Citation: Dasgupta, Partha (2000) 'Trust as a Commodity', in Gambetta, Diego (ed.) *Trust: Making and Breaking Cooperative Relations*, electronic edition, Department of Sociology, University of Oxford, chapter 4, pp. 49-72, <http://www.sociology.ox.ac.uk/papers/dasgupta49-72.pdf>.

<<49>>

4

Trust as a Commodity

Partha Dasgupta

TRUST, CREDIBILITY, AND COMMITMENT

Trust is central to all transactions and yet economists rarely discuss the notion.[1] It is treated rather as background environment, present whenever called upon, a sort of ever-ready lubricant that permits voluntary participation in production and exchange. In the standard model of a market economy it is taken for granted that consumers meet their budget constraints: they are not allowed to spend more than their wealth. Moreover, they always deliver the goods and services they said they would. But the model is silent on the rectitude of such agents. We are not told if they are persons of honour, conditioned by their upbringing always to meet the obligations they have chosen to undertake, or if there is a background agency which enforces contracts, credibly threatening to mete out punishment if obligations are not fulfilled - a punishment sufficiently stiff to deter consumers from ever failing to fulfil them. The same assumptions are made for producers. To be sure, the standard model can be extended to allow for bankruptcy in the face of an uncertain future. One must suppose that there is a special additional loss to becoming bankrupt - a loss of honour when honour matters, social and economic ostracism, a term in a debtors' prison, and so forth. Otherwise, a person may take silly risks or, to make a more subtle point, take insufficient care in managing his affairs, but claim that he ran into genuine bad luck, that it was Mother Nature's fault and not his own lack of ability or zeal.

<<50>>

I have given these preliminary examples so as to make seven points, each of which I will elaborate in what follows. Firstly, if there is an absence of suitable punishment, that is incurred loss, for breaking agreements or contracts, individuals will not possess the appropriate *incentives* to fulfil them; and since this will be generally recognized within the population, people will not wish to enter into transactions with one another. Thus, what could in principle be mutually beneficial relationships will not be initiated.

Secondly, the threat of punishment for errant behaviour must be *credible*, else the threat is no threat. To put it another way, the enforcement agency itself must be *trustworthy:* it will do what it says and only what it says. I should add, parenthetically, that the enforcement agency may be society at large. Social ostracism, and the sense of shame that society can invoke, are examples of such punishment. I should also add that at another extreme are cases where the 'enforcement agency' may be the injured party to the transaction. Here, as in game-theoretic models of reciprocal altruism (analysed by Friedman 1971; Trivers 1971; Aumann and Shapley 1976; Rubinstein 1979; Maynard Smith 1982; Axelrod 1984; Fudenberg and Maskin 1986), it is the injured party which punishes the 'defaulter' by ceasing to transact with him (see the fourth section of this chapter, and also Singer 1981, for a fine exploration of the link between ethics and sociobiology).

---

[1] <<49>> I am most grateful to Diego Gambetta for suggesting that I write on this topic, and to him, Kenneth Arrow, Paul David, John Hartwick, Dieter Helm, Hugh Mellor, Ugo Pagano, Peter Temin, and Menahem Yaari for illuminating discussions on the subject. I have also benefited greatly from the searching comments made by participants of the trust seminar and those of the seminar of the Sub-Faculty of Economics at Oxford University.

Thirdly, and this follows from the first and second points, trust among persons and agencies is interconnected. If your trust in the enforcement agency falters, you will not trust persons to fulfil their terms of an agreement and thus will not enter into that agreement. By the same token, you will not trust the enforcement agency - for example, the government - to do on balance what is expected of it if you do not trust that it will be thrown out of power (through the ballot-box or armed rebellion) if it does not do on balance what is expected of it. It is this interconnectedness which makes trust such a fragile commodity. If it erodes in any part of the mosaic it brings down an awful lot with it. This is *one* reason (there are others) why the medical and legal professions had, and in many places still have, such stern codes of conduct instilled into their members; they needed to break this intricate link, as it were, so that vital transactions concerning health and protection could be entered into even if enforcement costs were to rise due to an erosion of trust elsewhere in the economy, through rapidly changing social mores or for whatever other reason (see Arrow 1963).

Fourthly, and this is implicit in the third point, you do not trust a person (or an agency) to do something merely because he says he will do it. You trust him only because, knowing what you know of his disposition, his available options and their consequences, his ability and so <<51>> forth, you expect that he will *choose* to do it. His promise must be credible. That is why we like to distinguish 'trusting someone' from 'trusting someone blindly', and think the latter to be ill-advised.

Fifthly, and this follows from the fourth point, when you decide whether to enter into an agreement with a person, you need to look at the world from his perspective as it is likely to be when it comes to his having to fulfil his part of the agreement. This explains why the mathematician Richard Bellman (1957), and the game theorists Thomas Schelling (1960) and Reinhard Selten (1965; 1975), instructed us always to calculate *backwards*, against time, and not *forwards*, with time (I will elaborate on this point later on).

Sixthly, even though there are no obvious units in which trust can be measured, this does not matter, because in any given context you can measure its value, its worthwhileness (see below). In this respect, trust is not dissimilar to commodities such as knowledge or information.

Seventhly, and for the moment most importantly, I am using the word 'trust' in the sense of correct expectations about the *actions* of other people that have a bearing on one's own choice of action when that action must be chosen before one can *monitor* the actions of those others.

There are other uses of the word 'trust'. When on meeting an acquaintance one says 'I trust your family is well', one is expressing a hope that his family is well. I do not think that anything else is implied. On the other hand, when one says 'trust Frank Halin to say that one must economize on trust', one is merely saying that one's expectations of his attitude have been confirmed. But this use does not have the potency that the word actually possesses. For whether one is right or wrong about Hahn may have no bearing on one's actions. Trust is of much importance precisely because its presence or absence can have a strong bearing on what we choose to do and in many cases what we *can* do. The clause concerning the inability to *monitor* others' actions in my definition of trust is crucial. If I can monitor what others have done before I choose my own action, the word 'trust' loses its potency. I should emphasize that an inability to monitor the actions of others need not be due to the fact that my choice of action temporally precedes those of others. Certainly in many cases this will be so, such as my lending you a book, trusting that you will return it in five years' time. But there are a great many other cases where what I ought now to do depends on whether you have done what you said you would do and where I cannot *now*, or possibly ever, monitor whether you have actually done it. I will not burden you with examples of this sort. Nothing of analytical importance depends on this particular classification.

I have so far, in defining one's trust in others, talked of the significance of others' unobservable *actions* for choosing one's own course of <<52>> action. But there is another important class

of cases where trust, in the sense that I wish to use the term, comes into play. This is when others *know* something about themselves or the world which I do not, and when what I ought to do depends on the extent of my ignorance of these matters. An agreement between myself and such other people may call upon them to disclose their information. But can I trust them to be truthful; that is, can I trust them to send me the correct signals, those they would send if they were truly trustworthy? (They do not actually have to be *truthful* for me to rely on them. As long as I always know how to interpret their messages correctly I can trust them. Thus the ancient Cretan was as informative as a knowledgeable saint.) Examples of this class abound. When I ask my mother what she would like as a present I hope I can trust her to tell me the truth and not try to save me money. When I go to a second-hand car dealer I worry whether I can trust him not to sell me what Americans call a 'lemon'. The point is not so much whether the car is in fact a lemon (the dealer may not know) but whether the dealer *knowingly* sells me a lemon.[2]

Having distinguished these two broad categories around which the concept of trust revolves, I am going to restrict myself, in what follows, to the second; that is, those circumstances where an individual does not know fully the disposition (or motivation) of the person(s) with whom he is considering a transaction. The reason for this is that at the *analytical* level both categories raise the same issues. There are some derived notions of trust; for example, I trust that the value of the pound will not fall to zero tomorrow. But this is a derived notion, since the value of the pound tomorrow depends upon what we all do, what we all expect, what we all know, and so forth. My definition covers such cases automatically.[3]

Talk of 'contracts' and 'agreements' might suggest that I am, in discussing trust, taking an unduly legalistic attitude. I do not intend to do that, and I use 'contracts' and 'agreements' merely as props. A contract can be vague. In fact no contract, even if it is scrutinized by sharp lawyers, can detail every eventuality, if for no other reason than that no <<53>> language can cope with unlimited refinement in distinguishing contingencies. Thus trust covers expectations about what others will do or have done (or what messages they will transmit) in circumstances that are not explicitly covered in the agreement. In many cases there may not even *be* an agreement: can I trust people to come to my rescue if I am about to drown? Towards the end of this article, when I will conclude that trust is based on *reputation* and that reputation has ultimately to be acquired through behaviour over time in well-understood circumstances, it will be seen that none of these distinctions, between actions and message transmission, between legal contracts and implicit understandings, is of any analytical moment for the problem at hand. It is important to realize that when Sam Goldwyn remarked that a verbal contract is not worth the paper it is written on, he was only half right, and that all that is interesting in the concept of trust lies precisely in that half which was wrong.

In defining trust I have spoken of one's expectations regarding others' choice of actions that have a bearing on one's own choice of action. Now, of course, choice need not be based exclusively on self-interest and nothing I have said or will say supposes that it is so based. We are all at once both egoists and altruists, occasionally rising to the moment and doing what is the right thing to do and not what is in our personal interest, and unhappily often failing to so rise. Furthermore, it is often the case that the mere fact that someone has placed his trust in us makes us feel obligated, and this makes it harder to betray that trust. Again, at the general analytical

---

[2] <<52>> In the insurance literature, resource allocation problems arising from incomplete information on the part of one party regarding other parties' *actions* are called 'moral hazard', and those arising from incomplete information on their part regarding the others' *characteristics* (for example, other parties' disposition, or the knowledge they possess) are called 'adverse selection'. At a very general level, moral hazard and adverse selection raise very much the same sorts of issues (see for example Laffont and Maskin 1981).

[3] <<52>> Luhmann (this volume) suggests reserving the term 'confidence' for 'trust' in the ability of social institution (c. g. the market) to *function* as is expected of it. Likewise, it seems to me, we show 'confidence' in our doctor's *ability* to cure us of our ailments, in our teacher's *ability* to inspire us, in our civil servants' *ability* to take the correct decisions, and so on. Thus confidence stems from ability, and trust from a person's underlying disposition or motivation.

level it does not matter whether we see people - having imposed moral constraints on their available set of options - choosing in the light of self-interest, or whether they explicitly consider trade-offs between self-interest and the interests of others. The problem of trust would of course not arise if we were all hopelessly moral, always doing what we said we would do in the circumstances in which we said we would do it. This is, the problem of trust would not arise if it was common knowledge that we were all trustworthy. A minimal non-congruence between individual and moral values is necessary for the problem of trust to *be* a problem. So of course I shall assume that there is non-congruence: all I want to warn you against is the idea that noncongruence necessarily implies undiluted personal greed.

It is nevertheless possible to claim on the one hand that a person is untrustworthy and on the other that he can be trusted to do what he said he would on a given occasion. This is because on this occasion he may have the right incentive. 'Trustworthiness' concentrates on a person's overall disposition, his motivation, the extent to which he awards importance to his own honesty. Being able to trust a person to do what he said he would, on the other hand, requires us to know not only <<54>> something of his disposition, but also something of the circumstances surrounding the occasion at hand. If the incentives are 'right', even a trustworthy person can be relied upon to be untrustworthy. 'Every man has his price': repugnant though it is to our sensibilities, the cliché captures the view that no one awards an infinite weight to his own honesty. Even Yudhishthira in the epic *Mahabharata,* renowned for his trustworthiness, uttered on one occasion what in effect was a lie so as to throw off his unrighteous enemies. Yudhishthira, as it turned out, was a consequentialist. But the point is that his enemies did not know this. They thought that Yudhishthira could never be motivated to deviate from the path of truth. This is why the lie worked.

We wish to know the sort of person we are dealing with before we deal with him. But we will know it only imperfectly. We form an opinion on the basis of his background, the opportunities he has faced, the courses of action he has taken, and so forth. Our opinion is thus based partly on the theory we hold of the effect of culture, class membership, family line, and the like on a person's motivation (his disposition) and hence his behaviour. The opinion which is publicly formed and held is this person's reputation. Our problem in essence is to infer the person's qualities from such data. This will be the central theme in the remaining sections of this article.

Can one trust oneself? Here I will not address this question at any length. But I will discuss it briefly as a springboard for my central case - trust in others. There is the trivial sense of not trusting oneself to do the right thing in circumstances when one cannot think things through clearly, such as a state of intoxication or great emotional stress. There is not much to discuss here, except to note that in certain cases society commits itself to render invalid any agreements entered into in such states. The important class of cases that needs analysis is where a person does not trust the extent of his own commitment, his *ability* to carry out and see through his own projects. In many such cases he imposes routines on himself, at great inconvenience (to himself and on occasion to others), not merely to stay in mental and physical shape, as it were, but to persuade *himself* of his own commitment. The rituals of Hindu ascetics might usefully be understood in some such terms. An interesting instance of this trust in oneself arises in those all-too-familiar cases where one's values or preferences are expected to change over time. What should one do now if one expects (and in extreme cases knows) that one will acquire values later which are incongruent with one's current values? The right way to attack the problem, as has been shown by a number of economists, is to view it as an intertemporal game between one's temporal selves.[4] The classic example, as noted by Strotz (1955-56), is <<55>> that of Odysseus and the Sirens (for elaborations, see Elster 1979). From our point of view the interesting feature of the story is that Odysseus had the option of (quite literally) binding himself now to a future course of action (by excluding the option of trying to get to the Sirens). He

[4] <<54>> The classic article on the subject is Strotz (1955-56). Important subsequent contributions include Phelps and Pollak (1968), Blackorby et al. (1973), and Peleg and Yaari (1973). Analyses of such intertemporal games as these which are consistent with the backward-induction argument (see below) are provided in Dasgupta (1974), Hammond (1976), Yaari (1977), and Bernheim and Ray (1983).

exercised that option, and this enabled him to choose a further option (that of listening to the Sirens) which he would not have chosen had he not been able to so bind himself. Of course, it is also the case that Odysseus could trust his fellow wanderers not to hear his pleas during the critical period in which he begged for release. Otherwise he would not have had the option of binding himself. Odysseus's customary self had, as game theorists would call it, a first-mover advantage over his bewitched self.

The suggestion of a link between commitment and trust is made in this volume on several occasions. I think the link is a pretty obvious one. I also think it best not to dwell too long on it, because it tends to distract from trust itself. When one wonders whether to enter into a binding commitment one is really wondering which intertemporal game to play. This can be clearly seen in the case of an exchange of contracts on a house. The penalty (loss of money, honour, and so forth) for breaking a verbal agreement is usually less than the penalty for breaking the legal agreement that is entailed on exchange of contracts. All that commitment does is turn the original game into one in which the explicitly or implicitly agreed-upon courses of action are credible. If they are credible, then the participants can be trusted to carry them out. To use once again the game theorist's terminology, certain types of commitment (for example an exchange of contracts; and to people of honour, a promise) change the initial game into one in which mutually beneficial courses of action become equilibrium. strategies when in fact they were not so initially. (I will illustrate this presently.) Such commitments often involve the expenditure of resources, such as lawyers' fees and deposits. But the gain is a mutually beneficial transaction that can be trusted to be carried out. Presumably, each participant calculates the cost of such commitment and compares it to the 'gains' to him emanating from it. The cost of Commitment is then the price that has to be paid for each party to trust the other to fulfil the terms of the agreement. We are now solidly in the domain of economics. No doubt you were not sure when I would entice you into the economist's lair. But doubtless you trusted me to do so somehow.

Let me then summarize the argument so far. You do not trust a person to do something merely because he says he will do it. You trust him <<56>> because, knowing what you know of his disposition, his information, his ability, his available options and their consequences, you expect he will choose to do it. Commitment on his part merely alters the game that is to be played, and therefore alters your expectation of what he will choose to do and, what is subtly tied to this, simultaneously alters *his* expectation of what you will choose to do.

How is trust created? How is it perpetuated? Why is it not present when it is not? Why is it there when it is there, and why does it break down when it does? How can it be built up again when it breaks down? And finally what are we to make of the point made recently by Professor Albert Hirschman: that trust, like other moral resources, grows with use and decays with disuse?[5]

There are, it seems to me, two ways of approaching these questions, one of which is analytical and the other, for want of a better description, anthropological. The former pursues a general abstract route; the latter illuminates matters by compiling and analysing historical case studies. Several of the articles in this volume fall into the second category. What follows belongs rather to the first. But I make no pretence of answering these questions. The examples that follow are designed to indicate an avenue - a research strategy - for answering them. Along the way I hope the questions will be met in part.

SELF-FULFILLING EXPECTATIONS ABOUT HONEST AND DISHONEST BEHAVIOUR

Before presenting a detailed case (in the third and fourth sections) which, it seems to me, captures some of the crucial aspects of the formation of trust, I want to offer an intermediate example. This example at first blush seems promising for the problem at hand. In fact it

[5] <<56>> 'These are resources whose supply may well increase rather than decrease through use; second, these resources do not remain intact if they stay unused; like the ability to speak a foreign language or to play the piano, these moral resources are likely to become depleted and to atrophy if not used' (Hirschman 1984: 93).

captures something related, but something else. I want to discuss it because it is one of the first examples that comes to mind. It will also prove a useful prelude to my main example.

There is a large group of individuals in a society who meet pairwise at random in every period to transact. Each party at each encounter can do one of two things: X (transact honestly) and Y (cheat a little in the transaction). To have an interesting problem I must suppose that the parties choose their actions simultaneously. Each person would prefer to do X over Y if the person he encounters were also to choose X (it is nice to be honest when dealing with honesty), and he would prefer Y over X if <<57>> the person he meets were to choose Y (it is unpleasant to be a sucker). I now want to assume that each party prefers both to choose X rather than Y (it is preferable to have bilateral honesty than bilateral cheating).

In what follows I will suppose that while people appear to be the same there are in fact two types of persons in the population, one given to more honest urges (type 1) than the other (type 2). To make all this vivid I present in tables 4.1 and 4.2 the benefits (or pay-offs) to a person of each type when he transacts with someone. In table 4.1 the pay-offs to a person of type 1 are presented, and he is assumed to choose a row (the party he is transacting with is assumed to choose a column). In table 4.2 the pay-offs to a person of type 2 are presented, and he too is assumed to choose a row (the party he is transacting with is assumed to choose a column). For simplicity of exposition I have made the pay-off matrices the same except for the north-west box, and this is sufficient to indicate that a person of type I is more inclined towards honest dealing.

TABLE 4.1 *Pay-offs to a person of type 1, who chooses a row*

	X	Y
X	30	5
Y	5	10

TABLE 4.2 *Pay-offs to a person of type 2, who chooses a row*

	X	Y
X	20	5
Y	5	10

I want to suppose that the population is large, so that when two people meet at random to transact they cannot tell if they have met before. A person's past record of choices is thus not known by anyone (but see the fourth section). What then should an individual choose when he <<58>> randomly meets another for transaction? Clearly, a person's best policy in this example depends on what he *thinks* the other will do. But what thought should he rationally entertain?

Consider first a person of type 1. It is simple to confirm that his optimal choice is X if he regards the chance that the other party will choose X to be in excess of 1/6, and his optimal choice is Y if this chance is less than 1/6.[6] Now consider a person of type 2. It is equally simple, confirm that his optimal choice is X if he regards the chance that the other party will choose X with a probability in excess of 1/4, and his optimal choice is Y if this chance is less than 1/4.[7]

[6] <<58>> To confirm this, suppose this chance is p. Then for the person of type 1 to be *indifferent* between X and Y, p must satisfy the equation $30p + 5(1 - p) = 5p + 10(1 - p)$, yielding p = 1/6.

[7] <<58>> Because 1/6 is less than 1/4, a person of type 1 needs less 'persuasion' to choose X. It is in this sense that type 1 people in the example lean more frequently towards honest dealing.

Let us now suppose, for the sake of vividness, that the proportion of people of type 1 in the population is P, and suppose this is publicly and commonly known. I want to concentrate attention on stationary (or steady-state) behaviour patterns, and for simplicity I shall restrict myself to equilibria where persons of the same type choose the same option.

Notice at once that if each person expects the others to choose X then each will chose X, whereas if each expects the others to choose Y then each will choose Y. Thus both X and Y are equilibrium behaviour. The role of expectations is of course crucial: if everyone expects everyone to be honest then everyone will be honest, and if everyone expects everyone to cheat a little then everyone will cheat a little.[8] Note too that each is a stable expectation, in that a slight departure from either equilibrium expectation will bring the expectation back to equilibrium, provided expectations at each period adjust in the direction of the past period's average behaviour. We conclude, then, that if people are somewhat cynical about one another (others will choose Y with a high probability), everyone will emerge worse off than if people trust one another (others will choose X with a high probability).

The population ratio *P* has not entered into the equation so far. It has an interesting role to play if its value lies between 1/6 and 1/4. Consider any value in this range. Then it is clear that in addition to the two equilibria we have already identified there is another, where each person of type 1 chooses X, and of type 2 chooses Y, on every occasion. At this equilibrium it is common knowledge that type 1 people are always honest and that type 2 people always cheat a little. But at no encounter does either party know who he is dealing with. Behaviour according to type is therefore what emerges at this equilibrium.

<<59>>

The example on its own cannot, of course, tell us equilibrium will prevail in the long run, if indeed it is a stationary equilibrium towards which the society gravitates. For this we need more information about the dynamics of expectation formation. History clearly matters here, but it can matter in many ways and it is hard to tell how best to capture it.

Instructive though such an example might appear to be, it is not the right one for my purpose. For trust to be developed between individuals they must have repeated encounters, and they must have some memory of previous encounters. Moreover, for honesty to have potency as a concept there must be some *cost* involved in honest behaviour. And finally, trust is linked with reputation, and reputation has to be acquired. The example which follows attempts to take account of these issues.

AN EXAMPLE OF POSSIBLE COOPERATION

A salesman possesses a number of motor-cars. Their outward appearances are the same, but in fact some of them are reliable and the rest are 'lemons'.[9] The salesman knows each car intimately. In order to concentrate on what for us are the essential points, I shall assume that there is no haggling: the salesman has posted a fixed price in his showroom window. Any potential customer can, should he wish to, enter the showroom, pay the quoted price, and drive away in a car *the salesman picks out for him.* Alternatively, he can choose not to enter the showroom at all. I assume in this section that there is a single potential customer. Should this person not enter, the salesman sells no car, and in the event I assume (without loss of generality) that the *net benefit* (or *pay-off*) to each party is nil. However, should the person enter and pay the price, the salesman has two options: to hand over a reliable car (option A) or to saddle the customer with a lemon (option B). If a lemon is sold the net benefit (or pay-off) to

[8] <<58>> Multiple self-fulfilling expectations have received careful attention in Schelling (1978).

[9] <<59>> As mentioned earlier, a 'lemon' is a 'sour purchase', a commodity which is not worth the price paid for it. In a well-known article, Akerlof (1970) analysed a more complicated version of the example in this section.

the salesman is γ and that to the customer is $-\beta$, where $0 < \beta < 1$.[10] If a reliable car is sold the net benefit (or pay-off) to the salesman is α and that to the customer is $1 - \beta$.

<<60>>

Thus the customer prefers buying a good car to not buying a car at all, and he prefers this in turn to buying a 'lemon'. (This is reflected in the hypothesis that $1 - \beta > 0 > -\beta$.) I now assume that these four pay-off values, $-\beta$, $1 - \beta$, $-\gamma$, and α, are common knowledge.[11]

The tree describing this two-move game is depicted in figure 4.1. At the initial node of the tree (labelled 1) the customer chooses whether to enter and buy a car. (The customer thus has the first move.) Should he not enter, the game terminates. Should he enter, the game reaches the node labelled 2, the salesman chooses between A and B, and the game then terminates. Thus the customer has to choose between the two strategies 'enter' and 'not-enter'; and the salesman has to choose between the two strategies 'hand over a lemon' and 'hand over a reliable car'.

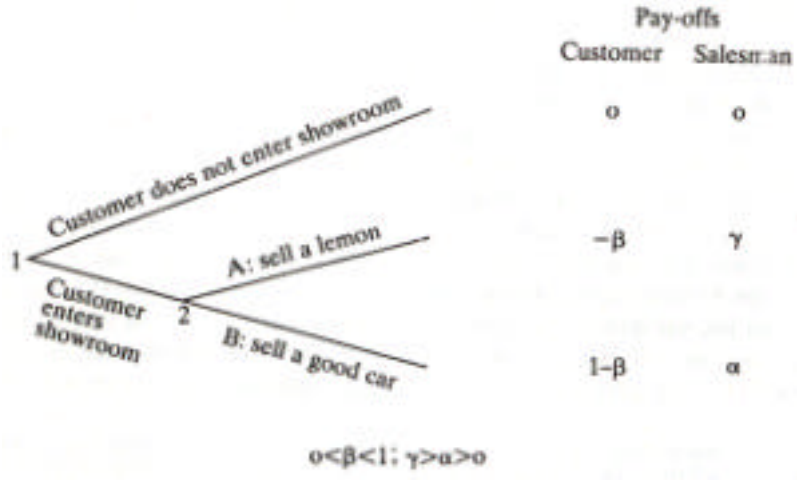

FIGURE 4.1 *Game tree: salesman dishonest*

I begin by considering a 'dishonest' salesman, I formalize his disposition by the hypothesis that γ exceeds α. To have an interesting problem I assume further that α is positive, that is $\gamma > \alpha > 0$. In other words, the salesman would choose to sell the customer a lemon rather than a reliable car, but he would choose the latter option over not selling a car at all. Since the pay-offs are common knowledge it is, among other things, common knowledge that the salesman is dishonest; that he would <<61>> choose A (selling a lemon) over B (selling a reliable car). So of course there will not be a transaction. The potential customer reasons *backwards*: 'If I enter the showroom and make a purchase the salesman will sell me a lemon and my pay-off will be $-\beta$, which is less than zero - my pay-off in not buying a car at all. So of course I should not enter

[10] <<59>> By hypothesis the customer realises subsequently that the car is a lemon, but by then there is nothing he can do. Since $-\beta$ is negative we are making the natural assumption that buying a lemon is worse than not buying a car at all. To be sure, salesmen usually offer warranties and the like, but they are a means of making the salesman's (or manufacturer's) claims about the reliability of the car credible. I am for the moment thinking of situations where such options are not available to the salesman, because I am discussing trust. Even when warranties are available, there is always the residual chance of being saddled with a lemon. Naturally, I want to consider such situations.

[11] <<60>> By this I mean that each knows them, that each knows that each knows them, that each knows that each knows that each knows them, and so on *ad infinitum*. For a non-technical account of the role that the common knowledge assumption plays in game theory, see Binmore and Dasgupta (1986). I should add that we do not require the full force of the common knowledge assumption in the simple two-move game I am considering now, but we will need it subsequently when we extend the game to discuss other matters.

the showroom'. And he will not. We have thus a unique equilibrium of the game where no transaction occurs and where each party enjoys a pay-off of nil. This is worse for both than the (non-equilibrium) outcome where a reliable car is sold and where the customer and the salesman enjoy β and α respectively. A mutually beneficial transaction is thus unrealized.[12]

Notice that the salesman would love to persuade the customer of his 'good intentions', but mere words would not, in this example, be sufficient. If he could at little cost commit himself to choosing B - conditional upon the customer entering the showroom - he would do so. For example, if he could offer the potential customer a legally valid contract in which he would be liable to the customer for an amount something in excess of $\gamma - \alpha$ should the customer subsequently find that he has been sold a lemon, he would offer it. And if he were to offer it the customer's optimal strategy would be to accept it, and to buy a car, confident of the fact that the salesman would now be handing over a reliable car, it being now in his interest to do so. I shall suppose that the salesman cannot make such commitments, because I want to study the link between 'trust' and 'the reputation for being trustworthy'. [13] For this I must study a game in which people have incomplete information. I therefore make a minimal alteration to the game in figure 4.1: I suppose that the potential customer is unsure whether the salesman is dishonest, as in the game in figure 4.1, or honest and trustworthy. The relevant payoffs for the dishonest salesman are as in figure 4.1; those for the honest type are as in the right-hand column in figure 4.2. Thus, to the honest salesman zero is the pay-off in not being able to sell a car, μ the pay-off in selling a lemon, and δ the pay-off in selling a reliable car. Naturally, I want to assume that δ is positive and that it exceeds μ. In other words, an honest salesman is one who would choose to sell a good car if he were given the option of selling a car and if he said it was a good car.[14]

<<62>>

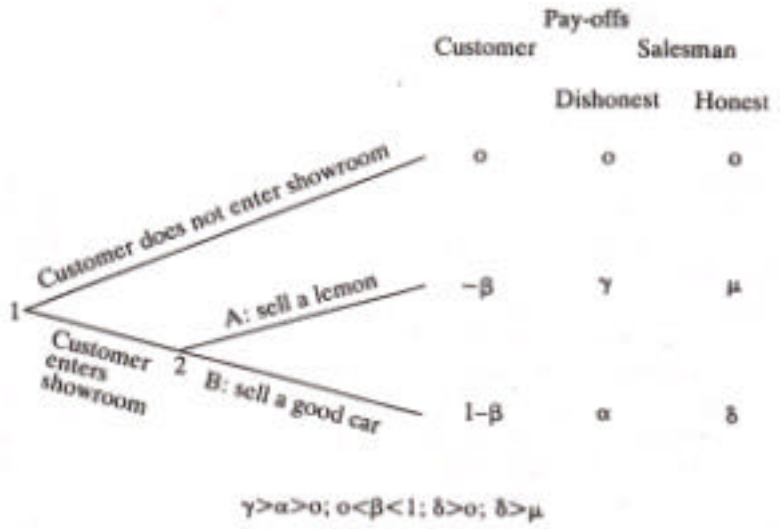

FIGURE 4.2 *Game tree: salesman dishonest or honest*

[12] <<61>> The idea of 'pay-offs' as represented by numbers such as β, α and γ is one which I am appealing to only for expositional purposes. Behind it is the more general idea that each party can *rank* the outcomes (the eventualities) of the game and that choices conform to these rankings. These rankings can be based on as complex a set of considerations as we would like to include in our model of the choosers. The game tree I am discussing in the text is not based on an identification of choice with preference, or of preference with welfare.

[13] <<61>> Warranties and the like reduce the need for personal knowledge on the part of the customer about the salesman's trustworthiness (i.e. his characteristics). But as noted earlier the need for trust cannot be eliminated entirely by such forms of commitment: the 'buck', is merely passed on to agencies that enforce contracts such as warranties. For this reason I assume that warranties do not exist in the game under consideration.

[14] <<61>> We do not need to specify whether μ is positive or negative.

I now suppose that the customer is unsure whether the salesman is trustworthy or untrustworthy, and that he imputes a (subjective) probability *p* to the salesman being honest. I want to think of *p* as being the 'reputation' the salesman has for honest dealing. This last move on my part needs elaboration, to which I now proceed.

Reputation is a capital asset. One can build it up by pursuing certain courses of action, or destroy it by pursuing certain others. Sometimes a reputation can be acquired by pure chance, as was the point of the short story *Luck* by Mark Twain. Sometimes it can be destroyed by misfortune, as in the case of Jean Valjean in *Les Misérables.* A reputation for honesty, or trustworthiness, is usually acquired gradually. This alone suggests that the language of probabilities is the right one in which to discuss reputation: a person's reputation is the 'public's' imputation of a probability distribution over the various types of person that the person in question can be in principle. Reputation is rarely of the all-or-nothing variety. And although a reputation for honesty may be acquired slowly, it can generally be destroyed very quickly. We will wish to see whether the framework we develop can accommodate this asymmetry.

At one level I am using the word 'reputation' somewhat loosely in the present customer-salesman example, for I am not explaining why *p* is what it is. The salesman in question may be a representative salesman of a population of salesmen, some of whom are known to be honest, some not. In this case *p* is not so much the salesman's as the *population's* <<63>> reputation. For example, the Gurkhas have a worldwide reputation for bravery on the battlefield. No doubt individual Gurkhas vary in their bravery. But this group reputation has been won over the years by the bravery displayed by thousands of Gurkhas. A young Gurkha today enjoys a reputation painfully acquired by the efforts of others. We impute a high probability to his being brave, not necessarily because he has done anything - he may have gone untested so far - but because he is a Gurkha, he is one of *them.* In the next section I will discuss the much more difficult problem of reputation acquisition.

Now consider again the game tree in figure 4.2, where the potential customer does not know for sure whether the salesman in question is honest. As before, the left-hand column contains the pay-offs to the customer in the three possible eventualities of the game, the middle column the pay-offs to the salesman should he be untrustworthy, and the right-hand column the pay-offs to the salesman should he be trustworthy. All the parameters of the game, α, β, γ, δ, 1 - β, and *p*, are common knowledge. And I assume for the sake of simplicity that the potential customer chooses his action on the basis of his *expected net benefit.* This too is common knowledge.[15] Should he not enter the showroom, his net benefit is nil. Should he enter, his net benefit is clearly $p(1 - 0) - (1 - p)\beta$, or in other words $p - \beta$. Whether he will choose to enter therefore depends upon whether *p* exceeds β.

I conclude that if the seller's reputation for honesty is high - specifically if *p* exceeds β - the potential buyer will engage in a transaction. If it is low - specifically if *p* is less than he will not enter and there will not be a sale. Each party will receive nothing in this case.

Simple as this example is, it nevertheless allows us to clarify two ideas at once. Firstly, if *p* is less than β to start with, an increase in *p* sufficient to tilt it over β is beneficial to all parties, in the sense that expected net benefits, in equilibrium, are thereby raised. The customer gains because the expected benefit from buying a car is now positive. But so does the salesman gain. If he is honest he gains δ, if dishonest then γ. Thus an honest salesman would be prepared to pay up to δ to find some way of increasing *p* from a figure below β to a figure above it. The

[15] <<63>> We do not need the expected net benefit hypothesis for this game. All we need to postulate is that the customer's preferences for entering the showroom are an increasing function of *p*.

dishonest salesman would, on the other hand, be prepared to pay up to γ. I conclude that 'reputation' is worth more to the honest salesman only if δ exceeds γ.

But we now have a mild paradox. If δ exceeds γ and if there is some mechanism in which by investing resources a salesman can tilt *p* from below β to a figure above it, the honest salesman can outbid his <<64>> counterpart, the dishonest one. To put it more accurately, if the salesman is willing to spend more than γ, and if he does so, he will reveal himself unambiguously as being honest (*p* will jump to the value of unity). But if γ exceeds δ there cannot be any such mechanism for raising *p* in this simple model. If the salesman is honest he will be willing to spend a maximum of δ. The dishonest salesman will be willing to spend more, but if he does he will reveal himself as being unambiguously dishonest! So of course he will not spend more. But if he does not, there is no way for the customer to distinguish the two possible types, and it is precisely because of this that the salesman (no matter which type he is) will not spend resources in improving his image.[16]

The second point this simple example brings to the surface is the underlying *externality* among persons that trust, or the lack of it, creates. Trust, as economists have noted before, is a public good, a social lubricant which makes possible production and exchange (see for example Akerlof 1970; Arrow, 1974). To see this externality, suppose for example that δ exceeds γ. In this case the honest salesman would be willing to pay a maximum of δ to have his reputation increase from a figure below β to one above it. But the customer would also be willing to pay something towards this increase - in fact to the extent of his own increase in expected net benefit. We may thus conclude that an honest salesman's willingness to pay for an increase in reputation falls short of society's willingness to pay. The point is that if *your* trust in me increases it certainly benefits me. But if there are good reasons for this increase in trust it benefits *you* as well. This latter benefit I do not take into account when I try to build up my own reputation. This is the source of 'market failure' and, in particular, why there might typically be an under-investment in trust formation.[17]

But what are the mechanisms through which one can buy and build one's reputation? Furthermore, how do we formalize Professor Hirschman's suggestion that trust grows with use? Although the questions are related the latter is easier to answer, at least in part. First of all, there is the fact that one feels a sense of obligation not to betray someone's trust if that someone has placed his trust in one. But this sense of obligation is not commonly overriding. If it were I would not be writing this essay now.

Secondly, there is the fact that society is not composed of culturally alienated beings. In dealing with someone you learn something not only <<65>> about him, but also about others in his society. You learn something about population statistics. Therefore, if you meet several honest persons and no dishonest ones you might want to revise your prior opinion of the society at large. It is unlikely that you would say you were especially fortunate to have met the few honest ones and so the chance that the remaining members are dishonest is now higher. You would display a good deal of ignorance of both social anthropology and statistical inference if you thought like that. By the same token, of course, if you were to meet several dishonest persons in your initial encounters you might also want to revise your prior opinion: the posterior

[16] <<64>> This argument depends critically on there being a single play of the game and a single salesman. In the following section I will relax each of these assumptions. Readers who are familiar with the insurance literature will note that the argument in the text implies that equilibrium is separating when δ exceeds γ and that it is pooling when γ exceeds δ.

[17] <<64>> Typically, but by no means necessarily. If there is competition among salesmen there may be excessive investment in reputation acquisition as each salesman tries to outdo his rivals - much as can happen in races for industrial patents. For a theoretical exploration of such races, see Dasgupta and Stiglitz (1980a; 1980b).

distribution you would hold would reflect a lowered opinion in this case. Such initial runs of bad experience - leading to termination in relationships of production and exchange - can occur with positive chance. Thus suppose there are many salesmen, of whom some are honest (their pay-offs are given by the right-hand column in figure 4.2) and the remainder are dishonest (the middle column in figure 4.2), and suppose the proportion of honest salesmen is *p*. Suppose there is a single customer who encounters salesmen at random every week (i.e. a different salesman each week) and has to decide each week whether to make a single purchase of the commodity in question. (The commodity must, in this example, be non-durable and not a motor-car!) Now suppose the customer to begin with does not know the 'true' population average *p*, but entertains a subjective probability distribution of the population average. If his prior estimate of the population average is high, specifically in excess of *p*, he will purchase the commodity in his first encounter. If the commodity is of good quality he will revise his initial estimate upward. If, on the other hand, the commodity is bad, he will revise it downward. The point is, however, that even if the *true* population average *p* exceeds β, he will, with a *positive* probability, have a run of bad encounters leading to so low a revised estimate of the population average that he will, as a rational agent, no longer wish to make a purchase. Once he ceases purchasing he will cease forever, because a non-purchase yields no information about the population at large. Thus, if *p* exceeds β there are two possible outcomes: one where after a finite number of purchases transactions cease, and another where the initial run turns out on average to have been favourable, so that purchases continue indefinitely and the customer's posterior estimate of the population average converges in the long run to the actual figure of *p*. (If *p* is less than β then of course the only long-run equilibrium is a cessation of purchase.)[18]

<<66>>

The third point which explains Professor Hirschman's observation is the fact that bonds develop among people who encounter one another repeatedly. A sharp dealer will not sell a defective piece of merchandise to a previous customer, and this is not necessarily because otherwise the customer would not come again. The point is psychologically deeper.

THE ACQUISITION OF A REPUTATION FOR HONESTY

While all this is true, none of it throws light on the idea of a person acquiring a reputation for honesty by his act. For this we need to discuss repeated plays of our basic game and to assume that the salesman's behaviour (whether he has sold a reliable car or a lemon) is recorded after each transaction, so that future customers know the seller's history.[19]

Suppose that the game in figure 4.2 is repeated over time, and suppose - although this is not essential, since the salesman's past behaviour is, by assumption, common knowledge - that a different potential customer appears in each period. But assume that the salesman remains the same since it is his reputation we are trying to model.

Consider first the extreme, but well-known, case of $p = 0$; customers (subjectively) know for sure that the seller's pay-offs are as in figure 4.1. If the game is repeated a finite number of

[18] <<65>> Formally, this model is akin to the model of a 'two-armed bandit' - two adjacent fruit machines - in statistical decision theory, where the gambler knows the true odds offered by one of the arms and not those of the other. For an excellent presentation of the two-armed bandit problem, see Rothschild (1974). The model in the text can be extended, with only a little difficulty, to include many (identical) buyers and, as in the example in the second section, to have customers and sellers meet each week at random. The natural assumption to make here is that customers never encounter the same salesman twice (or what is the same thing, have no memory of who the past salesmen were) and that each customer can see how many other customers are continuing to purchase the good. (This latter assumption allows each customer to glean something about others' experiences in the market.) If *p* exceeds β there is a unique long-run equilibrium, with each customer purchasing each week. If *p* is less than 0 the unique long-run equilibrium is a complete cessation of trade.

[19] <<66>> In actual markets, word of mouth is the most common mode in which such records are kept.

times, say T, where T is known in advance, no customer will ever enter the showroom. This follows from the familiar backward-induction argument.[20] Thus a move from a single encounter game to a finitely repeated game does not alter the conclusion if $p = 0$. In other words, our refusal to deal with a fly-by-night operator, if we *know* he is untrustworthy, will remain the correct strategy even if he were to cease being a fly-by-night operator and announce that he was <<67>> offering business for the next 50 years. We should still not deal with him.[21]

Now continue to assume that $p = 0$ but that the encounters are to go on forever. That is, set T equal to infinity.[22] If the salesman is known to discount his future benefits at a low enough rate then it is the case that the outcome, where at each period the customer enjoys $1 - \beta$ and the salesman α - that is, where at each date the customer enters the showroom and the salesman hands over a reliable car - is an equilibrium outcome even if there is no possibility of binding commitment of any sort. The point here is that customers and salesman have available to them what can be shown to be a set of credible strategies whereby future customers refuse to enter the showroom should the salesman ever sell a lemon. Since the salesman discounts future benefits at a low enough rate it is in his interest to sell reliable cars only: by reneging on his promise he enjoys a one-period gain of $\gamma - \alpha$ but loses an infinite flow of benefits, α at each period, for customers never again enter his showroom. The above, of course, does not constitute an argument, merely a strong indication that there may *be* an argument establishing the claim that there are credible strategies - that is, strategies which jointly satisfy the backward-induction argument - supporting the 'cooperative' outcome.[23]

But while this example fits in part the notion of trust as I have defined its it captures only one aspect of the idea of trust, namely that customers can trust the salesman to hand over a reliable car. It cannot convey the idea of the salesman trying to acquire a reputation for honesty - trying, through his behaviour, to alter people's perception of what kind of person 'he really is'. For notice that in the foregoing infinitely repeated game customers do not revise their opinion of him. They *know* he is dishonest (in the sense of having the pay-off structure of figure 4.1) and behaves honestly only because of the punishment (loss of customers) that will be meted out to him should he never renege. To represent the idea of reputation formation we will, if we are followers of the good Reverend Thomas Bayes, have to assume that customers are initially uncertain as <<68>> to whether the salesman is honest. They all assign a positive probability p that he is honest; that is, a probability p that his pay-off structure is the right-hand column of figure 4.2. We assume that this is common knowledge. (This initial assignment is provided historically, as it were, by the past behaviour of other salesmen. For example, imagine that we are all queueing outside a store in a Cairo bazaar.)

---

[20] <<66>> See for example Luce and Raiffa (1957) for this argument applied to finite repetitions of the Prisoner's Dilemma game. See also Axelrod (1984). The argument 'proceeds' as follows. There is no way a cooperative outcome can be realized within the structure of the game. The point is that no matter what promises the salesman makes he will, given that his pay-offs are as in figure 4.1, sell a lemon at the last date T should a customer come. This is common knowledge. So in effect $T - 1$ is the last date, when again he will sell a lemon should a customer arrive at his showroom. And so on.

[21] <<67>> This seemingly paradoxical result really tells us that the model is a bad one, not that we have a deep philosophical problem on our hands, because the result hangs on a number of highly unrealistic assumptions pertaining to human motivation and knowledge. Radner (1980) has shown how cooperative outcomes can be sustained even in finitely repeated games if individuals are not fine-tuned optimizers. In the text below I shall explore cooperation by introducing ignorance on the part of potential customers about the characteristics of the salesman in question, that is, by supposing that the initial reputation p is positive.

[22] <<67>> Actually, we need only assume that there is a chance that encounters will not terminate at any finite date.

[23] <<67>> Readers unfamiliar with infinitely repeated games will find an excellent account in Axelrod (1984). The most complete account to date of repeated games is Fudenberg and Maskin (1986); see also Friedman (1971).

To focus on the idea of reputation acquisition, I assume that p is less than β. We noted in the third section (figure 4.2) that if the salesman is in business for only one period the customer will not enter the showroom. So I suppose that the salesman is in the market for T periods, where T is some large positive integer.[24] Assume finally for simplicity of exposition that the salesman does not discount future benefits to himself.

Notice that given the data available to all potential customers, if the salesman were ever to sell a lemon (i.e. choose option A in figure 4.2) he would reveal himself as having the pay-off structure in the middle column of figure 4.2. The game would end then and there, because his reputation would plummet from whatever positive probability assessment customers had assigned him to a solid zero.[25] Once he was found out, no customer would ever come to him. (This is a finitely repeated game, so the backward-induction argument sketched in footnote 20 would now be valid.) Trust would be totally destroyed. Of course, a dishonest seller can reason the consequences of reneging. So can each potential customer. Equally, if the seller is in fact honest he will always sell a reliable car (i.e. choose option B in figure 4.2) when called upon. Each participant can reason thus, just as we have done. But what if the seller is actually untrustworthy? He will certainly sell a lemon in the final period should a customer come to him. He no longer cares for his reputation. In the final period he is truly a fly-by-night operator.

What of previous years though? Might it not be in a dishonest seller's interest to sell a reliable car in the initial period, simply to keep open in the public mind the possibility that he is honest and thus encourage future customers to enter his showroom? Furthermore, since customers can mimic this reasoning, may they not arrive in the early periods, <<69>> anticipating correctly that even if the seller is dishonest his behaviour will be honest; that is, he will choose B so as not to destroy his original reputation p of being honest? For the point to note is that, to an untrustworthy seller, playing B in any period yields a loss amounting to $\gamma - \alpha$ in that period. Set against this one-period loss is the gain of α at each period in the future that he would enjoy if, by choosing B, he could keep the customers' trust in him. Using techniques developed in the important work of Kreps and Wilson (1982) on games of incomplete information, we can show that such reasoning on the part of participants is mutually consistent if T is large. I shall not go into technicalities. But it can be shown that if T is large there is a set of *credible* strategies, one per participant, where for a large number of initial periods customers come to the showroom, one by one, and where the salesman sells a reliable car (i.e. chooses B) *irrespective of his true disposition*, thus maintaining his reputation at the initial estimate p. (The reason the estimate does not change is that behaviour in these periods is independent of the salesman's disposition, that is, independent of whether he is honest or dishonest. This being so, the fact that he sells reliable cars provides no information about his true disposition. Hence his reputation does not change.) The remaining periods are a bit harder to describe, since this involves the salesman, if *he is dishonest*, bluffing by choosing between A and B in random manner (if he is honest he always chooses B).[26] As long as during this set of periods customers see the salesman selling reliable cars, they update their assessment. His reputation as an honest salesman rises, and if the

---

[24] <<68>> I do not suppose that T is infinity (that is, an infinitely repeated game), because we would not then know from our analysis which part of our answer is based on the reputation effect.

[25] <<68>> I have not allowed in the model for the salesman to make a genuine mistake. In the world as we know it a salesman's reputation usually does not collapse on the basis of his selling, say, one lemon. This is because we like to think that even the salesman may not know the true characteristics of his own merchandise. The possibility of genuine mistakes or, more generally, bad luck on the part of the salesman can be incorporated into the present model by invoking the important recent investigation by Radner (1981). The effect of such considerations is as one might expect. In equilibrium customers will keep track of the time series of the ratio of lemons sold to the total number of cars sold. If this ratio ever exceeds some critical value - a value determined by among other things the chance at each stage that the salesman has made a genuine mistake - customers dump the salesman.

[26] <<69>> It may also involve potential customers randomizing over whether to enter. But for expositional ease I shall ignore this.

run of Bs continues it rises to a value in excess of β. (The reason the salesman's reputation increases along an uninterrupted run of Bs in those periods when the dishonest salesman is *known* to choose B randomly follows from Bayes's rule. The honest salesman will in any case choose B, and the dishonest salesman is known to choose B with a positive probability *less than one.* So if in fact B obtains, the possibility that the salesman in question is genuinely honest is enhanced.) If the salesman is truly honest, all is well: customers continue to come until the very end (but see footnote 26) and only reliable cars will be sold. But if the salesman is dishonest one of two things will happen. Firstly, in some period while the salesman is randomizing over A and B, A will be chosen by the 'luck of the draw'; customers will realize the salesman is dishonest, and no further exchange will occur. Alternatively, the random draws will continue to produce B in each period until, towards the end, it is the dishonest salesman's equilibrium strategy to choose A with probability one. In either event a trusting customer will at *some* date be saddled with a lemon and from then onwards transactions will cease.

<<70>>

I think this is the right sort of framework to capture the idea that people invest resources for the purpose of building a reputation for honesty. The gains from honest behaviour are built directly into the model. The fact that the dishonest salesman builds his reputation by going against his short-term interests is an important feature to highlight. This is his investment.

The first, and to my mind most serious, weakness of this model is that there is nothing which enables the honest salesman to distinguish himself from the dishonest one. Everything the former can do the latter can do as well. The dishonest type is the cancerous cell which 'acts' like the healthy one. It is here that the role of commitments can assume importance. If the stake is high enough the honest salesman will search for ways of persuading customers that he really is honest; he will be willing to spend resources to distinguish himself from his possible counterpart. This is not included in the foregoing model of reputation acquisition.

The second weakness is that the strategies I sketch for the various participants are not the only credible ones. It follows that the outcome I describe is not the only viable one. There are other, equally credible, sets of strategies that result in outcomes which do not have this form of reputation acquisition, where transactions are not undertaken for a great many periods. In short, such repeated games possess more than one equilibrium outcome (see Fudenberg and Maskin 1986). There is therefore no guarantee that the one I have outlined will prevail. It is here that historical accidents may matter. I have assumed a single salesman in the model. But there are usually many sellers in the world, some of whom are honest, some dishonest. And of course there are many customers. Chance events may set off transactions among some protagonists, the associated sellers embarking on the acquisition of reputation, while others - some of whom are truly honest - are left by the wayside for want of customers. It seems plausible to suppose that if a record is kept of salesmen's past behaviour (say through word of mouth) customers will gravitate towards salesmen with a good track record. But I do not know how to prove it. Adding competition among salesmen enlarges the initial game, enlarges the class of strategies from which participants can choose, and makes the problem very difficult to analyse.

One way to simplify the problem is to assume that participants have a very limited number of strategies at their disposal. An extreme case, explored by socio-biologists, is that where each participant has only one strategy available to him; that is, no one has any choice. In socio-biological models such as those of Maynard Smith (1982) and Axelrod (1984), a participant in a game *is* a strategy. Such models are a great deal easier to analyse because the long-run outcomes can be readily traced to the initial distribution of strategies in the population. The question socio-biologists <<71>> have asked is not: 'What outcomes are viable when participants are rational calculators?' They have asked instead: 'What strategies have survival value?' Mercifully, people in the world we know are not as restricted in their choices as participants in socio-biological models. But the moral must surely be that repeated games need

some form of 'friction' to generate predictable outcomes.[27] Moral codes are a form of friction. There are certain things, while feasible, that are 'not done'. This may well be a route to pursue in exploring the emergence of trust in analytical models.

REFERENCES

Akerlof, G. 1970: The market for 'lemons': qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, 488-500.

Arrow, K. J. 1963: Uncertainty and the economics of medical care. *American Economic Review* 53, 941-73.

Arrow, K. J. 1974: *The Limits of Organization*. New York: Norton.

Aumann, R. and Shapley, L. 1976: Long term competition: a game theoretic analysis. Hebrew University, Jerusalem: unpublished paper.

Axelrod, R. 1984: *The Evolution of Cooperation*. New York: Basic Books.

Bellman, R. 1957: *Dynamic Programming*. Princeton: Princeton University Press.

Bernheim, D. and Ray, D. 1983: Altruistic growth economies. IMSSS technical report no. 419, Stanford University.

Binmore, K. and Dasgupta, P. 1986: Game theory: a survey. In K. Binmore and P. Dasgupta (eds), *Economic Organizations as Games,* Oxford: Basil Blackwell.

Blackorby, C., Nissen, D., Primont, D., and Russell, R. R. 1973: Consistent intertemporal decision making. *Review of Economic Studies* 40, 239-48.

Dasgupta, P. 1974: On some alternative criteria for justice between generations. *Journal of Public Economics* 3, 405-24.

Dasgupta, P. and Stiglitz, J. E. 1980a: Industrial structure and the nature of innovative activity. *Economic Journal* 90, 266-93.

Dasgupta, P. and Stiglitz, J. E. 1980b: Uncertainty, industrial structure and the speed of R&D. *Bell Journal of Economics* 11, 1-28.

Elster, J. 1979: *Ulysses and the Sirens: studies in rationality and irrationality*. Cambridge: Cambridge University Press.

Friedman, J. 1971: A non-cooperative equilibrium for supergames. *Review of Economic Studies* 38, 1-12.

Fudenberg, D. and Maskin, E. 1986: A folk-theorem in repeated games with discounting and with incomplete information. *Econometrica* 54, 533-54.

Hammond, P. J. 1976: Changing tastes and coherent dynamic choice. *Review of Economic Studies* 43, 159-73.

<<72>>

Hirschman, A. 0. 1984: Against parsimony: three easy ways of complicating some categories of economic discourse. *American Economic Review* Proceedings, 74, 88-96.

Kreps, D. and Wilson, R. 1982: Reputation and imperfect information. *Journal of Economic Theory* 27, 253-79.

Laffont, J.4. and Maskin, E. 1981: The theory of incentives: an overview. University of Cambridge: unpublished paper.

Luce, R. D. and Raiffa, H. 1957: *Games and Decisions*. New York: Wiley.

Maynard Smith, J. 1982: *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Peleg, B. and Yaari, M. 1973: On the existence of a consistence course of action when tastes are changing. *Review of Economic Studies* 40, 391-401.

Phelps, E. S. and Pollak, R. A. 1968: On second-best national savings and game equilibrium growth. *Review of Economic Studies* 35, 185-99.

Radner, R. 1980: Collusive behaviour in non-cooperative epsilon equilibria of oligopolies with long but finite lives. *Journal of Economic Theory* 22, 136-54.

Radner, R. 1981: Monitoring cooperative agreements in a repeated principal-agent relationship. *Econometrica* 49, 1127-48.

[27] <<71>> By 'predictable' I mean narrowing down the set of viable (or equilibrium) outcomes.

Rothschild, M. 1974: A two-armed bandit theory of market pricing. *Journal of Economic Theory* 9, 185-202.
Rubinstein, A. 1979: Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory* 21, 1-9.
Schelling, T. 1960: *The Strategy of Conflict.* New York: Oxford University Press.
Schelling, T. 1978: *Micromotives and Macrobehaviour*. New York: Norton.
Selten, R. 1965: Spieltheoretische Behandlund eines Oligopolmodels mit Nachfragetragheit. *Zeitschriftjur die Gesampte Staatswissenschaft* 121, 401-24 and 667-89.
Selten, R. 1975: Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4, 25-55.
Singer, P. 1981: *The Expanding Circle: ethics and sociobiology*. Oxford: Oxford University Press.
Strotz, R. H. 1955-56: Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23, 165-80.
Trivers, R. 1. 1971: The evolution of reciprocal altruism. *Quarterly Review of Biology* 46, 35-57.
Yaari, M. 1977: Consistence utilization of an exhaustible resource - or how to eat an appetite arousing cake. Research memorandum no. 23, Hebrew University.