THE PERSISTENCE OF GOODNESS

Ashok S. Guha[*] and Brishti Guha[**]

Abstract

Experimental evidence and economic examples like Basu's (1984) taxi driver problem illustrate that many people are honest (or 'good') even when beyond the reach of the law, and without repeated interactions or reputation effects. We provide game theoretic underpinnings of the level of goodness in a population. For appropriate parameter ranges, a certain level of good *behaviour* will emerge as an evolutionarily stable equilibrium: virtue will not be driven out of the population, even in a Darwinian world of the survival of the fittest. The long-run equilibrium proportion of good behaviour is independent of the level of intrinsic goodness.

JEL Classifications: C73, C72, D03, D82

1. Introduction

One of the major puzzles of contemporary economics, highlighted by casual observation and confirmed by behavioral experiments, is that people often exhibit a degree of honesty or philanthropy or sense of fairness that far exceeds the predictions of conventional game-theoretic rationality. Standard explanations of such behavior run in terms of fear of the law or of retaliation by one's partner where a game is indefinitely repeated between two individuals or of loss of reputation that may hinder one's future dealings with others. However, people often seem to act contrary to their narrow self-interest in situations beyond the reach of the law, and where chances of a repeat encounter are minimal and anonymity rules out any reputation effect. BASU [1984] has illustrated this observation by asking why we usually pay taxi-drivers: we could after all simply walk away in most cases after reaching our destinations. We are unlikely to meet the taxi-driver again, and he cannot ruin our reputations since he cannot readily identify us to others. In a city like Delhi (rather than New York), one can in fact turn Basu's question on its head and ask why we board taxis in the expectation of being safely delivered to our destinations instead of being driven to a lonely spot and stripped of our possessions when the taxi-driver can evade the law by the simple expedient of a fake number plate.

[*] School of International Studies, Jawaharlal Nehru University New Delhi 110067. E-mail : ashoksanjayguha@gmail.com.

[**]Corresponding author. Department of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. E-mail : bguha@smu.edu.sg. Both authors would like to thank Kaushik Basu for feedback. We also thank two anonymous referees.

We could of course assert that goodness is an innate human trait – as HAUSER [2006] has persuasively argued, morality is part of our genetic heritage – and that knowledge of the distribution of unselfish virtue in the population enables us to trustingly conclude many transactions that we would otherwise shun in a society of universal suspicion and iron-clad contracts. Honesty and trust are what makes the world go round, or at least saves the taxi business from extinction. GHATAK AND INGERSENT [1984], for instance, stress the role of honesty and trust in determining credit transaction cost in informal money markets.

Hauser's answer unfortunately amounts to a mere restatement of Basu's Paradox. The really interesting question is how virtue can survive in a competitive environment – since opportunists stand to do better than the virtuous and can squeeze the latter out of the market. This is an exact analog of the question in evolutionary biology that bedeviled the protagonists of group selection: if some individuals exhibit a trait that fosters the welfare and multiplication of their species at the expense of themselves, why shouldn't their lower survival rate lead eventually to their effective eclipse? The analog is worth noting since evolutionary game theory – later a tool much used by economists – originated in the application of game theory to biological contexts, especially after MAYNARD SMITH (1972) and MAYNARD SMITH AND PRICE (1973) applied game theory to animal conflict. The analog can be best understood in terms of "replicator dynamics" subsequently formalized by TAYLOR AND JONKER (1978), which specified how given phenotypes in a population multiply. According to the replicator dynamics, the fraction of individuals in a population exhibiting a given trait grows in proportion to the extent by which the average fitness of these individuals exceeds the average fitness of the entire population. From this, it is evident that any trait that benefits others in the population, but lowers the individual's fitness on average below the population mean, will die out in time as individuals with this trait can replicate much more slowly than other individuals. Why then do we still observe some individuals with these "altruistic" traits?[1] Economists have adapted the replicator dynamics to their own purposes by assuming that the share of a population playing a given strategy increases for high-payoff strategies and decreases for strategies which yield low payoffs (SAMUELSON [2002]). Extending the analog to the problem we study, we may pose the following question: if virtue is a basic human trait, why shouldn't its role in economic transactions be reduced to insignificance – given that opportunists can on average extract higher payoffs than the virtuous? Indeed, why should virtue remain a basic human trait in a Darwinian world of survival of the fittest?

[1] An interesting possibility that Taylor and Jonker do not consider is what would happen if relatively more robust *groups,* rather than – or in addition to – relatively more robust *individuals*, were able to replicate faster than average. In this case it is plausible that a trait which lowers individual fitness but benefits the group might still survive if the group's increased rate of replication were fast enough to offset the presumably lower than average replication rate of the altruistic *individual*.

These are questions that possibly explain why, despite massive evidence of the innate goodness of a fraction of society (for instance ANDERSON [2000], GNEEZY [2005], HAUSMAN AND MCPHERSON[1996]) the dominant paradigm in economics in general and game theory in particular remains that of the opportunistic maximizer. Perhaps, it may be felt, innate goodness is an illusion: good behaviour is simply the outcome of behavioral constraints that we do not fully appreciate. Or, perhaps it is simply transient, a disequilibrium phenomenon on its way to extinction. These questions need to be explored and answered before economists will be prepared to incorporate the possibility of honesty in their theorizing.

MYERSON [2004] has in fact already produced a game-theorist's solution of the taxi-driver riddle. He argues that taxi-drivers demand and passengers pay fares at the established rate without allegiance to any principles of morality; they do so because payment at this rate for a taxi-ride is a convention that acts as a focal point for all taxi-drivers, passengers and, indeed, pedestrians, coordinating their behavior in a situation in which multiple equilibria exist. This is a solution that may not fully satisfy everyone. The coordination of behavior achieved through the convention is quite imperfect. While the majority of drivers and passengers follows it, a significant minority does not. Quite a few drivers do in fact rob passengers, and a not-entirely-negligible number of passengers escape without payment. Ideally, one would want an equilibrium in which different types of behavior coexist.

MUKHERJEE [1984] rejects game-theoretic analyses of the problem and proposes instead a solution in terms of bounded rationality, a rule of thumb adopted by the taxi-rider in the face of uncertainty, presumably about the possible combined impact of the strong arm of the taxi-driver and the long arm of the law in the event of non-payment.

This paper seeks to take some tentative steps towards a game-theoretic answer different from Myerson's. We suggest that a minimal level of good behaviour is essential for bad behaviour to be profitable, that long run processes drive a society towards a stable evolutionary equilibrium at this level and that, since bad behaviour at this point is no more profitable than good, bad people cannot compete the good people out of the market. In essence, our model is the well-known hawk-dove model of evolutionary game theory, and our innovation is to show how it fits into the taxi-driver riddle and, more generally, into the problem of the persistence of virtue.

We define good behaviour by an individual as activity that increases the payoffs of agents with whom he transacts. Intrinsic goodness implies that one persists in such activity even if it reduces one's own payoffs. 0pportunists on the other hand always seek to maximize their personal payoffs, which may or may not involve good behaviour. Bad behaviour of course reduces the payoffs of others. No

one is intrinsically bad. People are either innately good or opportunists. In what follows, we use the terms 'honesty' and 'cheating' as shorthand for good and bad behaviour.

2. A Model of Unilateral Moral Hazard

Visualize a two-good economy, peopled by two groups of individuals. Individuals in these two groups differ in their initial stocks of goods, creating opportunities for profitable trade between members of the two groups. All individuals in any one group are identical in tastes and endowment (though not in morality). However, any transaction between members of separate groups is beset by moral hazard.

We assume in our initial model that this hazard is unilateral – that only the second party has the opportunity to cheat. The first party may abandon the transaction, but if it does undertake the transaction, has no opportunity to cheat. The second party has the additional option of cheating, which offers a higher payoff than honesty. We could interpret this in terms of a sequential game where the first movers are group 1 individuals. Group 2 individuals respond to group 1 individuals' move. For example, a group 1 individual could be a trader who decides whether or not to extend credit to a customer drawn from a large population with whom he is unlikely to have repeat dealings. The group 2 individual would be the customer, who, if he gets the trade credit, may then either repay (if he is honest) or default (if he is opportunistic). Of course, it is easy to see that if all group 2 individuals were opportunistic, the game described above would reduce to a standard hold-up problem, so that no group 1 individuals would extend trade credit. In our setup, however, a fraction λ of the population (of either group) always acts honestly while the rest are opportunistic.

The value of λ is common knowledge and one believes that one's partner in any transaction is drawn at random from this population. For our purposes, λ, the initial proportion of intrinsically honest agents, is exogenous (we explore long-run evolution of honest behavior later in the paper). It could be a function of history: a society beset by wars, invasions or a history of mutual distrust may have a low λ ; one with a relatively peaceful history with successful mutual co-operation may have a higher one. Alternatively, λ could be a genetic trait or a joint product of genes and history. At the outset of the game, the players decide whether to transact. Those who decide against transacting, exit the game. Those who continue transact honestly or cheat and realize their payoffs.

Since opportunists in the second group can cheat with impunity, they always do so while the honest ones do not. Agents in the first group have no opportunity to cheat; their payoffs amount to $\alpha_1 > 0$ if their partners act honestly and $\alpha_2 < 0$ if their partners cheat. Their expected income from the transaction

$$\lambda\alpha_1 + (1-\lambda)\alpha_2 \geq 0$$

$$\text{iff } \lambda \geq -\alpha_2/[\alpha_1 - \alpha_2] = |\alpha_2|/[\alpha_1 + |\alpha_2|] = \underline{\lambda}$$

Below this honesty threshold, such transactions do not occur. Above it, they do and opportunists in the second group always cheat and capture a surplus over honest folk.

There are however two factors that may erode or disrupt such an equilibrium. We have assumed up to this point that there is no competition between honest and opportunistic individuals. If however they do compete for a scarce resource, this will drive the intrinsically honest out of the market. All group 2 agents who remain in the market will cheat; since their potential partners expect this they do not enter the market at all. We prove this later in the more general context of bilateral moral hazard and defer further consideration of the issue at this point.

Even if there is no competition between honest folk and opportunists to disrupt the static equilibrium, the market will eventually collapse in a dynamic model that allows for the long run evolution of the honesty coefficient λ. We model this process in an overlapping generations framework in which each generation lives for two periods. Individuals are not economically active in the first period: they simply learn from observation and parental and cultural example and precept. In the second period, they produce, transact and earn while cheating or acting honest; they also reproduce and bring up their children. At the end of the second period, they die. V_H is the expected payoff from honesty and V_C that from cheating.

We assume plausibly that evolutionary fitness is increasing in utility levels. To see why this assumption (a standard one made in evolutionary game theory) is reasonable, suppose that in contrast, fitness were not directly related to utility levels. Then, many people would maximize their utility by following strategies which make them unfit; hence they would die out. Thus, the argument is not that every one necessarily optimizes, but that of the pool of survivors we see in practice, the overwhelming majority will consist of optimizers, as others will have died out. See SAMUELSON [2002] for a similar discussion and justification of this assumption. Each parent produces $(1 + s)$ surviving children. s is a function of parental utility. Thus, an honest parent leaves $(1 + s(V_H))$ surviving children while an opportunist is survived by $(1 + s(V_O))$, where the opportunist's payoff is $V_O = max\ [V_H, V_C]$. Children conform to their parental types unless culturally conditioned to switch types during their childhood: v is the probability of such a switch and is a function of the honesty premium $V_H - V_C$ (or the 'dishonesty premium' $V_C - V_H$) in that period. Then, if H and O are the numbers of intrinsically honest and opportunistic individuals (with the appropriate time-subscripts), the long run dynamics are given by

$$H_{t+1} - H_t = \Delta H_t = s(V_{Ht})H_t + v(V_{Ht} - V_{Ct})O_t \quad if\ V_{Ht} > V_{Ct}$$

$\Delta H_t = s(V_{Ht})H_t - v(V_{Ct} - V_{Ht})H_t$ *if* $V_{Ht} < V_{Ct}$

$O_{t+1} - O_t = \Delta O_t = s(V_{Ht})O_t - v(V_{Ht} - V_{Ct})O_t$ *if* $V_{Ht} > V_{Ct}$

$\Delta O_t = s(V_{Ct})O_t + v(V_{Ct} - V_{Ht})H_t$ *if* $V_{Ht} < V_{Ct}$

Such a dynamical system partially separates the genetic and cultural factors in honesty, represented by the functions $s(\ )$ and $v(\ )$ respectively. The separation however is not complete. As long as cultural influences affect honesty, our assumption that children conform to parental type unless their culture induces a switch involves transmission of acquired characteristics to one's children, a process that cannot be purely genetic.

Now, since for all λ, $V_{Ht} < V_{Ct}$,

$$\Delta H_t/H_t = s(V_{Ht}) - v(V_{Ct} - V_{Ht}),$$

$$\Delta O_t/O_t = s(V_{Ot}) + v(V_{Ct} - V_{Ht})H_t/O_t > \Delta H_t/H_t.$$

The proportion of honest agents in the population falls continuously, so that the honesty coefficient ends up below the threshold $\underline{\lambda}$ and trade breaks down.

A model with unilateral moral hazard does not therefore offer a stable sustained solution to Basu's Paradox.

Bilateral Moral Hazard

Now add to our model the possibility that the first party may also cheat profitably. The two parties to a transaction each have three strategies open to them – to abandon the transaction, to transact honestly or to cheat. If they decide to transact, they must both prepare in advance for the strategy of their choice – so that they are pre-committed to this strategy without prior knowledge of the option selected by their potential partner. Basu's passenger, if he plans to cheat, must pack his gun but no money – and so must the taxi-driver if he has similar designs.

Let α_{ij} and β_{ij} be the surpluses from the transaction earned by the first- and second-group players respectively when the actions of the first and second players are indexed by i and j respectively ($i = h, c$ according as the first player acts honestly or cheats, j likewise for the actions of the second player).

We assume that $\alpha_{ch} > \alpha_{hh} > 0 > \alpha_{hc} > \alpha_{cc}$, $\beta_{hc} > \beta_{hh} > 0 > \beta_{ch} > \beta_{cc}$. Each agent finds the transaction worthwhile if his partner acts honestly, but each finds cheating more rewarding in that event. If his partner cheats, the transaction results in losses which would be maximal if both cheat (since this would lead to mutually

destructive conflict, perhaps a gun-fight in Basu's case); it would be preferable to cut one's losses by bowing out honestly – though it would have been even better not to enter the transaction at all.

The payoffs to mutual honesty α_{hh} and β_{hh} correspond to the utility levels of the two types in their standard offer curve equilibrium. They represent the surplus of these utility levels over the autarchy utility levels and are fixed as long as the indifference maps of the agents do not change. The other payoffs all reflect cheating and depend on the technology and equipment for cheating of each agent. We assume that they too are parameters for our purposes.

The structure of the game is as follows[2]:

1. Both parties decide whether to participate or not. If either withdraws, the game ends and they receive the payoffs (0, 0).
2. But if both decide to participate, they either act honest or cheat and receive the payoffs indicated above.

We solve the game by backward induction, assuming initially that both have decided to participate. The resulting subgame is a standard Harsanyi game of incomplete information with a solution in mixed strategies. The expected surplus of the first agent from acting honestly depends on his beliefs regarding the probability of the second agent's acting honestly. This, in turn, is the sum of the likelihood of the latter's being intrinsically honest (λ) and the probability of his being opportunistic but choosing to act honest.. Suppose that, if he is an opportunist, he chooses to act honest with probability p (i.e. opportunistic second group agents play a mixed strategy in which the probability of their acting honest is p). Then the first agent expects to encounter honest behaviour with probability $\lambda + (1 - \lambda)p = p + (1 - p)\lambda$ and cheating with probability $(1 - \lambda)(1 - p)$. The surplus he expects from acting honestly himself will therefore be

$$V_{1h} = (p + (1 - p)\,\lambda)\alpha_{hh} + (1 - \lambda)(1 - p)\alpha_{hc} \qquad (1)$$

while that from cheating is

$$V_{1c} = (p + (1 - p)\,\lambda)\alpha_{ch} + (1 - \lambda)(1 - p)\alpha_{cc} \qquad (2)$$

In a Bayes-Nash equilibrium, the two will be equal, yielding

[2] At first this game may seem to be very similar to the "sophisticated prisoners dilemma" in BASU [1995], in which two parties play a PD game with an option to enter or stay out. However, our game differs in the important respect that it is a hawk-dove rather than a PD game ; the best response to bad behavior /cheating in our model is to be good (honest) as doing otherwise will lead to an escalation of violence. In the PD, of course, the dominant strategy is to always be bad/cheat.

$$p(\lambda) = [(\lambda/(1-\lambda)(\alpha_{ch} - \alpha_{hh}) + (\alpha_{cc} - \alpha_{hc})]/(\alpha_{hh} - \alpha_{hc} - \alpha_{ch} + \alpha_{cc}) \quad (3)$$

Such an equilibrium is feasible for the subgame iff $1 \geq p^* \geq 0$, which implies

$$\lambda \leq (\alpha_{cc} - \alpha_{hc}))/(\alpha_{hh} - \alpha_{hc} - \alpha_{ch} + \alpha_{cc}) = \lambda_1$$

For higher values of the honesty coefficient, the first agent will expect a higher surplus from cheating regardless of the second's behaviour. If he is an opportunist, he will therefore invariably cheat.

Similarly, the second agent expects to encounter honest behaviour with probability $q + (1 - q)\lambda$ where he believes opportunistic first group agents will play honest with probability q. In Bayes-Nash equilibrium, the value of q that makes cheating and honesty indifferent for the second agent is

$$q(\lambda) = [(\lambda/(1-\lambda))(\beta_{hc} - \beta_{hh}) + (\beta_{cc} - \beta_{ch})] /(\beta_{hh} - \beta_{ch} - \beta_{hc} + \beta_{cc}) \quad (4)$$

The non-negativity restriction on q implies

$$\lambda \leq (\beta_{cc} - \beta_{ch}))/(\beta_{hh} - \beta_{ch} - \beta_{hc} + \beta_{cc}) = \lambda_2$$

Consider now the decisions of the players to opt out or stay in the game. Each stays in if the value of the subgame for him $V_{ih} = V_{ic}$ $(i = 1, 2)$ is non-negative. A little manipulation shows that

$$V_{1h}^* = V_{1c}^* = (\alpha_{hh}\,\alpha_{cc} - \alpha_{hc}\,\alpha_{ch})/(\alpha_{hh} - \alpha_{hc} - \alpha_{ch} + \alpha_{cc})$$

The denominator is negative, so that the non-negativity of $V_{1h}^* = V_{1c}^*$ requires

$$(\alpha_{hh}\,\alpha_{cc} - \alpha_{hc}\,\alpha_{ch}) \leq 0 \quad (5)$$

The non-negativity of $V_{2h}^* = V_{2c}^*$ implies

$$(\beta_{hh}\,\beta_{cc} - \beta_{hc}\,\beta_{ch}) \leq 0 \quad (6)$$

All this may be summed up in

Proposition 1. Given a mass λ of intrinsically honest individuals in the population, if $\lambda \leq min\,[\lambda_1, \lambda_2]$, and conditions (5) and (6) obtain, a Bayes-Nash equilibrium emerges in which opportunistic agents of both groups employ mixed strategies with p and q given by equations (3) and (4).

A notable feature of the equilibrium in which opportunists of both groups play mixed strategies is that the probabilities of a randomly chosen member of either group *acting* honest ($\lambda + (1 - \lambda)p$ and $\lambda + (1 - \lambda)q$) will be independent of the index of intrinsic honesty λ, provided the latter is less than λ_1 and λ_2.

$$\lambda + (1 - \lambda)p = (\alpha_{cc} - \alpha_{hc}))/(\alpha_{hh} - \alpha_{hc} - \alpha_{ch} + \alpha_{cc}) = \lambda_1$$

$$\lambda + (1 - \lambda)q = (\beta_{cc} - \beta_{ch}))/(\beta_{hh} - \beta_{ch} - \beta_{hc} + \beta_{cc}) = \lambda_2$$

Within this domain ($0 \leq \lambda \leq min[\lambda_1, \lambda_2]$) honest behaviour reflects not intrinsic honesty, but the structure of payoffs to honesty and cheating. In particular, if $\lambda = 0$, $p = \lambda_1$ and $q = \lambda_2$: a perfectly opportunistic population will display the same probability of honest behaviour as one with a positive fraction of intrinsically honest people (provided this fraction is not higher than the indicated threshold).

Taking into account the dependence of p on λ, V_{1h} and V_{1c} can be plotted against λ as in Fig. 1. For $\lambda \leq \lambda_1$, $V_{1h} = V_{1c}$ at a constant level. For $\lambda > \lambda_1$, $p = 0$ and both V_{1h} and V_{1c} are linear increasing functions of λ with V_{1c} increasing more steeply than V_{1h}. Likewise, we can construct V_{2h} and V_{2c} (analogously defined for group 2 individuals) as constant and equal functions of λ for $\lambda \leq \lambda_2$ and linearly increasing functions for $\lambda > \lambda_2$ with V_{2c} having the steeper gradient. In the figure, the solid lines denote the expected payoffs from honesty and cheating over each range of λ.

What if the conditions in Proposition 1 are violated? If $\lambda > max [\lambda_1, \lambda_2]$ but (5) and (6) continue to hold, cheating will be the dominant strategy for opportunists of both groups so that $p = q = 0$; the high proportion of honest agents will ensure that expected incomes from honesty on both sides of the market remain positive though cheating will be more rewarding than honesty for both groups.

On the other hand, if λ lies between λ_1 and λ_2, with (5) and (6) still holding, things will be different. If $\lambda_1 > \lambda > \lambda_2$, all opportunists in the second group will cheat ($p = 0$). Inserting $p = 0$ in equations (1) and (2), we infer that condition (5) ensures that honesty is the dominant strategy for first group opportunists ($q = 1$). In turn, it can readily be checked that $q = 1$ implies that, for $\lambda_1 > \lambda > \lambda_2$, $V_{2c} = \beta_{ch} > \beta_{hh} = V_{2h}$. *Vice versa* if $\lambda_2 > \lambda > \lambda_1$.

We have assumed so far that inequalities (5) and (6) are satisfied. If either of these is violated (say (5)), the relevant common horizontal level of the V_{1h} and V_{1c} functions will no longer lie in the non-negative quadrant as in Fig. 1. Fig. 2 illustrates this situation. First group agents will withdraw unless the level of intrinsic honesty is high enough for V_{1h} and V_{1c} to be non-negative. If it is high enough, they could participate in the market (provided second group agents do); however, all opportunists among them will cheat (since this is the more profitable option) and be

recognized as doing so. The honest first group agents participate if $V_{1h} \geq 0$ – which implies

$$\lambda \geq -\alpha_{hc}/[-\alpha_{hc} + \alpha_{hh}] = \lambda_3$$

If this condition is violated, they will not participate, it will be recognized that all first group players remaining in the market will cheat, ensuring a negative expected outcome for any potential partner. No transaction will therefore take place. If $\lambda \geq \lambda_3$, transactions are feasible because the fraction of behaviorally honest agents is high enough to raise the expected payoff of their partners to non-negative levels. Again, the solid lines in Fig. 2 indicate the expected payoffs from each course of action for any value of λ. There are no solid lines in the region $\lambda < \lambda_3$, since no transactions can take place here. Similarly, if (6) is violated, the participation of honest second group agents (and therefore the occurrence of any transaction) requires

$$\lambda \geq -\beta_{ch}/[-\beta_{ch} + \beta_{hh}] = \lambda_4$$

If these conditions are not fulfilled, the honest individuals in the group (or groups) that expect(s) negative income from honesty will withdraw from the game at the outset: only opportunists will remain in the group in the second period and they can all be expected to cheat, which means that their partner's payoffs will be negative in all cases. Anticipating this, the potential partners would prefer to withdraw from the game at the outset. Therefore, unless both conditions are satisfied ($\lambda \geq max\ [\lambda_3, \lambda_4]$), no transaction takes place. It can readily be checked that $\lambda_1 \geq \lambda_3$ iff condition (5) is fulfilled, and that $\lambda_2 \geq \lambda_4$ iff condition (6) is satisfied. We restate all this in

Proposition 2: If either (5) or (6) is violated and $\lambda \geq max\ [\lambda_3, \lambda_4]$, opportunists of both groups cheat, but the honest also break even. But if $\lambda < max\ [\lambda_3, \lambda_4]$, no transaction can take place.

We have assumed so far that there is no competition between the honest and opportunistic types. However, if, for instance, all agents require a scarce resource, competition for it between the honest and the opportunistic may drive down the return to everyone who uses it, depressing α_{hh} and β_{hh}, until honest agents can no longer break even in equilibrium. Since everyone knows that, in this event, only cheats will remain in the market, all trade will collapse.

Proposition 3: If cheating is the best strategy for opportunists, and if individuals of both types compete for an essential resource, trade collapses if the opportunists' requirement of the fixed resource equals or exceeds its supply.

Proof: To capture this process using our model, it suffices to restrict our attention to individuals from group 1 (a symmetric argument holds for those from

group 2). Suppose we are in a parameter range where $V_{1c} > V_{1h}$. Opportunistic individuals of group 1 cheat while intrinsically honest individuals act honestly. However, suppose all group 1 individuals require a fixed quantity of an essential resource without which they cannot transact. This fixed requirement is normalized to one. The aggregate supply of the resource is less than the number of opportunistic group 1 agents. The resource is allocated through a sealed-bid auction. The payoffs in this game are simply those of the previous game less the price of the scarce resource. The structure of this game is as follows:

1. The agents decide whether to play or to stay out.
2. Group 1 agents who opt in bid for the resource, and, if successful, either cheat or play honestly, provided they can find a partner from group 2. Group 2 agents who opt in either cheat or transact honestly

Solve the game as before by backward induction. Consider the subgame that would unfold if agents of both groups stay in initially. It is trivial that the unique Nash equilibrium price that the group 1 agents will offer for the resource will be the maximum that the opportunists among them can offer without their *net* expected payoffs turning negative: V_{1c}. No higher price can be offered by any agent involved in the auction while any lower price can be outbid by an opportunist who wishes to ensure his access to the resource. But if the equilibrium bid for the resource is V_{1c} and $V_{1c} > V_{1h}$, no honest first group agent can make non-negative profits after paying for the resource.

It follows that all honest first group agents will opt out at the outset. But then all second group agents know that any first group agent they may encounter will cheat them, ensuring that their profits are necessarily negative. So no second group agent will enter and trade will collapse. A parallel argument holds for group 2.

Combining Proposition 3 with our earlier results, we derive

Proposition 4. If opportunists and honest individuals compete for an essential resource, trade collapses for all parameter ranges except when conditions (5) and (6) hold and $\lambda \leq min\ [\lambda_1, \lambda_2]$.

Proof. When individuals of both types compete, Propositions 2 and 3, taken together, rule out any transaction if either condition (5) or condition (6) is violated.

If both conditions (5) and (6) hold, and $\lambda > max\ [\lambda_1, \lambda_2]$, opportunists of both groups find cheating to be their best strategy, so that Proposition 3 ensures the breakdown of trade. If with (5) and (6) holding, $\lambda_1 > \lambda > \lambda_2$, all opportunists of the second group find cheating profitable, so that, by Proposition 3, they drive all honest

members of their group out of the market and trade therefore collapses. Opportunists of the first group do likewise if $\lambda_2 > \lambda > \lambda_1$.

Thus, trade can be sustained iff (5) and (6) hold and $\lambda \leq min\ [\lambda_1, \lambda_2]$. In that event, neither type can oust the other, since cheating and honesty are equally profitable.

Nor will the long run evolution of the honesty coefficient erode this equilibrium as it would under unilateral moral hazard. Since there is no premium for either honesty or dishonesty in equilibrium, there will be no endogenous change in λ. Further, societies whose parameters do not permit such an equilibrium (societies, for example, with $\lambda > max\ [\lambda_1, \lambda_2]$) will in the long run retreat into autarky and be unable to enjoy the gains from trade. They will therefore rank lower in the scale of evolutionary fitness than a population that is other wise identical but in which (5) and (6) hold and $\lambda \leq min\ [\lambda_1, \lambda_2]$. The latter will multiply faster than the former. In the event of conflict between them, the former will very likely lose out. Thus, evolutionary selection between populations will in fact increase the likelihood of our encountering societies in which our equilibrium obtains.

4. Conclusion

We have shown the possibility, with certain configurations of payoffs, of the persistence of virtue. A coefficient of intrinsic goodness higher than a specified threshold cannot be sustained. However, once the coefficient drops down to the threshold, or below it, it will not change endogenously, since now good and bad behaviour are equally profitable. Game theory, while it sets a ceiling to the level of intrinsic virtue sustainable in a society, does not dictate any further erosion below this level, certainly not its extinction. The threshold is independent of the intrinsic goodness index and is determined entirely by the structure of payoffs. Good *behaviour* is thus independent of the exact value of the proportion of innately good individuals – provided this proportion is not too large – in the long run, though not in the short run.

References

ANDERSON, E. [2000], "Beyond homo economicus: new developments in the theory of social norms," *Philosophy and Public Affairs,* 29, 170-200.

BASU, K. [1984], *The Less Developed Economy,* Basil Blackwell: Oxford.

BASU, K. [1995], "Civil institutions and evolution, concepts, critiques and models," *Journal of Development Economics* 46, 1, 19-33.

GHATAK, S. and K. INGERSENT [1984], *Agriculture and Economic Development*, Johns Hopkins Press: Baltimore.

GNEEZY, U. [2005], "Deception and the role of consequences," *American Economic Review,* 95, 1, 384-94.

HAUSER, M. D. [2006], *Moral Minds,* Harper Collins: New York.

HAUSMAN, D.M., and M. S. MCPHERSON [1996], *Economic Analysis and Moral Philosophy,* Cambridge University Press: Cambridge.

MAYNARD SMITH, J. [1972], Game theory and the evolution of fighting, in : J. Maynard Smith (ed) *On Evolution*, Edinburgh University Press: Edinburgh.

MAYNARD SMITH, J., and G. PRICE [1973], "The Logic of Animal Conflict," *Nature*, 146, 15-18.

MUKHERJEE, B. [1984], "Between Rationality and Value Judgment: those Non-free Riders of Taxicabs," *Economic and Political Weekly*, 19, 27, 1057-1060.

MYERSON, R. B. [2004], "Justice, institutions and multiple equilibria," *Chicago Journal of International Law*, 5, 1, 91-107.

SAMUELSON, L. [2002], "Evolution and game theory," *Journal of Economic Perspectives*, 16, 2, 47-66.

TAYLOR, P.D and L.B JONKER [1978], "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, 40, 145-156.

FIG.1: Expected Payoffs to Honesty and Cheating when (5) and (6) hold

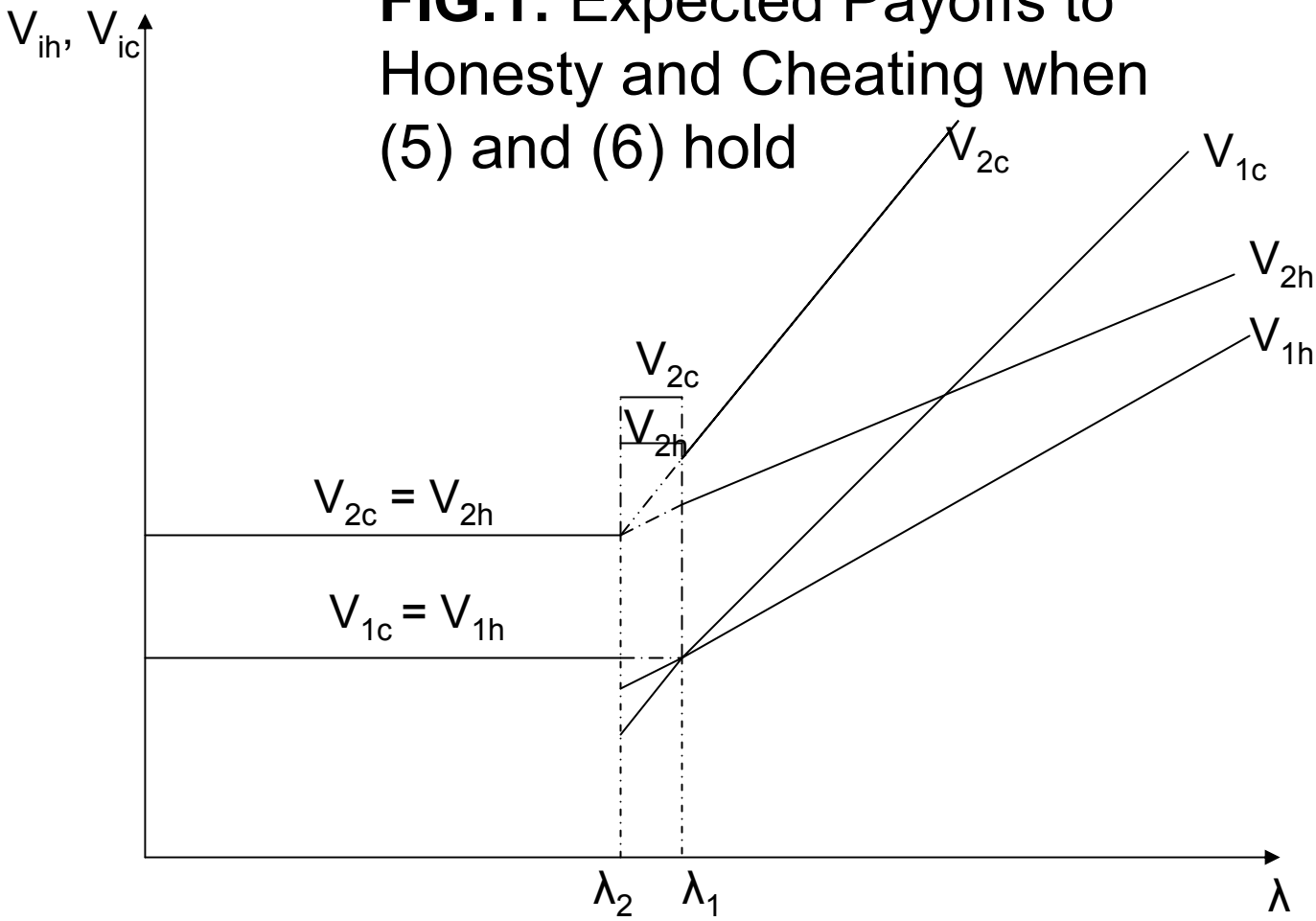

Fig 2 : Expected payoffs to honesty and cheating when (5) and/or (6) do not hold