

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/255682316>

# Can We Trust Trust? Diego Gambetta

Article · August 2000

---

CITATIONS

51

---

READS

5,005

1 author:



[Diego Gambetta](#)

Collegio Carlo Alberto

71 PUBLICATIONS 8,293 CITATIONS

SEE PROFILE

**Citation:** Gambetta, Diego (2000) 'Can We Trust Trust?', in Gambetta, Diego (ed.) *Trust: Making and Breaking Cooperative Relations*, electronic edition, Department of Sociology, University of Oxford, chapter 13, pp. 213-237, <<http://www.sociology.ox.ac.uk/papers/gambetta213-237.pdf>>.

<<213>>

## 13

### Can We Trust Trust?

Diego Gambetta

In this concluding essay I shall try to reconstruct what seem to me the central questions about trust that the individual contributions presented in this volume raise and partly answer.<sup>1</sup> In the first section, I briefly qualify the claim that there is a degree of rational cooperation that should but does not exist, and I shall give a preliminary indication of the importance of the beliefs we hold about others, over and above the importance of the motives we may have for cooperation. In the second section, I define trust and the general conditions under which it becomes relevant for cooperation. In the third, I discuss the extent to which cooperation can come about independently of trust, and also whether trust can be seen as a result rather than a precondition of cooperation. In the final section, I address the question of whether there are rational reasons for people to trust - and especially whether there are reasons to trust trust and, correspondingly, distrust distrust.

## I

The unqualified claim that more cooperation<sup>2</sup> than we normally get would be desirable is generally sterile, is often characterized by irritating <<214>> rhetorical flabbiness and, if preached too extensively, may even have the effect of making cooperation less attractive (Hirschman 1984a). Such a claim can be and is disputed in a variety of ways. Let us begin by considering whether we necessarily need more cooperation, keeping, for the moment, the distinction between cooperation and trust blurred and their relationship implicit.

'According to the trite observation' - Adam Smith wrote - 'if there is any society among robbers and murderers, they must at least ... abstain from robbing and murdering one another' (Smith [1759] 1976: 86; see also Saint Augustine in Dunn, this volume). This 'trite' observation serves a double purpose: it reminds us that basic forms of cooperation are inevitable if a society is to be at all viable, but it also points out, perhaps unwittingly, that there are instances of cooperation - notably those among robbers and murderers - that we may want to dispose of rather than improve. We may want *less* cooperation (and trust) rather than more, especially among those who are threatening us, and whose cooperation is a hindrance to ours. *A priori*, we cannot always say whether greater trust and cooperation are in fact desirable (Schelling 1984: 211).

---

<sup>1</sup> <<213>> In this essay converge not just several of the ideas which contributors have published in this volume but also those which were patiently expressed in conversation, and thanks to which my reflections on trust were shaped and reshaped. I am intensely grateful to all of them. Throughout the seminar, I relied constantly on the invaluable help of Geoffrey Hawthorn. The essay also benefits from exchanges with several other people at different stages of the seminar. In particular, I would like to express my appreciation to Luca Anderlini, Elisabetta Galeotti, Albert Hirschman, Caroline Humphrey, Alan Macfarlane, Andy Martin, Paul Ryan, Hamid Sabourian and Allan Silver. I am also particularly grateful to Heather Pratt for helping me to edit this volume as well as for polishing my awkward English.

<sup>2</sup> <<213>> In this essay 'cooperation' is meant in the broad sense of agents, such as individuals, firms, and governments, agreeing on any set of rules - a 'contract' - which is then to be observed in the course of their interaction (cf. Binmore and Dasgupta 1986:3). Agreements need not be the result of previous communication but can emerge implicitly in the course of interaction itself, and rules need not be written but can be established as a result of habit, prior successful experience, trial and error, and so on.

The problem, however, is not only that we may want less of it among our enemies, but also that we may not want it among ourselves, at least not all the time.<sup>3</sup> And it is not just that we may lazily wish not to have to cooperate, but that we may wish for something else instead, notably competition. The ideological stance which holds competition and the ‘struggle for survival’ to be the texture of life is largely inadequate: in so far as it draws upon analogies with the animal world for its legitimation it is quite simply wrong (Bateson 1986 and this volume; Hinde 1986) and if taken literally there is no need to go back to Hobbes to realize that it would make social life impossible, or at least utterly unpleasant. Yet a certain dose of competition is notoriously beneficial in improving performance, fostering technological innovation, bettering services, allocating resources, spreading the fittest genes to later generations, pursuing excellence, preventing abuses of power - in short, in enriching the human lot. The rationale for this view is that not only those who succeed in competition benefit, but that the positive influence of competition is likely to be more generally felt.

<<215>>

Even if we believe both that there are people whose cooperation we would wish to diminish and that the arguments in favour of competition are empirically as well as theoretically valid in a sufficiently large number of instances to make them relevant, it does not follow that, thus qualified the claim that there is a wide variety of cases where it would be desirable to improve cooperation is thereby invalidated. On the contrary, this claim holds in the relationships between as well as (to differing degrees) within countries, whether socialist or capitalist, developed or underdeveloped. The problem, stated in very general terms, seems to be one of finding the optimal mixture of cooperation and competition rather than deciding at which extreme to converge. Cooperation and competition are not necessarily alternatives; they can and do coexist in both the animal and the human world. Very few people, however, would venture so far as to claim that in the world as it is we have managed to get that balance right. Neither the Invisible Hand nor, as far as humans are concerned, natural evolution seems to help in selecting optimally between these two states, and we still seem profoundly ignorant of the ways in which we might master the causality that brings them about.

More important still, the possibility of competition may depend upon cooperation to a much larger extent than is generally acknowledged, especially in capitalist countries (Hirsch 1977): the most basic form of human cooperation, abstention from mutual injury, is undoubtedly a precondition of potentially beneficial competition.<sup>4</sup> As Robert Hinde (1986) has pointed out, there is a difference between outdoing rivals and doing them in, and within species, competing animals are considerably more inclined to the former than the latter. Even to compete, in a mutually non-destructive way, one needs at some level to *trust* one’s competitors to comply with certain rules.

This applies equally to political and economic undertakings, and the awareness of such a need is not, of course, particularly novel. In spite of the fact that Hobbes has come down to us as the theorist of the inevitability of coercion in the handling of human affairs, he himself was conscious of the decisive role of the growth of trust among political parties for building viable societies (see Weil 1986). So was Adam Smith with respect to economic life. His notion of self-interest is not only contrasted from ‘above’ with the absence of benevolence, as is predominantly stressed, but also from ‘below’ with the absence of predatory behaviour (Smith [1759] 1976: 86; see also Hont and Ignatieff 1983). And, finally, in a characteristically historical remark Weber observed that the universal diffusion of unscrupulousness in the pursuit of self-interest <<216>> was far more common in pre-capitalist societies than in their supposedly more competitive capitalist counterparts (1970: 17ff.).

---

<sup>3</sup> <<214>> There are also cases where we may not want universal cooperation for, beyond a certain threshold, additional cooperators could jeopardize the effectiveness of cooperation (cf. Elster 1986).

<sup>4</sup> <<215>> This does not entail a causal relationship whereby cooperation generates beneficial competition. It is more likely to be the reverse, i.e. harmful competition may be a motive for seeking cooperation.

The point, though, is not only that we may lack that basic form of cooperation which nurtures beneficial competition, as is often the case in underdeveloped countries. Nor is it just that we have competition where the majority would prefer cooperation, the international relations between superpowers being the foremost example (Hinde 1986). We may simply have the lack of mutually beneficial cooperation, with nothing to replace it. Game theory has provided us with a better understanding of why cooperation may not be forthcoming even when it would benefit most of those involved. As Binmore and Dasgupta put it in their survey of the subject: 'It is a major and fundamental error to take it for granted that because certain cooperative behaviour will benefit every individual in a group, rational individuals will adopt this behaviour' (1986: 24). Irrespective of individual rationality and motivation, cooperation may still fail to take place.

In this respect, one of the most interesting as well as threatening lessons of game theory is that even if people's motives are not unquestioningly egoistic, cooperation may still encounter many obstacles. This is a much more striking result than that which shows that rationality in the pursuit of self-interest may not suffice. Consider, for instance, the well-known case of the Prisoner's Dilemma and related games: the mere expectation that the second player might choose to defect can lead the first player to do so, if only in self-defence. The first player's anticipation of the second's defection may be based simply on the belief that the second player is unconditionally uncooperative. But, more tragically, it may also be based on the fear that the second player will not trust *him* to cooperate, and will defect as a direct result of this lack of trust. Thus the outcome converges on a sub-optimal equilibrium, *even if* both players might have been *conditionally* predisposed to cooperate (cf. Williams, this volume). The problem, therefore, is essentially one of communication: even if people have perfectly adequate motives for cooperation they still need to know about each other's motives and to trust each other, or at least the effectiveness of their motives. It is necessary not only to trust others before acting cooperatively, but also to believe that one is trusted by others.

The lack of belief should not be confused with the lack of motive for cooperation. Motives for cooperation are of course crucial (see Williams, this volume). Yet, the mirror image of the 'major and fundamental error' of taking rational cooperation for granted is another fundamental error: that of inferring, if cooperation does not come about, that there are no rational motives for cooperation, and that people actually *prefer* the lack of it. For example, the ubiquitous problem of traffic jams in cities is often taken as a sign of the predominance of poisonous <<217>> preferences for travelling by car over travelling by other means. Although to some extent this may be so, there are also strong grounds for believing<sup>5</sup> that the motives for cooperation - that is, using bicycles and public transport - are not absent. What is lacking is the belief that everybody else is going to cooperate, which generates the fear of being the only 'sucker' around to sweat on the pedals, and the corresponding unwillingness to cooperate oneself. Thus, rationally motivated cooperation may not emerge and, if it does not, it does not follow that rational motives compatible with an increase in collective welfare are absent, but more simply that not enough people trust others to act by those motives. Revealed preference may simply reveal the fact that they are conditional on our beliefs: if the latter change, the former may change accordingly.

Here, traditional game theory does not help, for it considers beliefs to be far more undetermined than they are in reality, and further assumes that they are common knowledge. As a result, game theory loses predictive power, for it can 'find' more equilibria - usually more uncooperative ones<sup>6</sup> - than in fact there are in the real world. But 'why should beliefs held by different individuals (or types of individual) be commonly known? The fact is that our understanding of

---

<sup>5</sup> <<217>> This suspicion is backed by evidence: in a recent referendum held in Milan nearly 70 per cent of the population - many more than the number of unconditional pedestrians - indicated a preference for closure of the city centre to private and non-residential traffic.

<sup>6</sup> <<217>> Thus Woody Allen, in *Hannah and her sisters*, says that the reason we cannot answer the question 'Why the Holocaust?' is that it is the wrong question. What we should ask is 'Why doesn't it happen more often?' Somewhat similarly, we should ask why uncooperative behaviour does not emerge as often as game theory predicts.

human psychology ... is hopelessly imperfect. In particular, we have little idea of how individuals actually acquire beliefs' (Binmore and Dasgupta 1986: 11). Among these beliefs, trust - a particular expectation we have with regard to the likely behaviour of others - is of fundamental importance.

## II

In this volume there is a degree of convergence on the definition of trust which can be summarized as follows: trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both *before* he can monitor such action (or independently of his capacity ever to be able to monitor it) *and* in a context in which it affects *his own* action (see Dasgupta and Luhmann in particular, this volume). When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him. Correspondingly, when <<218>> we say that someone is untrustworthy, we imply that that probability is low enough for us to refrain from doing so.

This definition circumscribes the focus of our interest in trust in several ways.<sup>7</sup> Firstly, it tells us that trust is better seen as a threshold point, located on a probabilistic distribution<sup>8</sup> of more general expectations, which can take a number of values suspended between complete distrust (0) and complete trust (1), and which is centred around a mid-point (0.50) of uncertainty. Accordingly, *blind* trust or distrust represent lexicographic predispositions to assign the extreme values of the probability and maintain them unconditionally over and above the evidence.<sup>9</sup> Next, the definition stresses the fact that trust is particularly relevant in conditions of *ignorance* or uncertainty with respect to unknown or unknowable actions of other (see Hart's definition of trust as something suspended between *faith* and *confidence*, and Luhmann on the distinction between the latter and trust, this volume). In this respect, trust concerns not future actions in general, but all future actions which condition our present decisions. Thirdly, by postulating that our own actions are *dependent* on that probability, it excludes those instances where trust in someone has no influence on our decisions. Finally, it limits our interest to trust between agents and excludes that between agents and natural events. At this level of abstraction, the definition could reflect trust in the intentions of others not to cheat us and in their knowledge and skill to perform adequately over and above their intentions.<sup>10</sup> The essays in this volume refer to both, although the former is generally more prominent.

The condition of ignorance or uncertainty about other people's behaviour is central to the notion of trust. It is related to the limits of our capacity ever to achieve a full knowledge of others, their motives, their responses to endogenous as well as exogenous changes. Trust is a tentative and intrinsically fragile response to our ignorance, a way of coping with 'the limits of our foresight' (Shklar 1984: 151), hardly ever located at the top end of the probability distribution. If we were blessed with an unlimited computational ability to map out all possible contingencies in enforceable contracts, trust would not be a problem (see Dasgupta and Lorenz, this volume).

Trust is also related to the fact that agents have a degree of freedom to disappoint our expectations. For trust to be relevant, there must be <<219>> the possibility of *exit*, betrayal, defection. If other people's actions were heavily constrained, the role of trust in governing our decisions would be proportionately smaller, for the more limited people's freedom, the more restricted the field of actions in which we are required to guess *ex ante* the probability of their

---

<sup>7</sup> <<218>> For an interesting review of different views on trust in the social sciences see Mutti (1987).

<sup>8</sup> <<218>> The probability distribution of expectations can also be seen as expressing the *reputation* of others (cf. Dasgupta, this volume).

<sup>9</sup> <<218>> *Loyalty*, in this context, can perhaps be seen as the maintenance of global trust - in a person, a party, an institution - even in circumstances where local disappointments might encourage its withdrawal.

<sup>10</sup> <<218>> For a discussion of these distinctions cf. Barber (1983) and Luhmann, this volume.



performing them. Trust can be, and has been, more generally defined as a device for coping with the freedom of others (see Luhmann 1979; Dunn 1984).

The rulers of a slave society - assuming that they do not mind what slaves *think* - can restrict their trust in the slaves and in the viability of their society to the belief that the slaves are not going to commit mass suicide. They simply trust to the fact - not invariably borne out by historical evidence - that most humans, even under extreme conditions, have a preference ordering which ranks life before death. Here, trust must be understood in the limited sense of trusting the effectiveness of coercion as a motive for cooperation (see Williams, this volume). By contrast, trust becomes increasingly salient for our decisions and actions the larger the feasible set of alternatives open to others.

The freedom of others, however, is not by itself sufficient to characterize the conditions in which the issue of trust arises. Our relationship with people who are to some extent free must itself be one of *limited freedom*. It is a freedom in the sense that we have to have a choice as to whether we should enter into or maintain a potentially risky relationship: it must be possible for us to refrain from action. If it were only others who enjoyed freedom, while we had no alternative but to depend on them, then for us the problem of trust would not arise: we would hope rather than trust (see Hart's discussion of *reliance*, and Luhmann's of *confidence*, this volume). The freedom is limited in the sense that if our feasible set of alternatives is *too* large the pressure to trust anyone in particular tends to be lower (see Lorenz's discussion of exit options in markets, this volume).

In conclusion, trusting a person means believing that when offered the chance, he or she is not likely to behave in a way that is damaging to us, and trust will *typically* be relevant when at least one party is free to disappoint the other, free enough to avoid a risky relationship, *and* constrained enough to consider that relationship an attractive option. In short, trust is implicated in most human experience, if of course to widely different degrees.

Cooperation frequently makes some demand on the level of trust, particularly of mutual trust (for the conditions under which this occurs cf. Williams, this volume). If *distrust* is complete, cooperation will fail among free agents. Furthermore, if trust exists only unilaterally cooperation may also fail, and if it is blind it may constitute rather an incentive to deception. However, depending on the degree of constraint, risk and <<220>> interest involved, trust as a precondition of cooperation can be subjected to demands of differing intensities: it may be *needed* to varying degrees, depending on the force of the mechanisms that govern our cooperative decisions in general and on the social arrangements in which those decisions are made.

When considered prescriptively, this conclusion further suggests that we can circumscribe the extent to which we need to trust agents or cope with them in case of distrust. A wide variety of human endeavour is directed towards this end: from coercion to commitment, from contracts to promises, with varying degrees of subtlety, mutuality, legitimation, and success, men and women have tried to overcome the problem of trust by modifying the feasible set of alternatives open not only to others, but also to themselves.

Coercion, or at least its credible threat, has been and still is widely practised as a means to ensure cooperation; in its extreme form, to ensure submission and compliance. But it falls short of being an adequate alternative to trust; it limits the extent to which we worry about, but does not increase trust. On the contrary: coercion exercised over unwilling subjects - who have not pre-committed themselves to being prevented from taking certain courses of action or who do not accept the legitimacy of the enforcement of a particular set of rights - while demanding less of *our* trust in others, may simultaneously *reduce* the trust that others have in us.<sup>11</sup> It introduces

---

<sup>11</sup> <<220>> This may establish some limit to the benefits of coercion to those who practise it, for as Veyne (1976) and Elster (1983) suggest, extremely successful and ubiquitous coercion can lead people to the extreme of making myths of their rulers.

an asymmetry which disposes of *mutual* trust and promotes instead power and resentment. As the high incidence of paranoid behaviour among dictators suggests, coercion can be *self-defeating*, for while it may enforce ‘cooperation’ in specific acts, it also increases the probability of treacherous ones: betrayal, defection, and the classic stab in the back.<sup>12</sup> (A more subtle way of constraining agents against their will is to enhance and exploit *their* mutual distrust. This has been known since antiquity as *divide et impera*, and is here explored in rich historical detail by Pagden; some of its consequences are taken up in my own paper on the mafia.)

Coercion does not have to be illegitimate, and may be employed for the purpose of enforcing rights which are commonly shared. In this case, instead of a unilateral action, coercion may itself be part of those cooperative arrangements intended to reinforce and reproduce a degree of trust in the observance of agreements previously reached with respect to those rights. But even if the controlled exploitation of coercive power <<221>> were considered legitimate, it would not generally constitute an exhaustive ‘functional equivalent’ of trust. It would still be true that societies which rely heavily on the use of force are likely to be less efficient, more costly, and more unpleasant than those where trust is maintained by other means. In the former, resources tend to be diverted away from economic undertakings and spent in coercion,<sup>13</sup> surveillance, and information gathering, and less incentive is found to engage in cooperative activities.<sup>14</sup>

Constraint is relevant not only for us in deciding how far we need to trust others, but also for others to decide how far they can trust *us*. It is important to trust, but it may be equally important to *be trusted*. *Pre-commitment*, in its various unilateral and bilateral forms, is a device whereby we can impose some restraint on ourselves and thus restrict the extent to which others have to worry about our trustworthiness (see Dasgupta, this volume). In the case of Ulysses it may have been used to combat lack of self-trust (Elster 1979; Dasgupta, this volume), but it is generally invoked to weaken the demand that our trustworthiness places on others. How effective it can really be is particular in the extreme, and the range of possibilities is far too wide to attempt to apply any general principle. Certainly pre-commitment can be positive if it sets external causes in motion: when two individuals keep keys to the same safe, for instance. But it can also be costly, and a cause of bitter regret: when one decides to wear a chastity belt and throw the key into the river. In general, pre-commitment acting on external causes might be defined as the mirror image of coercion, in that it more or less significantly shifts the problem of trust towards a small subset of options: the banker need *only* trust his partner not to murder him or rob him of his key, and the departing lover need *only* be convinced that there is no second key.

Contracts and promises represent weaker forms of pre-commitment, which do not altogether rule out certain actions, but simply make them more costly. Contract shifts the focus of trust on to the efficacy of sanctions, and either our or a third party’s ability to enforce them if a contract is broken. Promises are interesting in that the sanctions they imply may themselves take the form of trust: ‘When a man says *he promises any thing*, he in effect expresses a *resolution* of performing it; and along with that, by making use of this *form of words*, subjects himself to the penalty of never being trusted again in case of failure’ (Hume [1740] 1969: 574).

<<222>>

---

<sup>12</sup> <<220>> As I explain in my paper on the mafia, in spite of their constant efforts to over-determine the motives for cooperation, mafiosi are often obsessed with betrayal and deception.

<sup>13</sup> <<221>> See Brenner (1986) for an account of the reasons why in pre-industrial societies the economic surplus tended to be invested in improving extra-economic coercion.

<sup>14</sup> <<221>> Some of the cumbersome aspects of Italian *vis à vis* British bureaucracy can be explained by the fact that the former invariably starts from the assumption that the general public are an untrustworthy lot whose every step must be carefully checked. Quite apart from the likely effect of self-fulfilment, there is the suspicion that the cost of running such a system may far outweigh the cost of even a large amount of cheating.

As contracts and promises suggest, the relevance of trust in determining action does not only depend on constraint; it is a matter, in other words, not just of feasible alternatives, but also of *interest*, of the relative attraction of the feasible alternatives, the degree of risk and the sanctions they involve. The importance of interest is twofold: it can be seen to govern action *independently* of a given level of trust, but it can also act on trust itself by making behaviour more predictable. The former applies when we consider trust - in the sense of a given value  $p$  of the probability - as an assessment prior to an assessment of other people's interests, while for the latter to hold there has to be some degree of information about the interests of others. Here I shall engage in a preliminary consideration of the first case, postponing the more complex second case to the next section.

If we assume an *a priori* estimate of the probability that a person will perform a certain action - which is to say a given degree of trust, predicated on whatever evidence (friendship, membership of a group, style of clothing) other than the interests of that person - the question is: how high does that probability have to be for us to engage in an action the success of which depends on whether the other person or persons will act cooperatively? The answer is that the *optimal threshold* of the probability of believing we trust someone enough to engage in such action will not be the same in all circumstances. In this sense, actions which are dependent on other people's cooperation are *independent* of trust: for any given level of trust, they may or may not be initiated depending on our particular predispositions and interests. That is, we can not only expect the threshold to vary *subjectively*, as a result of individual predispositions (one's inclination to take risks or degree of tolerance of potential disappointment); we can also expect it to vary in accordance with *objective* circumstances (Good, Luhmann, this volume). For example, it will be higher when the costs of misplacing trust are potentially higher than those of not granting it at all and refraining from action: to walk about in a trench in sight of the enemy (to pick up the example discussed by Axelrod 1984) requires an extremely high degree of trust that the enemy will observe the implicit truce, and the costs of being wrong may prove much more serious than those of lying low. We may attach a certain value  $p$  to the probability that someone is trustworthy, but if he has a gun - or the atomic bomb - this will make a considerable demand on the value of  $p$  for us to act. Here the pressure *not* to trust, to let only a very high threshold of trust govern our action, is strong (see Good, this volume). By contrast, if we expect an action - subject for its success to a cooperative response - to yield higher returns than the alternative options, what we stand to lose from 'playing defection' may be great enough for us to proceed even when  $p$  is small.

<<223>>

Here the pressure to accept even a low level of trust as the basis of cooperative action is stronger. We may have to trust *blindly*, not because we do not or do not want to know how untrustworthy others are, but simply because the alternatives are worse.

An interesting case arises where there is *tension* between the intensity of our interest in acting and the value of  $p$ . We may either not know whether to trust, or even know that we distrust somebody, but if we were to refrain from engaging (or attempting to engage) in cooperation, our losses could nevertheless be unacceptably high: opting out is a feasible option, but we would pay for it dearly. The option of avoiding or exiting from a relationship, for instance, may be present with respect to any one individual agent, but not when considered in aggregate terms: we may have a choice as to which restaurant we assign a high enough probability of not giving us food poisoning, but there are circumstances in which we can hardly afford to distrust *all* restaurants without perhaps suffering unpleasant consequences. The pressure to lower the trust threshold and pick *one* of them is substantial. Moreover, if we do not have any firm idea as to whether we can trust a particular individual - the probability is set at 0.50 - we may do away with the problem by choosing at random. This too, as in the case of forced reliance on someone else, is more a matter of hope than trust - at least in the first instance, when we have no evidence either way.



If the pressure to act is great even when the trust threshold is lower than 0.50 - when we verge, that is, on distrust - the tension between action and belief can generate, by means of wishful thinking and the reduction of cognitive dissonance, a deceptive rearrangement of beliefs. Thus there are those who distrust entire categories of people except the member of that category with whom they have a special relationship. Da Ponte makes this point clearly in Mozart's *Così fan tutte*. Don Alfonso claims: 'E' la fede nelle femmine come l'Araba Fenice, che vi sia ciascun lo dice, dove sia nessun lo sa.' And the lovers respectively reply: 'La Fenice è Dorabella' and 'La Fenice è Fiordiligi'. They both boldly and in mutual contradiction claim that all women are unfaithful except their fiancées.

In conclusion, the above examples tell us not *how* a certain level of trust is reached, but only that once reached it may be effective for action yielding potential cooperation, in different ways depending on the constraints, and on the costs and benefits presented by specific situations. Clearly, the higher the level of trust the higher the likelihood of cooperation, but cooperative behaviour does not depend on trust alone, and the optimal threshold of trust will vary according to the occasion. In addition, the last example indicates that the tension between what we need and what we believe may be strong enough to generate irrational, <<224>> fideistic responses. Confidence, in the sense defined by Luhmann in this volume, might be described as a kind of blind trust where, given the constraints of the situation, the relationships we engage in depend or are seen to depend very little on our actions and decisions. In other words, confidence may also issue from wishful thinking and the reduction of cognitive dissonance; it would then be more akin to hope than trust. We still know little, though, about how a certain level of trust is or can be achieved and promoted.

### III

The first question is, why bother? Why should we bother about trust at all when cooperation can be generated by other means? One solution is in fact that of not bothering, and concentrating instead on the manipulation of constraints and interests as those conditions of cooperation on which we can intentionally and most effectively operate. We can aim to promote as much cooperation as possible by deploying some reasonable degree of coercion and by supporting arrangements which encourage cooperation through self-interest, thereby making small demands on trust (for a successful case see Hawthorn, this volume and for the limits of this approach see Dasgupta, Dunn, Lorenz, Hawthorn, and Williams, this volume). If we are lucky enough to live in a society which holds some moral and religious beliefs - a side effect of which is to motivate cooperation for its inherent virtues - we can make good use of them. But we cannot count on these being readily available.

This is not just *a* solution: it is possibly *the* standard solution, which, filtered through Machiavelli, Hobbes, Hume, and Smith, has been handed down to the present day as the most realistic, economical, and viable. Trust - like altruism and solidarity - is here deemed a *scarce resource*. Elster and Moene make extremely clear the argument for this solution with respect to economic reform (1988):

Indeed, some amount of trust must be present in any complex economic system, and it is far from inconceivable that systems with a higher level of general trust could come about. It would be risky, however, to make higher levels of trust into a cornerstone of economic reform. We may hope that trust will come about as the by-product of a good economic system (and thus make the system even better), but one would be putting the cart before the horse were one to bank on trust, solidarity and altruism as the preconditions for reform.

There is much to be said for this strategy - which might come under the heading of *economizing on trust* - and on the whole the papers in this <<225>> volume are not opposed to it, nor do they necessarily say that it does not work. They do, however, raise three important points which diverge from this approach. The first is that trust is *not* scarce in the sense of a resource that is depleted through use; the second that although it is often a by-product, this is not always so; and the third that there are extremely important cases where self-reinforcing arrangements

acting on interests are either too costly (or unpleasant) to implement, or unavailable in the first place because trust is in excessively short supply. I shall expand these points shortly, but first we need to take a step backward.

The most economical strategy of all is not that of not having to bank on trust, but that of not even having to bank on manipulating cooperative arrangements, on the assumption that rational - in the sense of optimal - cooperation evolves by itself. Pat Bateson's contribution to this volume explores some forms of cooperation in the animal world. The existence of cooperation among animals seems to suggest that cooperation may evolve without necessarily postulating trust, a belief which animals are unlikely to entertain. As Bateson argues, the emergent behaviour of social groups may contribute to their success: some of the features which make individuals successful in evolution may do so by working in conjunction with features developed by other individuals in the same group. Whether or not a group survives, in other words, depends on the emission and reception of signals which foster cooperation, in so far as cooperation improves the adaptive features of a particular group.

When transferred to the human world, the evolutionary approach might *a fortiori* suggest that trust would be better understood as a result rather than a precondition of cooperation. Trust would exist in societies and groups which are successful because of their ability to cooperate, and would consist in nothing more than trust in the success of previous cooperation. Cooperation could be triggered not by trust, but simply by a set of fortunate practices, random at first, and then selectively retained (with varying degrees of learning and intentionality).<sup>15</sup>

The idea that trust might follow rather than precede cooperation is reinforced by work in game theory, such as that of Axelrod (1984), which shows that even where trust is very limited and the chance of communication very slim - as between enemies facing each other across the trenches - cooperation may still evolve if other conditions obtain. The conditions Axelrod indicates with reference to a repeated Prisoner's Dilemma are (a) that the parties involved cannot escape confrontation <<226>> (they have a choice only as to whether they cooperate or compete); (b) that they know they will be locked in this situation for a long period, the end of which is unknown; and finally (c) that they have a low enough discount rate of future benefits. Under these conditions - even if the parties involved cannot commit themselves in advance, cannot monitor the relevant behaviour of the other party before the event, and do not have the slightest prior notion of whether and to what extent they can trust each other - cooperation may still be triggered by a random 'signal' which is then retained due to the success of its consequences.

Take, for instance, the cooperation of 'live and let live' which flourished between enemy soldiers in the First World War (Axelrod 1984). This might be accounted for in various ways. It may have arisen as the result of a soldier shooting - out of distraction, boredom, or nervousness - at some clearly non-human target in the opposite trench. Or it may be that soldiers on both sides stopped shooting at regular intervals because they happened to have their meals at the same hours of the day. Such random signals may eventually have been 'interpreted' by one side as an inclination on the other towards an implicit truce, and they may have responded with other signals; first just to test for possible misunderstandings, then with increasing conviction, until the exchange slowly assumed the features of a stable cooperative abstention from mutual injury. Furthermore, what may have emerged accidentally can subsequently be *learned*, and soldiers apparently learned several ways of signalling to the 'enemy' their predisposition to cooperate.

It is not so much that trust is not involved here, as that it would not seem to be a precondition of cooperation. In essence, the conditions lie partly in the objective circumstances, and partly in the accumulation of knowledge with reference to mutual interests and the potential satisfaction of

---

<sup>15</sup> <<225>> Hayek (e.g. 1978) should probably be regarded as one of the theoretical fathers of this view. Nelson and Winter (1982) too, though they do not mention trust as a feature of the economic success of individual entrepreneurs, would probably agree that cooperative arrangements and trust might at times be among the successful entrepreneurial practices positively selected by markets.

those interests through cooperative behaviour. The probability that the other party will not act in a harmful way is raised by the understanding that mutual interest makes defection costly enough to be deterred. The soldiers' belief that they can trust each other results from an inference concerning the effectiveness of their interests as a motive for rational action in initiating and maintaining cooperation. Thus  $p$  is raised in the course of cooperation itself, without any assumption being made as to prior levels: 'The cooperative exchanges of mutual restraint actually changed the nature of the interaction. They tended to make the two sides care about each other's welfare' (Axelrod 1984: 85). In the previous section we saw that interest, irrespective of trust, could make cooperation more likely simply by making action more pressing. On closer inspection we find that when that pressure is commonly shared *and* this fact is known to both sides, then cooperation is motivated and trust itself may increase as a result.

<<227>>

Hume made the case for this process absolutely clear:

When each individual perceives the same sense of interest in all his fellows, he immediately performs his part of any contract, as being assured that they will not be wanting in theirs. All of them, by concert enter into a scheme of actions, calculated for common benefit, and agree to be true to their words; nor is there anything requisite to form this concert or connection, but that every one have a sense of interest in the faithful fulfilling of engagements, and express that sense to other members of the society. This immediately causes that interest to operate upon them and interest is the *first* obligation to the performance of promises.

Afterwards a sentiment of morals concurs with interest, and becomes a new obligation upon mankind (1969: 574).

Axelrod's work suggests that this 'concert' may issue even from the most minimal chance 'to express that sense of interest', in a situation as apparently uncondusive to cooperation as that of war, where agents, if anything, are more likely to distrust each other.

There are several problems with respect to generalizing the evolutionary approach. By this I do not mean the objection that it is difficult to generalize from experiments where each agent is represented by one computer strategy, or that the conditions fostering a cooperative outcome - such as having some knowledge of each other's interests - are sometimes hard to meet (cf. Williams, this volume). Even discounting these, I would still argue that - with respect to the time span of reasonable relevance to any given generation - the spontaneous evolution of a cooperative equilibrium among humans is only *just as likely* as that of a non-cooperative one, unless some restriction is imposed on agents' beliefs.

Although Axelrod claims that cooperation can evolve without trust, the strategy of tit for tat (according to his experiments the optimal one in playing the Prisoner's Dilemma) is inconceivable in relation to humans without at least a predisposition to trust: when the game has no history a cooperative first move is essential to set it on the right track, and unconditional distrust could never be conceived as conducive to this. If one group of soldiers for some reason believes itself to be facing a mob of unrestrained warriors and trusts neither the latter's time preferences nor their rationality, 'peaceful' signals are more likely to be interpreted as a trap. This problem may be circumvented by assuming the presence of uncertain beliefs and a random distribution which accommodates the probability of the right initial move being made and being 'correctly' interpreted. Yet there is no reason why the appropriate conditional beliefs should typically be the case, and the optimal move may be hard to come upon by accident (while we may not want to have to wait for it to <<228>> come upon us). If it is true that humans are characterized by a lack of fine-tuning and a tendency to go to extremes (Elster 1983), the assumption that trust will emerge naturally is singularly unjustified.

Nor would a unilateral belief in blind trust be less damaging, for to protract trusting moves in the face of another's defection could lead to disaster rather than cooperation. In other words, if

one inputs into Axelrod's experiments either unilateral blind trust or blind distrust - whether unilateral or bilateral - the game will not culminate in a cooperative solution: the latter depends on the absence of these two lexicographic inclinations and the presence of a basic disposition towards *conditional* trust. The optimality of tit for tat as a strategy in playing a repeated Prisoner's Dilemma and generating a cooperative outcome also suggests that distrust should not be allowed to congeal after defection from the other side. It would be best to have no memory,<sup>16</sup> to forget that past defections may be repeated in subsequent moves. But in the case of individuals who *do* have some (finite) amount of memory, this is to prescribe as rational the avoidance of unconditional distrust even *after* the evidence would suggest it to be advisable. One inflicted 'wound', in this argument, should not be considered a sufficient reason for prolonged retaliation once the other side has been tamed by the first retaliatory response. Some degree of revealed preference for defection, in short, does not mean the absence of an interest in cooperation.

Cooperation is conditional on the belief that the other party is not a sucker (is not disposed to grant trust blindly), but also on the belief that he will be well disposed towards us if we make the right move. Thus tit for tat can be an equilibrium only if both players believe the other will abide by it, otherwise other equilibria are just as possible and self-confirming. To show that trust is really not at stake, Axelrod should have shown that *whatever* the initial move and the succession of further moves, the game tends to converge on tit for tat. What he does do is express a powerful set of reasons why - under certain conditions and even in the absence of trust generated by friendship, say, or religious identity - a basic predisposition *to trust* can be perceived and adopted as a rational pursuit even by moderately forward-looking egoists. We learn that, tentatively and conditionally, we can trust trust and distrust distrust, that it can be rewarding to behave *as if* we trusted even in unpromising situations.<sup>17</sup>

We might, of course, conceive of different moves - cooperative or defective - as simply appearing in random succession until the optimal blend is found and spreads through generations selected on that basis.

<<229>>

We might even agree that, in a certain number of billion years, the universe is likely to contain only those planets - or underdeveloped countries - whose inhabitants happened to hit on the right sequence of cooperative moves and to behave, to just the right extent, *as if* they trusted each other. On the other hand, we would also like Earth - and *our* children - to be among them. Evolution has bestowed upon us the mixed blessing of being able to generate intentionally the *as if* behaviour. Knowing this, we can hardly avoid the responsibility of considering trust a choice rather than a fortunate by-product of evolution.

#### IV

The strategy of economizing on trust does not of course imply that we should wait for cooperation to evolve by itself; it just claims that we should set our sights on cooperation rather than trust. We should, in other words, promote the right conditions for cooperation, relying above all on constraint and interest, without assuming that the prior level of trust will eventually be high enough to bring about cooperation on its own account. In a 'dissenter's confession', Hirschman (1984b) - although he disagrees with taking the strategy of economizing on trust too far (1984a) - makes an important case for breaking vicious circles in underdeveloped countries by implementing systems with technologically severe constraints capable of generating, rather than merely presupposing, trust: 'According to my way of thinking the very attitudes alleged to

---

<sup>16</sup> <<228>> Strictly speaking, to have one-period memory and act only in response to the other player's last move.

<sup>17</sup> <<228>> For an excellent example see the collection of essays edited by Kenneth Oye (1986).

be preconditions of industrialization could be generated on the job and “on the way”, by certain characteristics of the industrialization process’ (p. 99).<sup>18</sup>

However, attractive as this strategy may be, it begs the crucial question of what is to be done either when  $p$  is so low that conditions suitable for cooperation are not available in the first place, or when  $p$  is not high enough to sustain potentially beneficial cooperation where those conditions are too complex, costly, or unpleasant to be a conceivable alternative to trust. The arms race might fall under the latter heading (see Hinde 1986), and most under-developed countries under the former (see Arrow 1972: ‘Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence’).<sup>19</sup>

<<230>>

What is at issue is not the importance of exploring in greater depth the causality of those forms of cooperation which are independent of trust, but the fact that *economizing on trust* is not as generalizable a strategy as might at first appear, and that, if it is risky to bank on trust, it is just as risky to fail to understand how it works, what forces other than successful cooperation bring it about, and how it relates to the conditions of cooperation. Considering the extremely limited literature on this crucial subject it seems that economizing on trust and economizing on understanding it have been unjustifiably conflated.

Motives for cooperation and factors other than trust that render those motives effective must be taken into account, and further research in line with Axelrod’s study is likely to prove fruitful in the very near future.<sup>20</sup> Good, in this volume, considers experimental evidence which suggests the fundamental importance of long-term arrangements, of the absence of potentially aggressive devices, of the lack of ambiguity in what people cooperate about, and of a step by step increase in the risk involved in cooperation. Each of these conditions, by affecting constraints and interests, can also affect cooperation irrespective of a given level of trust, and when successful can serve to reinforce trust itself. Yet as Williams shows (this volume), such conditions cannot be assumed: formal structures and social reality have a distressing tendency to diverge, sometimes sharply. It is in the space of this divergence that trust must be wedged, and if *la fortuna* does not help, then intentionality based on the *as if* game must be brought into play. The issue now is the extent to which intentionality *can* be invoked.

*Prima facie*, trust would seem to be one of those states that cannot be induced at will, with respect either to oneself or to others. In the former case this is because rational individuals cannot simply decide to believe that they trust someone if they do not (Williams 1973); in the latter because they cannot easily set out intentionally to impress someone of their trustworthiness (see Elster 1983: 43: ‘[They are states which] appear to have the property that they can only come about as the by-product of actions undertaken for other ends. They can never, that is, be brought about intelligently or intentionally, because the very attempt to do so precludes the state one is trying to bring about’). There is a sense in which trust may be a by-product, typically of familiarity and friendship, both of which imply that those involved have some knowledge of each other and some respect for each other’s welfare. Similarly, trust may emerge as a by-product of moral and religious values which prescribe honesty and mutual love.

<<231>>

---

<sup>18</sup> <<229>> See also Hirschman (1967 and 1977). Hawthorn discusses the limits of this approach in this volume.

<sup>19</sup> <<229>> On the relevance of trust for economic development see also Banfield (1958) and Mathias (1979).

<sup>20</sup> <<230>> In his more recent work Axelrod stresses much more forcibly the crucial importance of beliefs and of the interpretations of situations that actors entertain (see Axelrod and Keohane 1986).



Personal bonds and moral values can only function as encouragements to action and cooperation if they are concepts in which we believe. Motive and belief belong together here: they are part of a person's identity, they arise out of his passions and feelings. It is constituent of the definition of personal bonds that - within the limits of character and skill - one trusts one's friends more deeply than strangers (see Hart and Hawthorn).<sup>21</sup> If it were not, the motive for acting on the basis of personal bonds would vanish. A similar argument applies to values: one can hardly act, say, out of fear of God, if one does not have faith in his existence.

Furthermore, neither of these potential sources of trust can be brought about at will: I cannot will myself to believe that X is my friend, I can only believe that he is. Nor can they easily be manipulated *in order to* bring about mutual trust and fruitful cooperation: if X detects instrumentality behind my manifestations of friendship, he is more likely to reject me and, if anything, trust me even less (Elster 1983). Thus we may explain the genesis of trust as an effect of moral and religious feelings - a point Weber notoriously pursued - or we may invoke the bonds of friendship as its prime source. But as rational individuals we cannot expect to induce these feelings simply because they may be useful, nor can we build our lives on the expectation of fooling ourselves or others systematically or for any length of time. It would seem to be a matter of social luck should such feelings happen to exist, and we cannot trust to luck. Moreover, personal bonds and values cannot be trusted as the foundation of cooperation in complex societies, not just because they cannot be wilfully generated, but also because of their fragility in a disenchanted world, and because they necessarily operate on a limited scale: I may believe in God, but I can still have my doubts as to whether you do too; even in Iran religious faith is backed by massive coercion, just to remind those who misbehave, by way of appetizer, what God's punishment will be like.

Yet at one extreme this view clashes with the rich historical and anthropological catalogue of cases where the bonds of friendship and familiarity have been extended in ritualized and codified forms - sometimes far beyond the socially narrow limits of 'true' friends and believers (Eisenstadt and Roniger 1984) - and where people, relying on limited and even truly scarce resources of familiar bonds, have acted on the fiction of <<232>> the *as if* behaviour to the point where skilful and intentional pursuit can hardly be told from the random appearance of more or less fortunate practices (see Gellner, Hart, Hawthorn, and Lorenz, this volume). This applies equally to the 'good guys' and the 'bad', to the Florentine bankers of the late Middle Ages (Becker 1981) and to the mafia networks of 'friends of friends'. Fragile as these 'aristocracies' - as Hawthorn calls them - may be, always in need of external reinforcement both by sanctions and by success (in satisfying at least some people's interests), they often represent the only means of instigating a cooperative relationship that would otherwise fail because of uncertainty or frozen distrust. Among the rotating credit associations of Mexicans, studied by Velez-Ibanez (1983), there is an explicit 'cultural construct' known as *confianza en confianza*, trust in mutual trust: within the boundaries of 'aristocracies', trust is set to a value which is high enough for tentative cooperation not to be inhibited by paralysing suspicion.<sup>22</sup>

Conceptual and theoretical arguments likewise suggest that the maintenance of trust via the extension of associations based on personal bonds might be seen to involve an element of rational pursuit. Trust, although a potential spin-off of familiarity, friendship, and moral values, must not be confused with them, for it has quite different properties (Luhmann, this volume). It is perhaps such confusion which has led to the conclusion that we should treat it as both an eminently scarce resource and purely a by-product. This does not take adequate account of our

---

<sup>21</sup> <<231>> Trust in friends very much depends on what we need to trust them about. In pre-modern societies trust and friendship belonged together much more extensively. To the extent to which modern society has relaxed our dependence on friends for acquiring and maintaining resources, and has diffused and 'specialized' our relationships beyond localized familiarity, it has also given us a greater freedom to include among our friends some highly unreliable persons. The point though is that we are free not to depend on them. With respect to certain actions, we may actually trust others much more than our *friends*. For an extensive discussion of friendship and trust see Silver (1985).

<sup>22</sup> <<232>> See also Coleman (1984) who refers to similar associations in South-East Asia and Japan.

ability to act, to simulate, try out, learn, apply and codify signals and practices which may initially be predicated on unintentional states, but which could be duplicated in the *as if* behaviour form far beyond their source. Trust, of course, can take on the connotations of a passion (Dunn, this volume; Mutti 1987), reinforced or undermined by feelings of affection, dislike, and irrational or intuitive belief (none of which can be induced at will). To economize on the latter - to assume that the right set of feelings and beliefs may simply not obtain - may be justified. But even in the absence of that set and of the 'thick' trust which may accompany it, and even if we do not believe in the social viability or in the moral desirability of those 'aristocracies' which bank on them (cf. Williams, this volume), it still does not follow that we should economize on trust or relegate it to the status of by-product.

The limits of this approach are probably more interesting than the approach itself. There is a wide range of anecdotal, historical, and socio-psychological evidence to suggest that our capacity for self-delusion far exceeds rational optimistic expectations, and that we can indeed make ourselves and others 'believe'. Trust is of historical interest to us <<233>> precisely in those cases where it is misplaced: it could not exist without the betrayal (Shklar 1984), deception, and disappointment our foolishness sustains. But let us start from the charitable assumption that the typical case is that of persons who tentatively adopt rational strategies in the formation of their beliefs and who present a degree of healthy resistance to fooling themselves and others. Fundamentally, we expect rational persons to seek evidence for their beliefs and to offer that evidence to others. Within limits (Williams, Lorenz, this volume), we can increase (or decrease) our *p* by gathering information about the characteristics and past record of others, and whenever the gaps left by asymmetric information and uncertainty appear detrimental to us, we can try to bridge them by rationally enhancing our reputation for trustworthiness, pre-committing ourselves, and making promises. A reputation for trustworthiness is not just tangential to a good economic system: it is a commodity intentionally sought by - and a constant concern of - anyone who aims at such (see Dasgupta, this volume; Akerlof 1984). Interest may generate the pressure to behave honestly, but reputation and commitment are the means by which others are assured of the effectiveness of that pressure: 'A dealer is afraid of losing his character, and is scrupulous in observing every engagement. When a person makes perhaps 20 contracts in a day, he cannot gain so much by endeavouring to impose on his neighbour, as the *very appearance* of a cheat would make him lose' (Smith [1723] 1978: 538-9, my italics).<sup>23</sup> Conditions favourable to honesty and cooperation - that is, a healthy economy - *and* the reputation for trustworthiness must reinforce each other for a 'concert of interests' to be played. It may be hard to bank on altruism, but it is much harder to avoid banking on a reputation for trustworthiness: as all bankers (and used-car dealers) know, a good reputation is their best asset.

However, if evidence could solve the problem of trust, then trust would not be a problem at all. It is not only that the gathering and exchange of information may be costly, difficult, or even impossible to achieve. Nor is it just that past evidence does not fully eliminate the risk of future deviance. The point is that trust itself affects the evidence we are looking for. While it is never that difficult to find evidence of untrustworthy behaviour, it is virtually impossible to prove its positive mirror image (Luhmann 1979). As Tony Tanner has suggested (personal communication), this aspect of the nature of trust is implicit in Shakespeare's *Othello*. Othello asks Iago for 'ocular proof' of Desdemona's unfaithfulness. He could not conceivably have asked for direct evidence of her fidelity: the only ocular proof of (at least future) fidelity is a dead body. And given Othello's expectations, it is all too easy <<234>> for Iago to 'find' the evidence required. Doubt is far more insidious than certainty, and distrust may become the source of its own evidence.

Trust is a peculiar belief predicated not on evidence but on the lack of *contrary* evidence - a feature that (as Pagden (this volume) shows) makes it vulnerable to deliberate destruction. In contrast, deep distrust is very difficult to invalidate through experience, for either it prevents people from engaging in the appropriate kind of social experiment or, worse, it leads to

---

<sup>23</sup> <<233>> I am grateful to Eduardo da Fonseca for bringing this passage to my attention.

behaviour which bolsters the validity of distrust itself (see my paper on the mafia, this volume). Once distrust has set in it soon becomes impossible to know if it was ever in fact justified, for it has the capacity to be *self-fulfilling*, to generate a reality consistent with itself. It then becomes individually 'rational' to behave accordingly, even for those previously prepared to act on more optimistic expectations. Only accident or a third party may set up the right kind of 'experiment' to prove distrust unfounded (and even so, as Good argues in this volume, cognitive inertia may prevent people from changing their beliefs).

These properties indicate two general reasons why - even in the absence of 'thick' trust - it may be rational to trust trust and distrust distrust, that is, to choose deliberately a testing value of  $p$  which is both high enough for us to engage in tentative action, and small enough to set the risk and scale of possible disappointment acceptably low. The first is that if we do not, we shall never find out: trust begins with keeping oneself open to evidence, acting *as if* one trusted, at least until more stable beliefs can be established on the basis of further information.<sup>24</sup> The second is that trust is not a resource that is depleted through use; on the contrary, the more there is the more there is likely to be (see Dasgupta for a demonstration, this volume; also Bateson 1986). As Hirschman suggests (1984a; also Hirsch 1977), trust is depleted through *not* being used.

The latter can be taken to mean different things. Firstly, trust may increase through use, for if it is not unconditionally bestowed it may generate a greater sense of responsibility at the receiving end. When we say to someone: 'I trust you', we express both a belief in and an encouragement to commitment by the trust we place in the relationship (Mutti 1987). The concession of trust, that is, can generate the very behaviour which might logically seem to be its precondition.<sup>25</sup> Secondly, if behaviour spreads through learning and imitation, then sustained distrust can only lead to further distrust. Trust, even if *always* misplaced, can never do worse than that, and the expectation that it might do at least marginally better is therefore plausible. However, while the previous reasons <<235>> can motivate rational individuals to trust - at least to trust trust - this reason alone cannot, for though everyone may concede it, if the risk of misplacing trust is reputed to be high, no one wants to be the first to take it. It is enough, however, to motivate the search for social arrangements that may provide incentives for people to take risks.

More generally, trust uncovers dormant preferences for cooperation tucked under the seemingly safer blankets of defensive-aggressive revealed preferences. True, there are cases - such as in Sicily - where these preferences, if they were ever awake, have been sleeping so long that they may well be dead, reduced to the ashes of a historical reduction of cognitive dissonance. If this were the case, we might just as well give up, abandoning the place to its fate or simply trusting to collective disaster to generate the reasons for change. But the point is that if we are not prepared to bank on trust, then the alternatives in many cases will be so drastic, painful, and possibly immoral that they can never be lightly entertained. Being wrong is an inevitable part of the wager, of the learning process strung between success and disappointment, where only if we are prepared to endure the latter can we hope to enjoy the former. Asking too little of trust is just as ill advised as asking too much.

#### REFERENCES

- Akerlof, G. 1984: *An Economic Theorist's Book of Tales*. Cambridge: Cambridge University Press.  
 Arrow, K. J. 1972: Gift and exchanges. *Philosophy and Public Affairs* 1, 4, 343-62.  
 Arrow, K. J. 1978: Uncertainty and the welfare economics of medical care. In P. Diamond and M. Rothschild (eds), *Uncertainty in Economics*, New York: Academic Press.  
 Axelrod, R. 1984: *The Evolution of Cooperation*. New York: Basic Books.

<sup>24</sup> <<234>> On the importance of acting as if for a solution of the Prisoner's Dilemma and related games cf. Sen (1974). On the 'suspension of distrust' see also Silver 1987.

<sup>25</sup> <<234>> On the self-fulfilling nature of beliefs cf. Schelling (1978).

- Axelrod, R. and Keohane, R. O. 1986: Achieving cooperation under anarchy: strategies and institutions. In K. Oye (ed.), *Cooperation under Anarchy*, Princeton: Princeton University Press.
- Banfield, E. C. 1958: *The Moral Basis of a Backward Society*. Glencoe: Free Press.
- Barber, B. 1983: *The Logic and Limits of Trust*. New Brunswick: Rutgers University Press.
- Bateson, P. P. G. 1986: Sociobiology and human politics. In S. Rose and L. Appignanesi (eds), *Science and Beyond*, Oxford: Basil Blackwell.
- Becker, M. 1981: *Medieval Italy*. Bloomington: Indiana University Press.
- Binmore, K. and Dasgupta, P. 1986: Game theory: a survey. In K. Binmore and P. Dasgupta (eds), *Economic Organizations as Games*, Oxford: Basil Blackwell.
- Brenner, R. 1986: The social basis of economic development. In J. Roemer (ed.), *Analytical Marxism*, Cambridge: Cambridge University Press.
- Coleman, J. S. 1984: Introducing social structure into economic analysis. *American Economic Review Proceedings*, 74, 84-8.

<<236>>

- Dunn, J. 1984: The concept of trust in the politics of John Locke. In R. Rorty, J. B. Schneewind, and Q. Skinner (eds), *Philosophy in History*, Cambridge: Cambridge University Press.
- Eisenstadt, S. N. and Roniger, L. 1984: *Patrons, Clients and Friends: interpersonal relations and the structure of trust in society*. Cambridge: Cambridge University Press.
- Elster, J. 1979: *Ulysses and the Sirens: studies in rationality and irrationality*. Cambridge: Cambridge University Press.
- Elster, J. 1983: *Sour Grapes: studies in the subversion of rationality*. Cambridge: Cambridge University Press.
- Elster, J. 1986: The norm of fairness. Chicago: unpublished paper.
- Elster, J. and Moene, K. (eds) 1988: *Alternatives to Capitalism*. Cambridge: Cambridge University Press.
- Hayek, F. A. 1978: The three sources of human values. L. T. Hobhouse Memorial Trust Lecture, The London School of Economics and Political Science.
- Hinde, R. A. 1986: Trust, cooperation, commitment and international relationships. Paper given at the meeting of Psychologists for Peace, Helsinki, August.
- Hirsch, F. 1977: *Social Limits to Growth*. London: Routledge and Kegan Paul.
- Hirschman, A. O. 1967: *Development Projects Observed*. Washington: Brookings Institution.
- Hirschman, A. O. 1977: *The Passions and the Interests: political arguments for capitalism before its triumph*. Princeton: Princeton University Press.
- Hirschman, A. O. 1984a: Against parsimony: three easy ways of complicating some categories of economic discourse. *American Economic Review Proceedings*, 74, 88-96.
- Hirschman, A. O. 1984b: A dissenter's confession. In G. M. Meier and D. Seers (eds), *Pioneers of Development*, New York, Oxford University Press for the World Bank.
- Hont, I. and Ignatieff, M. 1983: Needs and justice in the *Wealth of Nations: an introductory essay*. In I. Hont and M. Ignatieff (eds), *Wealth and Virtue: the shaping of political economy in the Scottish Enlightenment*, Cambridge: Cambridge University Press.
- Hume, D. [1740] 1969: *A Treatise of Human Nature*. Harmondsworth, Middlesex: Penguin Books.
- Luhmann, N. 1979: *Trust and Power*. Chichester: Wiley.
- Mathias, P. 1979: Capital, credit and enterprise in the Industrial Revolution. In P. Mathias (ed.), *The Transformation of England*, London: Methuen, 88-115.
- McKean, R. N. 1975: Economics of trust, altruism and corporate responsibility. In E. S. Phelps (ed.), *Altruism, Morality and Economic Theory*, New York: Russell Sage Foundation.
- Mutti, A. 1987: La fiducia. *Rassegna Italiana di Sociologia*, 2.
- Nelson, R. and Winter, S. 1982: *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Harvard University Press.
- Oye, K. (ed.) 1986: *Cooperation under Anarchy*. Princeton: Princeton University Press.

<<237>>

- Sen, A. 1974: Choice orderings and morality. In S. Koerner (ed.), *Practical Reason*, Oxford: Basil Blackwell.
- Schelling, T. C. 1978: *Micromotives and Macrobehaviour*. New York: Norton.
- Schelling, T. C. 1984: Strategic analysis and social problems. In *Choice and Consequence*, Cambridge, Mass.: Harvard University Press.
- Shklar, J. N. 1984: *Ordinary Vices*. Harvard: The Belknap Press.
- Silver, A. 1985: Friendship and trust as moral ideals: a historical approach. Unpublished paper, American Sociological Association meeting, Washington DC, 26-30 August.
- Silver, A. 1987: Friendship in social theory: personal relations in classic liberalism. Unpublished paper, New York: Columbia University.
- Smith, A. [1723] 1978: *Lectures on Jurisprudence*. Oxford: Oxford University Press.
- Smith, A. [1759] 1976: *The Theory of Moral Sentiments*. Oxford: Clarendon Press.
- Velez-Ibancz, G. 1983: *Bonds of Mutual Trust*. New Brunswick: Rutgers University Press.
- Veyne, P. 1976: *Le pain et le cirque*. Paris: Editions du Seuil.
- Weber, M. 1970: *The Protestant Ethic*. London: George Allen and Unwin.
- Weil, F. D. 1986: The stranger, prudence, and trust in Hobbes's theory. *Theory and Society*, 5, 759-788.
- Williams, B. A. O. 1973: Deciding to believe. In *Problems of the Self*. Cambridge: Cambridge University Press.