

Blood, Bribes and the Crowding-Out of Altruism by Financial Incentives

Paul Seabright
IDEI, Université de Toulouse-1 & CEPR¹

8th February 2002

Abstract

This paper develops a model of two phenomena that have been claimed by psychologists and experimental economists to constitute important exceptions to the standard economic theory of human motivation. The first (the “discontinuity hypothesis”) 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 claims 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 (the crowding-out hypothesis”) 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. 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. The discontinuity hypothesis, but not crowding-out, is observed in a signaling game in which individuals announce the prices at which they will perform a civic action. That is, even though the distribution of types is continuous, announced prices are discontinuous, with some individuals announcing zero prices, and others announcing strictly positive prices, and with nobody announcing very low positive prices. Crowding-out, by contrast, 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 is higher than when rewards are positive but small.

Keywords : crowding effect, intrinsic motivation, assortative matching, economic psychology

JEL codes: A12, C70, D10, D60, H41, J22

¹ I am grateful to Emmanuelle Auriol, Ernst Fehr, Jennifer Gann 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 fee make them less willing to do something they would gladly have undertaken for free?

These questions are part of a much wider set of issues about the acceptability, either legally or ethically, of voluntary transactions between consenting adults. Social life is thick with prescriptions and proscriptions that govern the many transactions people can undertake, and one of the ways in which economic theory can cast light on such rules is to examine whether they can be given a rationale in terms of calculative reasoning. This does not imply that calculative reasoning provides an adequate causal explanation for how such rules could have come about (sometimes it does, sometimes it doesn't), but it helps to distinguish those rules that are stable under calculative examination from those that are not. When one examines them closely, some rules, however well entrenched historically, just seem to have no point.

In this spirit, the purpose of this paper is to examine whether there is a point to certain rules that either prohibit legally or stigmatize ethically² some kinds of voluntary transaction among consenting adults, not because the behaviour itself is considered undesirable, but because it is thought that it should not be made subject to a transaction. Many kinds of action that would be considered intrinsically acceptable or even admirable are considered "tainted" or repugnant when undertaken for a reward. An example is the donation of a kidney. Many individuals donate kidneys to those requiring them for transplants, but in almost all cases the two individuals are close relatives and the donation is unilateral. Gann (2001) 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 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 because the sale would violate the US National Organ Transplant Act, passed in 1984, which prohibits the sale of human body parts. The Act itself is silent regarding the reason for the prohibition, but the language used during the congressional hearings debating it leaves no doubt as to the motivation of its sponsors: '*if...organs of living people should be offered for buying and selling, then I think this would represent a major degradation for humankind. ...this*

² The distinction is important: some laws prohibit behaviour that would not be considered unethical if it were not illegal (such as selling a car above a certain age or opening a shop after a certain hour in the evening). Some ethical systems stigmatize behaviour which they do not imply should be illegal (such as adultery or the expression of certain political opinions). Legal prohibitions and ethical maledictions therefore need a distinct justification even when referring to the same kind of behaviour.

“free-market” sale of an individual’s organs is morally offensive and ethically indefensible’.

Indeed, the moral repugnance which we feel at the thought of selling part of our bodies appears to be near-universal. The UN and the European Union have, respectively, encouraged and instructed their member countries to prohibit the sale of human body parts. The World Health Organization has interpreted the Universal Treaty on Human Rights as prohibiting the sale of human organs.”

It is often somewhat unclear whether the objection is to a particular kind of behaviour featuring in a transaction at all, or more specifically to the monetary character of the transaction. This is a subtle issue because it has long been pointed out in anthropological writing on gift-giving that gifts are typically not altruistic gestures performed with no hope of a return, but are often part of elaborate reciprocal structures. Marcel Mauss wrote in 1925 in *The Gift* that in “archaic societies” gifts are “in theory voluntary, in reality given and returned obligatorily”³, and that gift exchange constitutes the major mechanism of circulation of goods in such societies. Nevertheless, he argued that they were different from monetary transactions in commercial societies for a variety of reasons, including that the nature of obligation incurred on receipt of a gift was not determined wholly or even primarily by the nature of the goods received but owed a great deal to the status and other social and affective links between donor and recipient. At times he seemed to believe that the complexity of these links was under threat from market transactions, and that modern societies would reach a phase “of purely individual contract, of the market where money circulates, of sale proper, and above all of the notion of price reckoned in coinage weighed and stamped with its value”. At other times he rejoiced in the fact that “a considerable part of our morality and our lives themselves are still permeated with this same atmosphere of the gift, where obligation and liberty intermingle. Fortunately, everything is still not wholly categorized in terms of buying and selling. Things still have sentimental as well as venal value”.

Given that gifts entailed reciprocal obligations it is not entirely clear how they escaped being considered venal, though it appears to have been part of the strategy of reciprocity that its venality should be dissimulated, or at least softened. Nathalie Davis has shown in her study of *The Gift in Sixteenth Century France* (2001) that multiple conceptions of gift-giving have coexisted for centuries, with complex rules of reciprocity cloaked by more high-minded ideals of “gratuitous and non-calculating values”, and of a liberality whose only reward was the gratitude of the recipient. Seen in this light, what makes financial transactions more offensive than others might be considered not their venality per se, but the crassness and lack of subtlety with which that venality is communicated. Selling a good is like giving it to someone and simultaneously reminding them of their debt to you, a debt that (it is implied) sensitive and intelligent individuals ought to have been aware of without a reminder, and might resent being brought so vulgarly to their attention.

Gifts may indeed have been a more delicate and refined way than commercial trade to undertake economic exchange. Like many delicate and refined arts, therefore,

³ Mauss (2000).

they gave an advantage to those who had the aristocratic privileges that made their mastery easier to acquire. As Davis's study makes clear, the rich and privileged could sometimes give away a great deal, but they rarely gave away real control over important resources. She writes: "gifts opened channels of communication across boundaries of status and literacy. They gave expression to the highly strained but genuine reciprocity between unequals in the social and economic order". She hardly needs to add that they did nothing to change that order.

In the remainder of this paper I shall ignore the distinction between monetary and non-monetary rewards and ask whether we can understand why certain kinds of behaviour are considered intrinsically unsuitable for being performed in exchange for an explicit reward. Two main claims have been made in the literature, and as will be seen these are quite distinct even though they are often confused. The first is that there is an intrinsic discontinuity between performing actions for a reward, however small, and performing them for free; we may see some individuals willing to perform them 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. The second is that offering a reward for actions that some people would perform for free may actually reduce the willingness of some individuals to perform the actions at all. Call these two claims the "discontinuity hypothesis" and the "crowding-out hypothesis" respectively.

The discontinuity hypothesis and the crowding-out hypothesis are quite distinct from each other. The former is a claim about the distribution of values of the willingness to accept payment in return for performing some action (WTA) across the population. It claims 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. This claim is compatible with an entirely orthodox interpretation of WTA.

The crowding-out hypothesis, by contrast, is a hypothesis about the nature of the willingness to accept of a given individual, and casts doubt on the very concept of WTA according to which higher associated payments always make a given option more attractive. It implies that some or even most individuals are willing to perform actions for free, unwilling to do so for a small payment but (perhaps) willing to do so for a large one.

In section 2 I briefly survey evidence for the crowding-out hypothesis, and discuss explanations that have been advanced for such phenomena. 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 discontinuity hypothesis a part of the explanation for the crowding-out hypothesis). In section 3, therefore, I propose a simple economic model to explain both the discontinuity hypothesis and the crowding-out hypothesis as arising from more elementary motivations. I also show that the circumstances that favour the discontinuity hypothesis do not necessarily favour the crowding-out hypothesis, which is observed under somewhat more restrictive circumstances. Section 4 summarises and concludes.

In contrast to the crowding-out hypothesis, the discontinuity hypothesis as such has been subjected to much less careful investigation. The evidence for it is more casual

and anecdotal. 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.

2. Markets and the “crowding-out” of reciprocity

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 discontinuity of WTA across the population. 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. 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 motivated⁴, 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 she did not have that particular duty⁵). However, there is something

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.

⁵ A similar point underlies Sen’s famous distinction between “sympathy” and “commitment” (Sen, 1977).

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 phenomenon⁶), 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 consequences⁷. 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 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 are⁸. Now it might be thought

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.

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, 2002). I discuss this possibility briefly in Seabright (2001) in relation to the theory of consumer

that this 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 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

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 find out about their own characteristics through consumption decisions: I do not know how fit I am until I go to the gym, I do not know whether I like caviar till I try it, and so on".

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.

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.

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. An individual i has a twice continuously differentiable utility function $U_i(m_i, c_i, b_i, \mathbf{q}_i, \mathbf{q}_j)$, 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
- \mathbf{q}_i is the individual’s type
- \mathbf{q}_j is the type of the individual with whom she is matched.

We suppose that the utility can be written as follows:

$$(1) \quad U_i(m_i, c_i, b_i, \mathbf{q}_i, \mathbf{q}_j) = m_i - c_i + b_i(\mathbf{q}_i) + V_i(\mathbf{q}_i \mathbf{q}_j)$$

with $\partial b_i(\mathbf{q}_i)/\partial \mathbf{q}_i > 0$, $\partial V_i(\mathbf{q}_i \mathbf{q}_j)/\partial \mathbf{q}_i > 0$, $\partial V_i(\mathbf{q}_i \mathbf{q}_j)/\partial \mathbf{q}_j > 0$ and $\partial^2 V_i(\mathbf{q}_i \mathbf{q}_j)/\partial \mathbf{q}_i \partial \mathbf{q}_j > 0$.

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 $V(\cdot)$ from the private activity is a function of the product of the types of 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 action⁹.
- 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 x_i after which participation takes place, then individuals are matched with others according to their expected type conditional on their participation decision.

In both games we look for a Perfect Bayesian Equilibrium. We suppose that there is a distribution of types \mathbf{q} along an interval $[\mathbf{q}^L, \mathbf{q}^H]$, where $\mathbf{q}^L < c < \mathbf{q}^H$. We assume that \mathbf{q}^H is sufficiently high that some individuals will wish to announce zero prices, while \mathbf{q}^L is sufficiently low that some individuals will wish to announce positive prices. We also assume $b_i(\mathbf{q}_i) = \mathbf{q}_i$ for simplicity, 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 $v^H(\mathbf{q}_i)$ as the expected utility of an individual of type \mathbf{q}_i in the second period if she pools with all weakly higher types $\mathbf{q}_j \geq \mathbf{q}_i$, that is, if she is matched at random with one of the set of all individuals with types (weakly) higher than her own. Similarly define $v^L(\mathbf{q}_i)$ as the expected utility of an individual of type \mathbf{q}_i in the second period if she pools with all (weakly) lower types $\mathbf{q}_j \leq \mathbf{q}_i$.

We first show that announced prices in the signalling game are discontinuous in \mathbf{q} :

Proposition 1 (Discontinuity):

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

⁹ This means we do not model the matching process explicitly but draw on the standard findings in the literature.

Proof:

I first assume, and later prove, that all and only the individuals with a type above some threshold value \mathbf{q}^* announce zero prices. In effect they announce that they are willing to participate in the activity for free (as a “gift”). Then in a PBE each will be matched in the second period with an individual whose expected is the mean type of all those whose type is higher than \mathbf{q}^* . The expected utility of such an individual is:

$$(2) \quad E[U_i(m_i, c_i, b_i, \mathbf{q}_i, \mathbf{q}_j) | p_i = 0] = m_i - c + \mathbf{q}_i + \int_{\mathbf{q}^*}^{\mathbf{q}^H} V(\mathbf{q}_i, \mathbf{q}_j) d\mathbf{q}_j$$

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 $\mathbf{f}_i(p)$ as follows:

$$(3) \quad V(\mathbf{q}_i, \mathbf{f}_i(p)) \equiv E[V(\mathbf{q}_i, \mathbf{q}_j) | p_j = p]$$

Now write her expected utility from announcing a price p_i as follows:

$$(4) \quad E[U_i(m_i, c_i, b_i, \mathbf{q}_i, \mathbf{q}_j) | p_i > 0] = m_i + p_i - c + \mathbf{q}_i + V(\mathbf{q}_i, \mathbf{f}_i(p_i))$$

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

$$(5) \quad -1 = \frac{\partial V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{f}_i} \cdot \frac{\partial \mathbf{f}_i}{\partial p_i}$$

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

$$(6) \quad \frac{\partial^2 V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{f}_i \partial \mathbf{q}_i} > 0$$

from which we know that equation (5) will not be satisfied by the same value of p_i at any two different values of \mathbf{q}_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:

$$(7) \quad E[U_i(m_i, c_i, b_i, \mathbf{q}_i, \mathbf{q}_j) | p_i > 0] = m_i + p_i - c + \mathbf{q}_i + V(\mathbf{q}_i, \mathbf{q}_i)$$

Define individual of type \mathbf{q}^* 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 that positive price is determined by the following equation:

$$(8) \quad m_i - c + \mathbf{q}^* + \int_{\mathbf{q}^*}^{\mathbf{q}^H} V(\mathbf{q}^*, \mathbf{q}_j) d\mathbf{q}_j = m_i + p_i - c + \mathbf{q}^* + V(\mathbf{q}^*, \mathbf{q}^*)$$

and thus

$$(9) \quad \int_{\mathbf{q}^*}^{\mathbf{q}^H} V(\mathbf{q}^*, \mathbf{q}_j) d\mathbf{q}_j = p_i + V(\mathbf{q}^*, \mathbf{q}^*)$$

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

It remains to be shown, first, that if an individual of type \mathbf{q}_i announces a zero price, then all individuals of type $\mathbf{q}_k > \mathbf{q}_i$ also do so (the assumptions on \mathbf{q}^L and \mathbf{q}^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

$$(10) \quad \frac{\partial^2 \mathbf{f}_i(p_i)}{\partial p_i \partial \mathbf{q}_i} < 0$$

Thus if any individual of type \mathbf{q}_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 \mathbf{q} 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 \mathbf{q}^* , 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 \mathbf{q}^* 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 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 \mathbf{q}^* , 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 $p^* - c + \mathbf{q} + V(\mathbf{q}, \mathbf{q}) = 0$. Proposition 2 shows that under separation, rents to participation are strictly increasing in \mathbf{q} , so that all participants of higher type than the marginal participant receive strictly positive rents.

Proposition 2 (Rents under separation increasing in \mathbf{q}):

$$R(\mathbf{q}) \circ \mathbf{q} + V(\mathbf{q}, \mathbf{q}) + p_i(\mathbf{q}) - c \text{ is increasing in } \mathbf{q}.$$

Proof:

From equation (5) we know that in equilibrium

$$(11) \quad \frac{\partial p_i}{\partial \mathbf{f}_i} = -\frac{\partial V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{f}_i}$$

Totally differentiating $V(\cdot)$ yields

$$(12) \quad \frac{\partial V}{\partial \mathbf{q}_i} = \frac{\partial V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{q}_i} + \frac{\partial V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{f}_i} \frac{\partial \mathbf{f}_i}{\partial \mathbf{q}_i}$$

Substituting (11) in (12) yields:

$$(13) \quad \frac{\partial p_i}{\partial \mathbf{q}_i} = \frac{\partial V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{q}_i} - \frac{dV}{d\mathbf{q}_i}$$

Totally differentiating the expression for rent, substituting (13) and using the envelope theorem yields:

$$(14) \quad \frac{\partial R(\mathbf{q}_i)}{\partial \mathbf{q}_i} = 1 + \frac{\partial V(\mathbf{q}_i, \mathbf{f}_i(p_i))}{\partial \mathbf{q}_i} > 0$$

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 \mathbf{q} 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 \mathbf{q} than their direct gains from participation.

Proposition 3 (Crowding Out):

If $v^H(\mathbf{q}) - v^L(\mathbf{q}) - \mathbf{q}$ is an increasing function of \mathbf{q} , there exists a PBE of the screening game in which at a price $p! = 0$ all and only individuals with $\mathbf{q}_i \geq \mathbf{q}^*$ participate in the civic activity, while for a range of strictly positive prices $p! > 0$ there exist threshold values of \mathbf{q} such that only individuals with values lower than the threshold participate, and the proportion of individuals participating in the civic activity is strictly lower than at $p! = 0$.

Proof:

At price $p! = 0$ all and only individuals with $\mathbf{q}_i \geq \mathbf{q}^*$ will participate by definition of \mathbf{q}^* . Now define a price p_L which is the lowest price at which the lowest-value type \mathbf{q}_L will participate, and consider a price $p^+ > p_L$. Define \mathbf{q}^+ as the value of \mathbf{q}_i at which individual i is just indifferent between participating and not participating, conditional on believing that only individuals with values $\mathbf{q}_i < \mathbf{q}^+$ will participate. Then we can set the benefits of not participating equal to the benefits of participating, as follows:

$$(15) \quad v^H(\mathbf{q}^+) = v^L(\mathbf{q}^+) + \mathbf{q}^+ - c + p^+$$

which implies

$$(16) \quad v^H(\mathbf{q}^+) - v^L(\mathbf{q}^+) - \mathbf{q}^+ = p^+ - c$$

From the fact that $v^H(\mathbf{q}) - v^L(\mathbf{q}) - \mathbf{q}$ we can confirm that individuals with $\mathbf{q} \geq \mathbf{q}^+$ will indeed participate while those with $\mathbf{q} < \mathbf{q}^+$ will not, and thus that this strategy constitutes a PBE. This also implies that \mathbf{q}^+ 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 \mathbf{q}^C such that the proportion of individuals with $\mathbf{q} < \mathbf{q}^C$ is the same as the proportion of individuals with $\mathbf{q} > \mathbf{q}^*$, and define p^C such that $\mathbf{q}^+ = \mathbf{q}^C$ when $p^+ = p^C$. For $p_L < p! < p^C$ the proportion of individuals participating will be strictly lower than the proportion participating at $p! = 0$. QED.

Figure 3 illustrates. The two thick lines show the benefits of participating (shallow line) and not participating (steep line); reduction in p^+ mean that \mathbf{q}^+ can be set arbitrarily close to \mathbf{q}_L .

So to summarise, the discontinuity hypothesis is observed in the signalling game, but crowding-out is not. The reason is that for crowding-out 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, crowding-out 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 as the result of signalling behaviour between individuals even when there is no intrinsic discontinuity in individual's 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 themselves 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 offered¹⁰.

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).

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.

¹⁰ 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.

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 (2002) : “Intrinsic and Extrinsic Motivation”, University of Toulouse, mimeo.
- Bessen, J. (2001): “Open source software: free provision of complex public goods”, Research on Innovation, mimeo, downloadable from:
<http://www.researchoninnovation.org/opensrc.pdf>
- Davis, N. (2001): *The Gift in Sixteenth Century France*, Madison, University of Wisconsin Press.
- 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.
- Mauss, M. (2000): *The Gift: the form and reason for exchange in archaic societies*, translated by W.D. Hall, New York, W.W. Norton & Company.
- 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

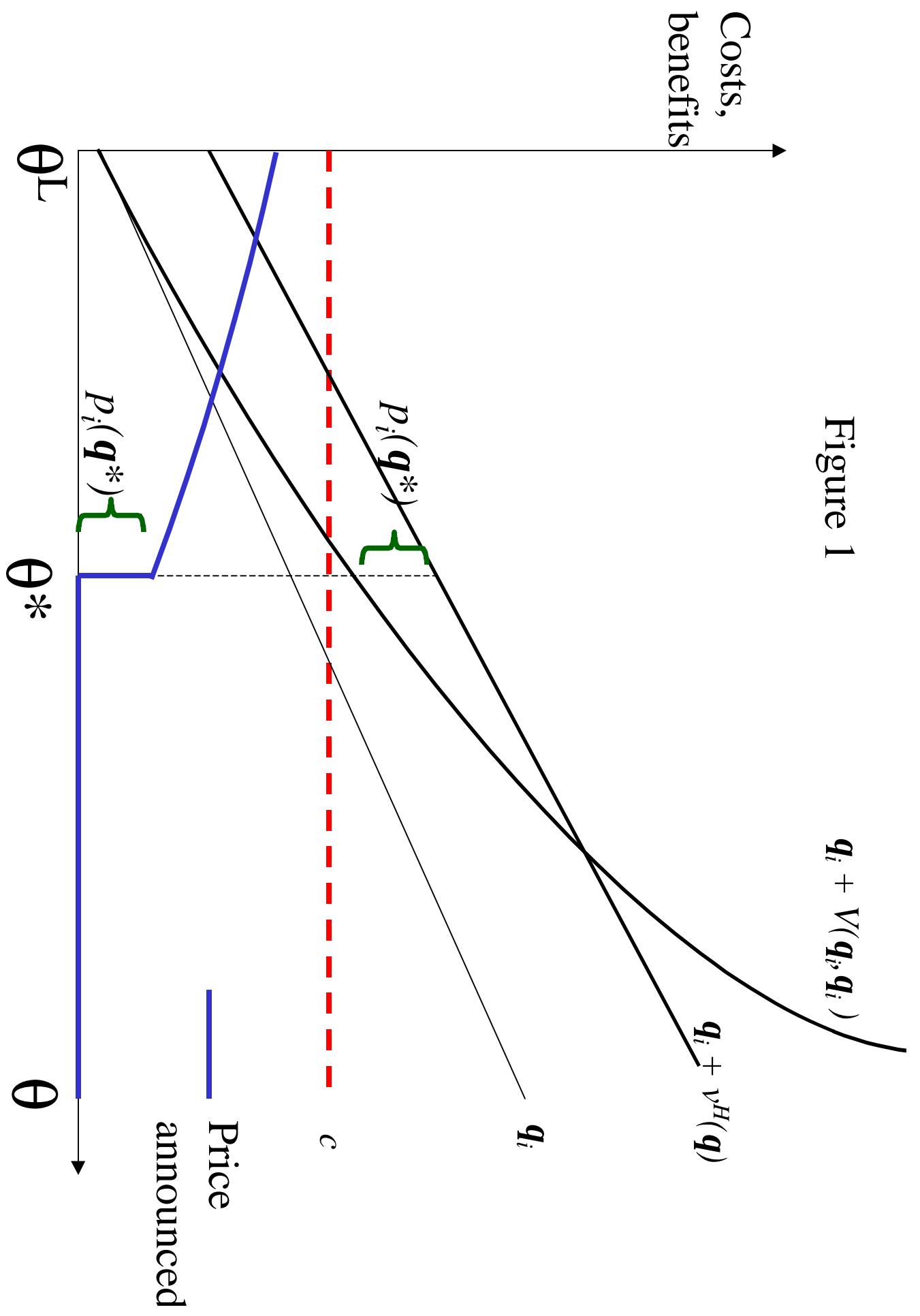


Figure 2

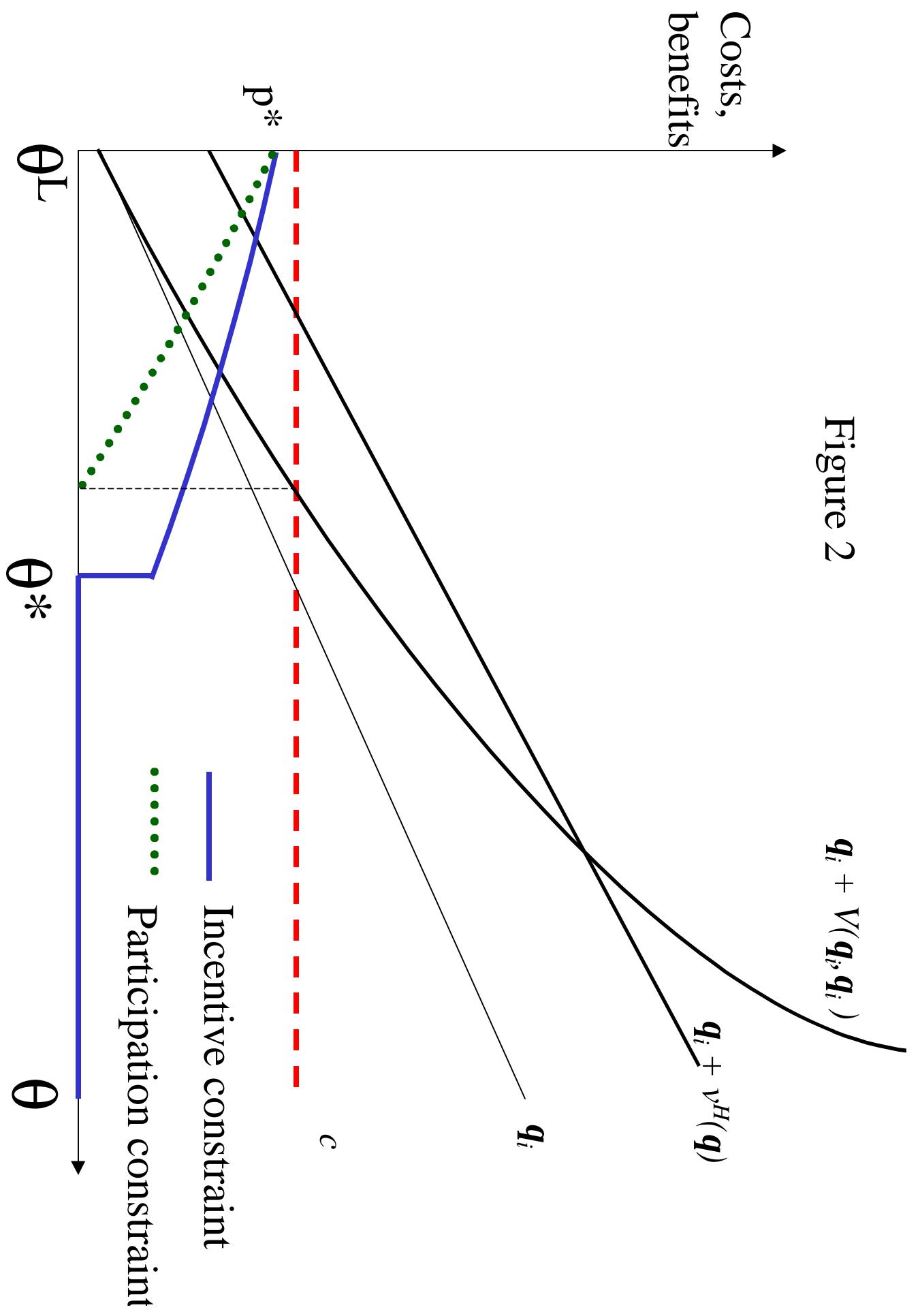


Figure 3

