarly define $v^L(\theta_i)$ as the expected utility of an individual of type θ_i in the second period if she pools with all (weakly) lower types $\theta_j \leq \theta_i$. We can also define $w^H(\theta_i,\theta_k)$ and $w^L(\theta_i,\theta_k)$ as the expected utility of an individual of type θ_i in the second period if she pools with all (weakly) lower types than type θ_k (not necessarily her own). Of course, $v^H(\theta_i)= w^H(\theta_i,\theta_k)$ and $v^L(\theta_i)= w^L(\theta_i,\theta_k)$ whenever $k=i$.

We first show that announced prices in the signalling game are discontinuous in θ:

Proposition 1 (Type Discontinuity):

In any equilibrium of the signalling game, there exists a threshold value θ^* such that all individuals of type $\theta > \theta^*$ announce a zero price, and all individuals of lower type announce prices that are strictly greater than zero.

Proof:

I first assume, and later prove, that all and only the individuals with a type above some threshold value θ^* announce zero prices. In effect they announce that they are willing to participate in the activity for free (as a "gift"). Then in equilibrium each will be matched in the second period with an individual chosen at random from all those whose type is higher than θ^*. The expected utility of an individual announcing a price of zero is:

(2) $$E\left[U_i\left(m_i, c_i, b_i, \theta_i, \theta_j\right) \middle| p_i = 0\right] = m_i - c + \theta_i + \frac{1}{1 - F(\theta^*)} \int_{\theta^*}^{\theta^H} V\left(\theta_i, \theta_j\right) f\left(\theta_j\right) d\theta_j$$

An individual who announces a positive price, however, faces a calculation of a somewhat different kind. In equilibrium she will be matched with another individual who has announced the same price and who therefore reveals herself to be the same type. To see this, first define $\phi(p, \theta_i)$ as follows:

(3) $$V(\theta_i, \phi(p, \theta_i)) \equiv E\left[V\left(\theta_i, \theta_j\right) \middle| p_j = p\right]$$

Intuitively $\phi(p, \theta_i)$ can be interpreted as the "certainty equivalent type" with whom, if type θ_i were paired, she would receive the same utility as the expected utility she receives from being matched randomly with all those with whom she in fact pools. Now write her expected utility from announcing a price p_i as follows:

(4) $$E\left[U_i\left(m_i, c_i, b_i, \theta_i, \theta_j\right) \middle| p_i > 0\right] = m_i - c + p_i + \theta_i + V\left(\theta_i, \phi\left(p_i, \theta_i\right)\right)$$

If she is maximising her utility we can write the first-order condition as follows:

(5) $$-1 = \frac{\partial V\left(\theta_i, \phi\left(p_i, \theta_i\right)\right)}{\partial \phi} \cdot \frac{\partial\left(\phi\left(p_i, \theta_i\right)\right)}{\partial p_i}$$

However, we also know from the conditions on the utility function that

$$(6) \qquad \frac{\partial^2 V(\theta_i, \phi(p_i, \theta_i))}{\partial \phi \partial \theta_i} > 0$$

from which we know that equation (5) will not be satisfied by the same value of p_i at any two different values of θ_i. Given that her announcement perfectly reveals her type she will therefore be matched with an individual of her own type. In equilibrium her expected utility can therefore be written as follows:

$$(7) \qquad E\left[U_i(m_i, c_i, b_i, \theta_i, \theta_j) \middle| p_i > 0\right] = m_i + p_i - c + \theta_i + V(\theta_i, \theta_i)$$

Define individual of type θ^* as one who is just indifferent between announcing a zero price and announcing the positive price that maximises her expected utility. For such an individual, setting equations (2) and (7) equal yields:

(8) $$\frac{1}{1-F(\theta^*)}\int_{\theta^*}^{\theta^H} V(\theta^*,\theta_j) f(\theta_j) d\theta_j = p_i + V(\theta^*,\theta^*)$$

from which it follows that p_i must be strictly greater than zero. Denote this value by $p_i(\theta^*)$.

It remains to be shown, first, that if an individual of type θ_i announces a zero price, then all individuals of type $\theta_k > \theta_i$ also do so (the assumptions on θ^L and θ^H ensure that such an individual exists, and also that not all individuals announce zero prices). The assertion follows from differentiating equation (5) and using equation (6) to show that

(9) $$\frac{\partial^2 V(\phi(p_i,\theta_i))}{\partial p_i \partial \theta_i} < 0$$

Thus if any individual of type θ_i is dissuaded from announcing a positive price by the reduction in the expected quality of her match, an individual of higher type will be even more dissuaded. The fact that the utility function is linear in money ensures that the higher type cannot be recompensed for this by a higher marginal utility of money. QED.

Figure 1 illustrates. The broken horizontal line represents the cost of participating in the civic activity while the two positively sloped dark lines represent the benefits (summed over both periods) under alternative assumptions about matching in the second period. The convex line shows the benefit of participation under the hypothesis that the individual is matched precisely with another of her type (call this "perfect matching"). The straight line is the benefit of participation under the assumption that the individual is matched at random with the set of individuals that are of weakly higher type than her (call this "pooling"). (The fact that one is drawn convex and the other straight is unimportant; what matters is that the former is steeper than the latter.) For the highest values of θ the individual would prefer perfect matching, but is unable to find a way to signal her type because of the non-negativity constraint on announced prices. For values below where the two dark lines cross the individual prefers pooling with other (higher) types, until we reach types below θ^*, at which point the positive price that the individual could announce while separating from the types lower than her own is just high enough to outweigh the benefits of pooling. At θ^* this price is strictly positive because here the benefits of participation under perfect matching must be lower than the benefits under pooling with individuals of higher type. Here, precisely, we see the Type Discontinuity hypothesis, describable by the phrase of Gneezy & Rustichini: "pay enough or don't pay at all".

As drawn, this price yields significant rents to the individual of type θ^*, since her gross benefits of participation are significantly higher than the costs. How large precisely will these rents be? The answer depends on p^*, the threshold price announced by the public authority. At that price the marginal participant will be the one for whom participation rents

are zero, namely the one for whom $p^* - c + \theta + V(\theta,\theta) = 0$. Proposition 2 shows that under separation, rents to participation are strictly increasing in θ, so that all participants of higher type than the marginal participant receive strictly positive rents.

Proposition 2 (Rents under separation increasing in θ):

$R(\theta) \equiv \theta + V(\theta,\theta) + p_i(\theta) - c$ is increasing in θ.

Proof:

From equation (5) we know that in equilibrium

$$(10) \qquad \frac{\partial p_i}{\partial \phi_i} = \frac{\partial V(\theta_i, \phi(p_i, \theta_i))}{\partial \phi_i}$$

where we write $\phi_i \equiv \phi(p_i, \theta_i)$.

Totally differentiating V(.) yields

(11) $$\frac{dV}{d\theta_i} = \frac{\partial V(\theta_i, \phi_i)}{\partial \theta_i} + \frac{\partial V(\theta_i, \phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \theta_i}$$

Substituting (10) in (11) yields:

(12) $$\frac{\partial p_i}{\partial \theta_i} = \frac{\partial V(\theta_i, \phi_i)}{\partial \theta_i} - \frac{dV}{\partial \theta_i}$$

Totally differentiating the expression for rent, substituting (12) and using the envelope theorem yields:

(13) $$\frac{\partial R(\theta_i)}{\partial \theta_i} = 1 + \frac{\partial V(\theta_i, \phi_i)}{\partial \theta_i} > 0$$

Q.E.D.

Figure 2 illustrates, showing that the participation constraint is steeper than the incentive constraint, leaving participants who announce prices lower than p^* with a strictly positive rent. The participation constraint (the minimum price at which individuals will participate) reaches zero at the point where the horizontal line representing cost of participation cuts the curve representing benefits under separation. Here the actual announced price as determined by the incentive constraint is still strictly positive.

What about crowding out? In the signalling game there is no crowding out, as can be seen from noting that the proportion of individuals whose participation constraint is satisfied is strictly increasing in the threshold price p^*. This is because the participation decision as such is not serving any signalling function. However, matters are quite different when we come to the screening game.

Recall that in the screening game individuals are paid the price announced by the public authority, not the price they announce themselves. This means that, unlike in the signalling game, their participation decision is the only way they have to signal their type. Proposition 3 shows that under these different conditions, and given an additional assumption about payoffs, there will indeed be crowding out. This is because individuals with high θ can signal this fact only by agreeing to participate when prices are zero and refusing to do so when prices are positive. For this to be a rational strategy, their gains from successful signalling have to increase faster in θ than their direct gains from participation. Proposition 3 states more precisely what it means for this latter condition to hold.

Proposition 3 (Payment Discontinuity):

Define θ^* as in Proposition 1. If $\partial[w^H(\theta,\theta_k) - w^L(\theta,\theta_k)] / \partial\theta > 1$ for all θ_k, there exists an equilibrium of the screening game in which, if the authority announces a price $p' = 0$ all and only individuals with $\theta_i \geq \theta^*$ participate in the civic activity, while if it announces a price within an interval of strictly positive prices there are threshold values of θ, one for each price within the interval, such that all and only individuals with values lower than the threshold participate, and the proportion of individuals participating is strictly lower than at $p' = 0$.

Proof:

At an equilibrium with price $p' = 0$ all and only individuals with $\theta_i \geq \theta^*$ will participate by definition of θ^*. Now define a price p_L which is the lowest price at which the lowest-value type θ_L will participate, conditional on being the only participant. Consider a price $p^+ > p_L$. Define θ^+ as the value of θ_i at which individual i is just indifferent between participating and not participating, conditional on believing that only individuals with values $\theta_j < \theta^+$ will participate. Then we can set the benefits of not participating equal to the benefits of participating, as follows:

(14) $v^H(\theta^+) = v^L(\theta^+) + \theta^+ - c + p^+$

which implies

(15) $v^H(\theta^+) - v^L(\theta^+) = \theta^+ + p^+ - c$

Now consider the incentives for individuals with $\theta' \geq \theta^+$. For each such individual to participate requires that

(16) $w^H(\theta', \theta^+) - w^L(\theta', \theta^+) \geq \theta' + p^+ - c$

We can re-write equation (15) as

(17) $w^H(\theta^+, \theta^+) - w^L(\theta^+, \theta^+) \geq \theta^+ + p^+ - c$

which we can subtract from (16) to yield

(18) $[w^H(\theta', \theta^+) - w^H(\theta^+, \theta^+)] - [w^L(\theta', \theta^+) - w^L(\theta^+, \theta^+)] \geq \theta' - \theta^+$

If $\partial[w^H(\theta,\theta_k) - w^L(\theta,\theta_k)] / \partial\theta > 1$ holds for all θ_k then it holds in particular for $\theta_k = \theta^+$, implying that $\partial[w^H(\theta,\theta^+) - w^L(\theta,\theta^+)] / \partial\theta > 1$ which in turn implies equation (18). Thus we

can confirm that all individuals with $\theta \geq \theta^+$ will indeed participate while all those with $\theta < \theta^+$ will not, and thus that this strategy constitutes a Nash equilibrium.

Next, from the properties of the *V(.)* function (specifically the positive cross-partial derivative) we know that

(19) $v^H(\theta^+) - v^L(\theta^+) > w^H(\theta', \theta^+) - w^L(\theta', \theta^+)$ for $\theta' > \theta^+$

This implies that θ^+ is an increasing function of p^+. By setting $p^+ - p_L$ positive but arbitrarily close to zero we can therefore make the proportion of individuals participating arbitrarily small. Define θ^C such that the proportion of individuals with $\theta < \theta^C$ is the same as the proportion of individuals with $\theta > \theta^*$, and define p^C such that $\theta^+ = \theta^C$ when $p^+ = p^C$. For any p' in the range $p_L < p' < p^C$ the proportion of individuals participating will be strictly lower than the proportion participating at $p' = 0$. QED.

An important feature to note about this equilibrium is that it is a single equilibrium in which individuals' strategy depends on the price announced. It should not be interpreted in terms of the existence of multiple equilibria (one for each of the prices announced), the possibility of which is a distinct question.

Figure 3 illustrates the equilibrium with crowding-out. The two thick lines show the benefits of participating (shallow line) and not participating (steep line); reductions in p^+ mean that θ^+ can be set arbitrarily close to θ_L.

So to summarise, Type Discontinuity is observed in the signalling game, but Payment Discontinuity is not. The reason is that for Payment Discontinuity to be observed (in this setting) requires participation itself (rather than the price of participation) to act as a signal of an individual's type. Nevertheless, in a screening game in which price signalling plays no role, Payment Discontinuity is observed in an entirely intuitive way.

4. Concluding remarks

This paper has suggested that a qualitative and discontinuous difference between gifts and sales, or free participation in civic activities and participation at a price, can emerge from signalling behaviour between individuals even without discontinuity in individuals' types. It is the result of the fact that individuals like to associate with others, and that society's matching processes tend to associate like with like.

It will be evident therefore what answers the model gives to the two pairs of situations set out at the end of section 1 above. A politician might be thought to wish to signal himself as someone who derives large benefits from performance of his public duties, and who by virtue of this signal would expect to spend time in the future among other such people. By asking for a small bribe he effectively signals himself as someone whose benefits from fulfilling his public duties are small – substantially smaller than they might otherwise

have been estimated to be. A small bribe thereby causes him a significant reputational loss. If he asks for a bribe at all, it will be a large one.

Similarly, someone who asks for a small price for donating their blood thereby reveals themself to be at best a reluctant altruist. To the extent that donating blood signals this fact to others, the reluctant altruist may thereby be substantially worse off.

Crowding-out of civic motivation by price incentives may also occur, though it does not necessarily occur under the same conditions. Specifically, it occurs when the participation decision itself (rather than the announcement of the price at which one is willing to participate) serves as a signal of an individual's type. One useful empirical implication of the result is that it implies that crowding-out is less likely to be observed when the context of the decision allows individuals many other means of signalling their commitment to civic virtue other than merely refusing to participate when a positive fee is offered[11].

There remain many interesting questions for further research, including the experimental testing of the results (the model implies that the observation of both discontinuity and crowding-out should be quite sensitive to the nature of the signalling opportunities open to subjects).

[11] This may provide an explanation for the discrepancy in the results of Frey et.al. (1996) between the situation in which respondents were asked to state their willingness to accept payment for a waste disposal facility, and the situation in which they voted in secret ballot on whether or not to accept an actual offer. Under a secret ballot the opportunity to use a participation decision for signalling purposes was very limited.

One final observation is that the phenomena explored in this paper are by no means exotic or pathological. Most of us spend most of our time in association with others, with whom our interactions are not governed entirely or even mainly by either market relations or explicit reciprocity. We do many things for which the immediate return is not calculated, but we value associating with those we like or admire. Understanding the difference between explicitly reciprocal interaction and implicit association is an important task for any satisfactory theory of social life.

Bibliography

Acemoglu, D. (1998): "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence", NBER working paper no. 6658.

Arrow, K.J. (1972): "Gifts and Exchanges", *Philosophy and Public Affairs*, 1(2), 343-62.

Bénabou, R. (1994): "Human Capital, Inequality, and Growth: A Local Perspective", *European-Economic-Review*; 38(3-4), April 1994, pages 817-26.

Bénabou, R. & J. Tirole (2003): "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*.

Bessen, J. (2001): "Open source software: free provision of complex public goods", Research on Innovation, mimeo, dowloadable from: http://www.researchoninnovation.org/opensrc.pdf

Deaton, A. (1995): "Inequality within and between Households in Growing and Aging Economies", Quibria,-M.-G., ed. Critical issues in Asian development: Theories, experiences and policies. Hong Kong; Oxford and New York: Oxford University Press for the Asian Development Bank, 1995, pages 18-38.

Dow, W. & T. Philipson (1996): "An Empirical Examination of the Implications of Assortative Matching on the Incidence of HIV", *Journal-of-Health-Economics*; 15(6), December 1996, pages 735-49.

Fehr, E. & A. Falk (2001): "Psychological Foundations of Incentives", University of Zurich, mimeo.

Frey, B. & R. Jegen (2001): "Motivation Crowding Theory", *Journal of Economic Surveys*, 15, 589-611.

Frey, B., F. Oberholzer-Gee & R. Eichenberger (1997): "The Old Lady Visits Your Backyard: a tale of morals and markets", *Journal of Political Economy*, 104 (6), 1297-1313.

Frey, B. & F. Oberholzer-Gee (1997): "The Cost of Price Incentives: an Empirical Analysis of Motivation Crowding-Out", *American Economic Review*, 87 (4), 746-755.

Gann, J. (2001): "Hey Brother Can You Spare a Kidney?", University of Toulouse, mimeo.

Ghatak, M. (1999): "Group Lending, Local Information and Peer Selection", *Journal of Development Economics*; 60(1), October 1999, pages 27-50.

Gneezy, U. & A. Rustichini (2000a) : "Pay Enough or Don't Pay At All", *Quarterly Journal of Economics*, 115(2), 791-810.

Gneezy, U. & A. Rustichini (2000b): "A Fine is a Price", *Journal of Legal Studies*, 29, 1-17.

Kahneman, D. & A. Tversky (1990): "Rational Choice and the Framing of Decisions", in Cook,-Karen-Schweers; Levi,-Margaret, eds. *The limits of rationality*. Chicago and London: University of Chicago Press, 1990, pages 60-89.

Kremer, M.: "The O-ring theory of economic development", *Quarterly Journal of Economics*, 108(3), 551-575.

Kremer, M. & E. Maskin (1996): "Wage Inequality and Segregation by Skill", NBER working paper no. 5718.

Lerman, R. (1996): "The Impact of the Changing US Family Structure on Child Poverty and Income Inequality", *Economica*; 63(250), Suppl. 1996, pages S119-39.

Mailath, R., L. Samuelson & A. Shaked (2000): "Endogenous Inequality in Integrated Labor Markets with Two-Sided Search", *American-Economic-Review*; 90(1), March 2000, pages 46-72.

Sen. A.K. (1977): "Rational Fools: a critique of the behavioral foundations of economic theory" in Hamouda,-Omar-F.; Rowley,-J.-C.-R., eds. *Paradoxes, ambiguity and rationality. Elgar Reference Collection. Foundations of Probability Econometrics and Economic Games*, vol. 2. Cheltenham, U.K. and Lyme, N.H.: Elgar; distributed by American International Distribution Corporation Williston, Vt., 1997, pages 181-208.

Shimer, R. & L. Smith (2000): "Assortative Matching and Search", *Econometrica*; 68(2), March 2000, pages 343-69

Solow, R. (1971): "Blood and Thunder", *Yale Law Journal*, 80(2), 170-83.

Titmuss, R. (1970): *The Gift Relationship*, London, Allen & Unwin.

Weiss, Y. (1993): "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens Upon Divorce", Tel Aviv Foerder Institute for Economic Research Working Paper 15/93, December 1993, pages 74.

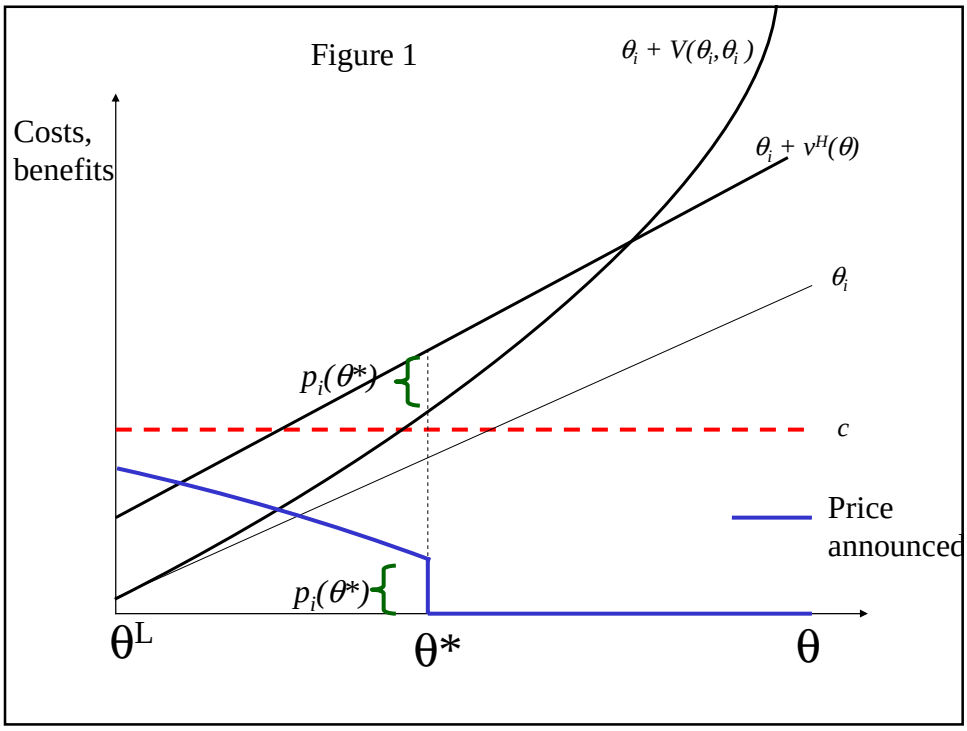
Figure 1
$\theta_i + V(\theta_i,\theta_i)$
Costs, benefits
$\theta_i + v^H(\theta)$
θ_i
$p_i(\theta^*)$
c
Price announced
$p_i(\theta^*)$
θ^L
θ^*
θ

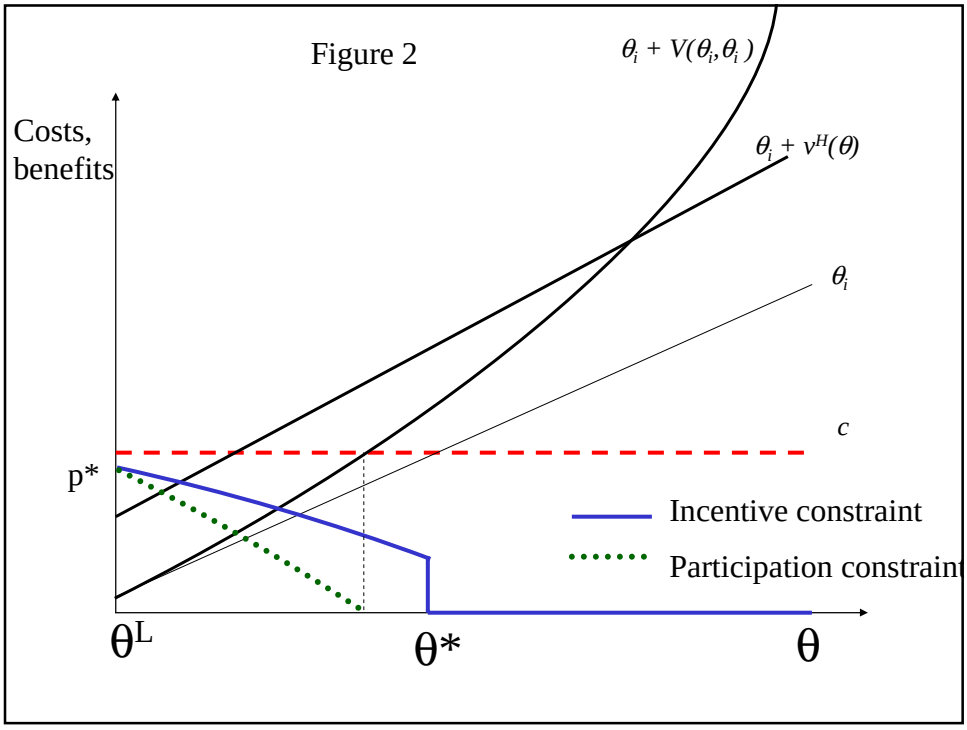
Figure 2
$\theta_i + V(\theta_i,\theta_i)$
Costs, benefits
$\theta_i + v^H(\theta)$
θ_i
c
p*
Incentive constraint
Participation constraint
θ^L
θ^*
θ

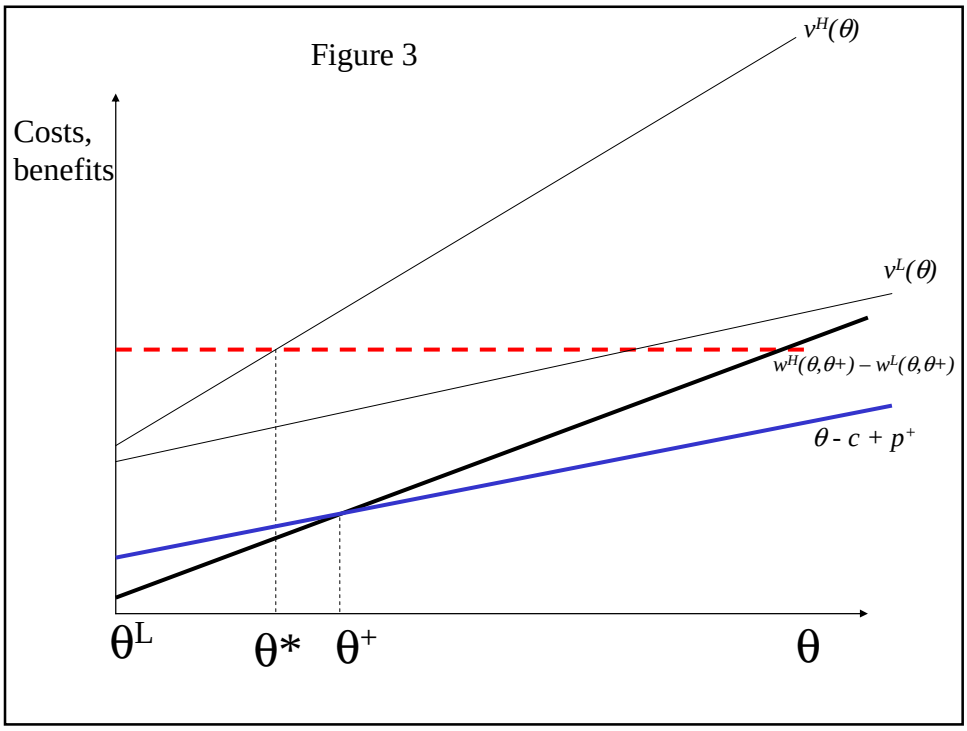
Figure 3
Costs, benefits
$v^H(\theta)$
$v^L(\theta)$
$w^H(\theta,\theta+) - w^L(\theta,\theta+)$
$\theta - c + p^+$
θ^L
θ^*
θ^+
θ