Universiteit Antwerpen
Universitaire Faculteiten Sint-Ignatius Antwerpen
Faculteit Toegepaste Economische Wetenschappen
Vakgroep Ontwikkelingsstudies

HUMAN NATURE AS WE KNOW IT

Inter-and intrapersonal perspectives on opportunism and commitment

Tom De Herdt

October 1999

Abstract

In the New Institutional Economics, an argument has been made to found theoretical models on "human nature as we know it". Though this argument has served to defend the assumption of opportunism, it can also be used to explore other models of man. In particular, we scrutinise what kind of intuitively convincing assumptions can be made about the human mind so as to allow for the theoretical existence of social norms. The paper is structured as a review of solutions to the "trust-game". First, this game is proposed, and compared to other, related games. Then, we scrutinize some common solutions to the game, and the (explicit or implicit) models of man they build on. Third, we explore a solution which combines temporary preference theory and the classical Smithian insights about the social embeddedness of men. Finally, we illustrate the argument by reviewing some well-known arguments in development economics.

Contents

Introduction[1]

In one of the first attempts to define the "New Institutional Economics" (NIE) as a new economic sub-discipline, Oliver E. Williamson argued that, unlike in many of the conventional approaches, the NIE "expressly acknowledges the importance played by *human* factors in attempting to grapple with problems of economic organization" (1975:2). Williamson frequently (e.g. 1987:2-6, 1986:173) refers, at this point, to "human nature as we know it", a phrase he picked up in Frank Knight's *Risk, uncertainty and profit* (1946 [1923]:270).

In practice, however, "human nature as we know it" has been operationalized as "opportunistic" and "boundedly rational", though the latter is perhaps somewhat optional. Williamson himself e.g. concedes that "bounded rationality can take many forms", and that "The main ramification of bounded rationality, for purposes of studying economic organization, is that all complex contracts are unavoidably incomplete" (1993:11). In other words, "bounded rationality" is, according to Williamson, not so much characterizing human nature as the nature of *contracts* as we know them. So, Williamson knows human beings as opportunists. The conventional economist's rational egoist is just a pussycat if compared to the lier, the cheater and the subtle deceiver of the practitioners of the NIE.

To be sure, Williamson's argument is rather that "human nature as we know it" does not *exclude* opportunism: "I merely assume that some individuals are opportunistic some of the time and that differential trustworthiness is rarely transparent *ex ante*" (1987:64). In a world where (1) contracts are unavoidably incomplete, and hence where there is a *possibility* to shirk, and (2) where even only a small minority will act opportunistically even only some of the time, so the argument proceeds, it can be supposed that contracts will be designed in such a way that the risk of opportunism will be minimized. In other words, man is modeled as an "opportunist" only to verify a worst-case scenario: To see whether contracts are efficient, we can limit our analysis to the way in which the "opportunists" would react.

In what follows I broadly concur with Williamson that human nature as we know it does indeed *not exclude* opportunism. However, I don't think that this would imply that positive predictions about institutional form and evolution should be based on the assumption of opportunism. Charles Sabel (1995:65) recently cited Talleyrand's dictum that the most suspicious people make the biggest dupes, arguing that some current strands in NIE might be among them. In a similar vein, Sen (1976) accused rational beings (the economist's model rather than the real-life beings we daily see and talk to) of being rather foolish.

[1] This paper is the provisional end-result of an enquiry I started in 1994, which implies *inter alia* that the list of readers of intermediate products is fairly large. Special thanks go to Stefaan Marysse and the participants of the seminar on Social Capital at UFSIA (December 1998). The author remains fully responsible for remaining errors.

In order to make my argument, my focus will be on the *majority* of human beings: If it isn't their nature, what is it that makes them oftentimes cooperate even without being forced to do so? If it isn't their nature, what is it that makes others bet on their cooperation without having the material guarantees ? And why will some of this majority act opportunistically *only some of the time* ?

If it isn't *nature*, it must be *artifact*. Herbert Simon (1981) introduces the latter concept to denote all kinds of material and immaterial things which involve human design. In our case, the answer must come from those artifacts called *institutions*, or, to take an already classical definition, the "humanly devised constraints that shape human interaction" (North 1990:3). This definition might be question-begging: It is clear that in a world pervaded by opportunities to shirk, constraints would be very helpful, but it is not at all clear why rational people would abide by them, once they have been devised. **If, at the one hand, we interpret institutions as ingenious inventions of perfectly rational human beings, at the other we need to model human beings as somewhat naive, myopic rule followers,** even if the latter's stupidity seems to be of a different kind than that of Sen's rational fool.

According to some, this question has not been put on the table because there were, as it were, two tables: In economics, man is thought of as an intentional, wilful *homo economicus* whose creations could only be understood by reference to his intentions. *Homo sociologicus*, at the other hand, enacts internalized roles regardless of the consequences (Dupuy 1992:65, Elster 1989:97, Vanberg 1988:147, Granovetter 1985:485, 1992:6). Though the image is cute, I would hesitate to organize all economists and sociologists around "their" table. Coleman's (1990) actors seem to be far more rational, for example, than the "boundedly rational" people inhabiting Simon's organizational artifacts (1981). In any case, it is the solution to this question which will interest us, but being (or aspiring to be) an economist myself, I will depart from the economists' side of the table.

The chapter is structured as follows.

I begin with presenting the trust game. I propose to use this game, rather than the more well-known prisoners' dilemma, as an archetypical illustration of a social dilemma, i.e. a game where "individuals in interdependent situations face choices in which the maximization of short-term self-interest yields outcomes leaving all participants worse off than feasible alternatives" (Ostrom 1998:1).

If "social capital" can be defined as those circumstances which enable individuals to overcome the dilemma, we should distinguish, with Collier and Gunning (1999), *public* from *civic* social capital; while the latter is embodied in private social interaction, the former is embodied in institutions of government. None of these forms of capital is owned individually, but the former involves state action, while the latter has rather the characteristics of a collectively owned good.

In order to explore *public* social capital, I discuss North's concept of "third party enforcement". North's argument is that, given economies of scale in enforcement, a *neutral* state will be the best solution to social dilemma's. The emergence of *neutrality* is, however, difficult to theorize. North called it the *fundamental dilemma of economic development.* Williamson (e.g. 1975) and Kreps (1995) argue that third party enforcement is perhaps too costly an option in *some* cases, and hence they point to the potentially limited relevance of North's "solution". In international comparative perspective, however, such critiques are better considered as uninteresting details: If economic development is, broadly conceived, a question of cooperation and trust, those who attribute African decline to low "public social capital" have a Nobel prize winner to sustain their argument.

Section 3.3. presents some rudimentary models which try to say something about *civic* social capital, viz. what kind of it seems to be needed to sustain economic development as economists know it, and how we might model the dynamics of *civic* social capital.

To be sure, though both sections are meant to expose solutions to the trust game, they are at the same time question-begging. More in particular, I make explicit in what way "nonwealth-maximizing" determinants of human behaviour have been incorporated in economic thinking and why I find the result unsatisfactory.

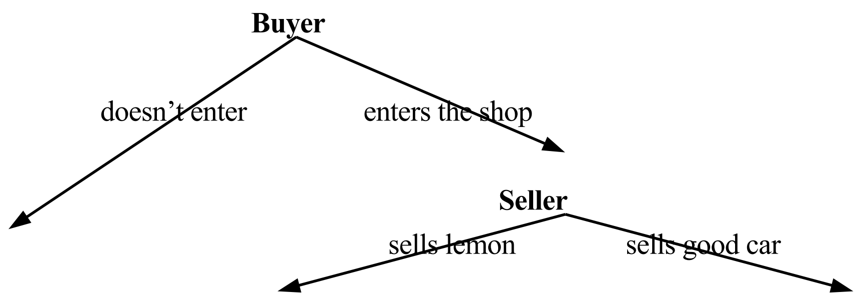
Section 3.4. concentrates on this problem. In a first step I introduce temporary preference theory, which is essentially an alternative and psychologically founded view of the way in which human beings articulate several distinct motives for action. The alternative we explore belongs to the family of economic theories which propose a dual preference structure (Hirschman 1984b, Sen 1976, Margolis 1982, Harsanyi 1977, Sabel 1993). In these theories, man is modelled as a reflexive person, someone who can make a difference between what he prefers and what he would like to prefer, but also someone who is not always successful in his attempts to better himself. This model of man seemed not only closer to my personal intuitions about human beings as I know them (to begin with myself) relative to the homunculi we reviewed above, it also connects well with empirical and theoretical contributions in other areas of the social sciences (e.g. as reviewed in Elster 1986). I use it as a basis to explore the ways in which nonwealth-maximizing motives can be incorporated into such a framework. Finally, I review the trust-game from this alternative perspective.

I conclude with discussing two well-known illustrations in the development literature, which will permit me to explore the interplay between *public* and *civic* social capital and to revisit North's *fundamental dilemma*.

3.1. The problem of credible commitment.

One of the fundamental assumptions made by the NIE is that economic exchange is costly. Exchange is costly because there is uncertainty about the value of the goods or services we buy or sell. Goods or services, or agents we hire, are not valuable in themselves; we get utility only from the diverse attributes of a good or service, or from the multitude of separate activities that constitute an agent's performance. And if goods, services and agents are readily observable, attributes and agents' activities are not. The problem is complicated by the presence of asymmetry in information: typically, the sellers of goods and services have "inside" information of the quality of the goods and services they sell (North 1990:30). Hence, buyers must dedicate resources to measure a particular asset's value. Those resources are wasted from the point of view of society. They are "burned" in the process, and thus they lower the gains to be made from trade.

In the form of an extended game structure, the problem of exchange in an anonymous market can be presented as what has become known as the *trust game* (Kreps [1986] 1996:231, Dasgupta 1988:60)[2]:



outcomes:

Buyer	0	$-\delta$	$1-\delta$
Seller	0	a	b

with $0<\delta<1$ and $0<b<a$

-- Figure 3.1. The trust game--

The game is exemplified by the market of second-hand cars. The customer can decide to enter the shop, in which case he must buy a car, or not enter at all. The seller can choose between selling him either a "lemon" or a good car. The pay-offs[3] are specified so that the seller prefers to sell lemons to good cars, and selling to not selling at all, while the buyer will prefer buying a good car to not buying, and not buying to buying a lemon. While it is clear that both parties can gain from transacting, once the buyer has entered the shop each party's interests are conflicting. Supposing

[2] See also the very similar game representing the "Time consistency problem" (Bates 1996:273). After the fact, it might have been better to copy Bates' name of the game, as it connects very well with the solution we will work out below, as well as with the problem of economic development more generally.

[3] The payoffs do not necessarily reflect private utilities; they can best be seen as having the function of ranking outcomes. Choices will conform to these rankings, but the rankings themselves "can be based on as complex a set of considerations as we would like to include in our model of the choosers" (Dasgupta 1988:61)

the potential buyer has full information over the game structure, and supposing he knows nothing about the seller except his payoffs, he will not enter the shop, and autarchy will be the outcome. The selling and buying of a good car will, however, be a (non-equilibrium) beneficial outcome for both. The example makes clear what is at stake: exchange will not take place, and thus the (Pareto) optimum will not be reached, unless the seller's opportunity to behave opportunistically is *credibly* curbed, in one way or another.

Before examining the several solutions to this problem, we'd like to discuss the appropriateness of the game as specified above.

First, the economic literature about social capital almost invariably invokes the "prisoner's dilemma" (PD) in order to discuss the twin problems of trust and absence of opportunism (e.g. Putnam 1993, Platteau 1994, Basu 1995). The trust game is, in fact, a one-sided version of this game: in PD's, both players need both to have trust and to be co-operative, in our game both players are confronted with only one of these problems: Looked at it from the seller's side, the problem of exchange in an anonymous market is a problem of curbing opportunism: why is it that the seller will not opt for maximising his interest ? Second, looked at it from the buyer's side, the problem of exchange is a problem of trust; even if a seller will really be committed to sell a good car, he must still be able *convince* the buyer. Both problems of opportunism and distrust need be solved in order to obtain the co-operative outcome, but they should be carefully distinguished, even if they are often causally related in theoretical models, and most probably also in practice. If not, the following story, as told by Sen, would not appear so absurd: "'Where is the railway station ?' he asks me. 'There', I say, pointing at the post office, 'and would you please post this letter for me on the way ?' 'Yes', he says, determined to open the envelope and check whether it contains something valuable" (Sen 1983 [1976]:96). Whether or not being too rational is foolish, the point of the story seems to be that it is foolish to have an irrational amount of trust in another's good conduct in a world populated by rational (in this case: opportunistic) actors.

Second, the trust game may perhaps represent an extreme situation in that the insecurity about the seller's move makes the exchange prohibitive: Unhonored trust would make the buyer worse off than before. A less extreme situation would be one where $\delta<0$: in that case, a rational buyer would invariably choose to enter the shop, since he will always be better off, whatever the move of the seller. For rational-choice theorists, a situation where $\delta<0$ would be trivial to discuss[4].

Third, it is a one-shot game. It is a folk theorem that "in settings where individuals know each other well, can observe one another's behaviours, and are in continuous interaction...any pattern of collective behaviour, including co-operation, can be sustained". Thus, if it is by no means necessary that, in those conditions, co-operation will be the likely result (Kreps 1990:506-512), it can at least be maintained once come into existence. To see this, suppose that transactions between the same buyer and the same seller are continuously and in(de)finitely repeated. In this

[4] See, however, Akerlof (1970) for an alternative analysis where $0<c<d$ is an important possibility.

case, and supposing the sellers' applied discount factor to actualize benefits from future dealings is sufficiently low, sellers will want to sell the good car, even if, in the short term, they will be better off selling a lemon. However, this kind of reasoning cannot solve the problem of social order underlying the free market society, because one of its characteristics is precisely *anonymity*. Micro-motives may be the basis for co-operation in traditional societies, within extended families or ethnic groups, but as long as the moral norm to cooperate doesn't exceed the boundaries of the group of (in-)direct acquaintances, this entails "a contradiction between two conditions of a market order: that required for mitigating opportunism and that of independence (or atomization) of the agents" (Platteau 1994:553). In other words, if there is an answer to this "impossibility theorem", as Platteau called it, we must be able to look for solutions, not of an in(de-)finitely repeated trust game, but of a one-shot trust game. These solutions must be sought in macro-motives, i.e. general motives that override the motives offered by the particular transaction itself.

3.2. Public social capital.

The most obvious mechanism to curb opportunism is commonly known as state intervention: Buyers can trust sellers to be honest because they know that dishonest sellers will be sanctioned by a state who guarantees *fair trade*: "Third-party involvement would involve a neutral party with the ability, costlessly, to be able to measure the attributes of a contract and, costlessly, to enforce agreements such that the offending party always had to compensate the injured party to a degree that made it costly to violate the contract" (North 1990:58). Indeed, the presence of a "neutral" party would be a guarantee for both parties that they can safely trade with eachother, even if they don't know eachother's identity. It must of course be admitted that the above description of the role of a third party is too ideal; in reality, third party monitoring and enforcement will never be costless and far from perfect. For these reasons private *institutional arrangements* will continue to exist in the midst of the anonymous market (Williamson 1987, 1993). It can be argued, however, that there are "immense scale economies in policing and enforcing agreements by a polity that acts as a third party and uses coercion to enforce agreements" (North 1990:58), and hence that the state has a crucial role to play to solve the problem of social order underlying the free market economy.

The problem with this solution is, however, that the state itself is also the result of co-operation. As we cannot assume that the state can explain itself, we cannot assume either that egoistic micro-motives can provide for an explanation, for "if the state has coercive force, then those who run the state will use that force in their own interest at the expense of the rest of society... How does one get the state to behave like an *impartial* third party ?" (North 1990:58, our emphasis). According to North, the conundrum of the creation of a neutral state out of private interests is "**the fundamental dilemma of economic development**". A way out of this conundrum can only be found if we introduce, some way or another, nonwealth-maximizing motives.

Suppose that agents are, albeit to different degrees, culturally predisposed to be honest. To be sure, they are still maximizing their utility functions, valuing costs against benefits. Only, their utility functions include not only considerations of self-interest, but also non-egoistic values like altruism or fairness. It is thus contended that, "while nonwealth-maximizing motivations influence choices", the degree to which choices are indeed influenced by these motivations depends on "the price to be paid for one's convictions" (North 1990:26)[5]. In our example, the difference between a and b is the price of honest behavior. If b-a approaches zero, nearly all sellers will behave honestly[6].

This is the theoretical foundation of North's theory of third-party enforcement: If the costs to be honest are too high in *economic* markets, the *political* market can be organized in such a way that people can express their conviction without any cost to themselves (b-a approaches zero). According to North, universal suffrage and a properly functioning parliament, in which there is an "incentive of an incumbent's opponents to bring his or her deficiencies before constituents and hence reduce agency problems" (North 1990:109) can indeed contribute to better outcomes.

Still, "there is a vast gap between better and efficient outcomes", because of two reasons that (can) prevent the political system to act for the common good. First, not all individuals have equal access to the decision-making process. The price paid to use *voice* can differ between different (groups of) actors. As a result, only "organizations with sufficient bargaining strength will use the polity to achieve objectives when the payoff from maximizing in that direction exceeds the payoff from investing within the existing constraints" (1990:79). Second, while individuals can be supposed to have an opinion about what is good for all, it cannot be taken for granted that their opinion will exactly coincide with what is *effectively* good for all[7], since, given environmental complexity and limited human processing capability, individuals lack the necessary information about the consequences of their choices, and must lean on subjective ideas, or ideologies. As an example, North cites the work of ECLA and dependency theorists, whose conclusions, according to North, not only rationalize the structure of Latin American economies, but, by promoting de-connection, also reinforce the existing pattern of underdevelopment in Latin America (North 1990:99-100).

Note that while North recognizes that rational actor models "are inadequate to deal with many issues confronting social scientists and are the fundamental stumbling block preventing an understanding of the existence, formation and evolution of institutions" (1990:24), he does not introduce an alternative behavioral theory. Instead, he contents himself with some simple hypotheses that will allow him to continue to use rational actor models. While he refuses to

[5] cost-sensitivity is also discussed by Williams (1988) and Elster (1989:187-8) but especially Elster (1998:64-7).

[6] Except of course the predators who derive pleasure from another's sorrow.

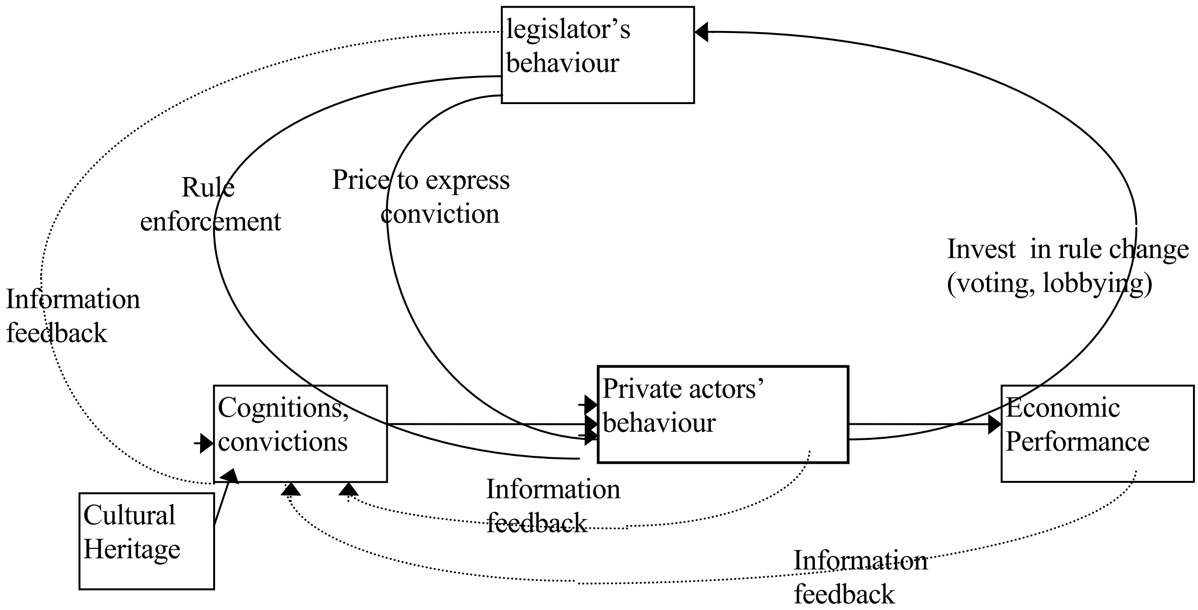
[7] For a contrasting view, see Harsanyi's idea of moral value judgments (1977), which is allegedly based on Adam Smith's idea that "the moral point of view is essentially the point of view of a *sympathetic* but *impartial* observer" (Harsanyi 1977:48-9). Adam Smith's moral theory is somewhat more sophisticated, however; see Jeffrey T. Young (1992).

formulate a theory about cultural heritage[8], this does not hinder him to discuss the role of individuals' convictions in an indirect way: "We can still take an important forward step by taking explicit account of the way institutions alter the price paid for one's convictions and hence play a critical role in the extent to which nonwealth-maximizing motivations influence choices" (1990:26). Unlike ordinary tastes, convictions are nonwealth-maximizing. Like ordinary tastes, however, expressing a conviction is as cost-sensitive as buying an apple. While he continues to set the individual, maximizing actor at the center stage of his analysis, this does not imply, however, that the outcome will always be optimal from the social point of view: Individuals' processing capacities are too limited to allow for a correct insight into the means-ends relationships governing the economy.

This methodological strategy has some important consequences for his view on institutional change. In developing societies[9], individuals' rights are enforced by formal institutions, set up by the individuals themselves. Formal institutions *may* have, however, partly created themselves, as they gradually lowered the price for expressing individuals' convictions. Individuals' convictions are largely exogenously determined (by cultural inheritance), though they can be partially refined by improved information feedback, made possible by a change in the institutional matrix. Note the (dialectic) circularity in the argument: Economic and political institutions and culture (comprising "cognitions" (mental constructs, ideologies) as well as "convictions" (values, opinions)) are to a significant degree self-reinforcing, and institutional change can only occur "at the margin". The same circularity can of course also lock a society into a stagnating path; in *underdeveloped* societies, actors don't get a third-party off the ground, or the existing state is organized in function of the rulers' private interests only. People will subsequently "rationalize" stagnation with theories (like the dependency theory) that will in turn prevent change towards a more productive path, and reinforce the pattern of stagnation.

[8]At most, he formulates some general idea about it: "Although we are not yet able to explain precisely the forces that shape cultural evolution, it is obvious that the cultural characteristics of a society change over time and that accidents, learning and natural selection all play a part" (North 1990:87).

[9]to be contrasted with underdeveloped, or stagnating societies, see below



-- Figure 3.2. Path-dependent institutional change **--**

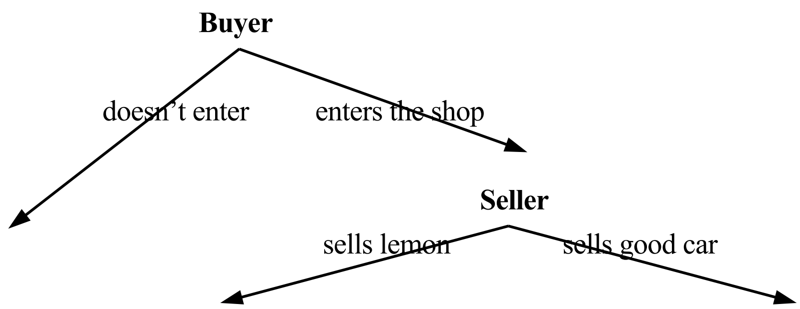
We have tried to represent this reasoning by a graphical schema. The circle of private actors' behaviour determining legislators' behaviour affecting private actors' behaviour (through informational feedback) may be either vicious or virtuous, locking-in societies in either stagnation or economic development. Nowadays, this type of circular reasoning has found relatively wide acceptance among the New Institutional Economists, under the heading of "path dependency" (see however Williamson 1994). The "old" institutionalists, like Veblen (and Myrdal) would have used the concept of "cumulative causality" (see Stanfield (1995) for a comparison between the "new" and the "old" institutionalism). For a game-theoretical account of this type of processes, see Kreps (1996).

As a result, North's theory is more apt to explain the *reproduction* of a developing society's institutional matrix than to explain its *emergence*. Entirely consistent with himself, he calls economic development "a fundamental dilemma". It could be argued, of course, that the path to economic development can be introduced exogeneously, e.g. by stimulating the polity to introduce universal suffrage. The problem is, however, that, while *formal* rules can change relatively quickly, human minds, private organizations and interest groups cannot be manipulated that easily. As a result, "the common imposition of a set of rules will lead to widely divergent outcomes in societies with different institutional arrangements" (1990:101).

Interestingly, Robert Putnam's study of institutional change in Italy seems to confirm this last insight. According to Putnam, economic as well as political performance are determined by a factor he calls "civicness". In the North of Italy, citizens conceive of the State as a "neutral third party", while, in the South, they do not. The obvious (Northian) way to "create" civicness in the South is reform of formal institutions. However, Putnam's study concludes that "where institution building (and not mere constitution writing) is concerned, time is measured in decades" (Putnam 1993:184)). A subsequent change in "civicness" can only be achieved over centuries. Putnam, as well as North, share the view that a society's *cultural heritage* is a crucial determinant of economic development. They also share the view that political institutions are partial determinants of belief systems, ideologies, etc. Putnam goes further than North, however, in that he attempts to fill in the blind spot in North's contribution. We discuss his contribution hereafter.

3.3. Civic social capital

Suppose now that *all* buyers *trust* second-hand car sellers to be honest, and that *all* second-hand car sellers *are* acting morally. The latter means, in terms of our model, that the sellers' ranking of payoffs are altered: selling a lemon, they will experience a feeling of guilt, and this negative sensation will more than neutralize the additional profit. As a result, there will be a market for second-hand cars, and the problem of opportunism is no problem at all. These situations are quite extreme, however. It is more realistic to begin with the assertion that, if we suppose agents to be opportunists, this "does not imply that most economic agents are opportunistic most of the time" (Williamson 1993:12). Thus, assuming the absence of a state but the presence of a society in which buyers transact consecutively with several sellers, the game structure is altered as follows:



outcomes:

Buyer	0	$-\delta$	$1-\delta$
Dishonest Seller	0	a	b
Honest Seller	0	f	g

with $0<\delta<1$, $0<b<a$ and $0<f<g$

-- **Figure 3.3.** Trust game with honest seller --

We have specified two types of sellers, the first being opportunistic, the second being honest, in the sense that the latter has a moral inclination towards selling reliable cars, so that $g>f$. From here, we can follow two different paths to model either the emergence of trust or the emergence of honesty.

3.3.1. Socialized honesty, experienced trust

The first model assumes honesty to be the product of a socialization process: people enter the socio-economic battlefield with ideas and convictions that have been "learned" during childhood, and which won't change during the rest of their lifetime. Hence, there are the "honest types" and the "opportunists". The model runs as follows: Suppose that "society is not composed of culturally alienated beings. In dealing with someone, you learn something not only about him, but also about others in his society. You learn something about population statistics" (Dasgupta 1988:64-5). In other words, by frequently transacting, the initial game of incomplete information can be transformed in a game of imperfect information: buyers do not know beforehand if the seller they are confronted with is honest or not, but the buyer's decision will then depend on his estimate p^e of the probability p that the seller is indeed honest. He will enter if

$$p^e(1-\delta) - (1-p^e)\delta > 0,$$

or
$$p^e > \delta \tag{3.1}$$

Further, p^e can be considered as a variable, to be adapted according to the experience buyers have had in previous rounds, such as modeled in the algorithm presented in figure 3.4.[10]. In terms of this figure, p^e can be seen as a buyer's estimate of the *proportion* of honest sellers in the society. On the basis of this model, we can predict that exchange will emerge *if and only if*

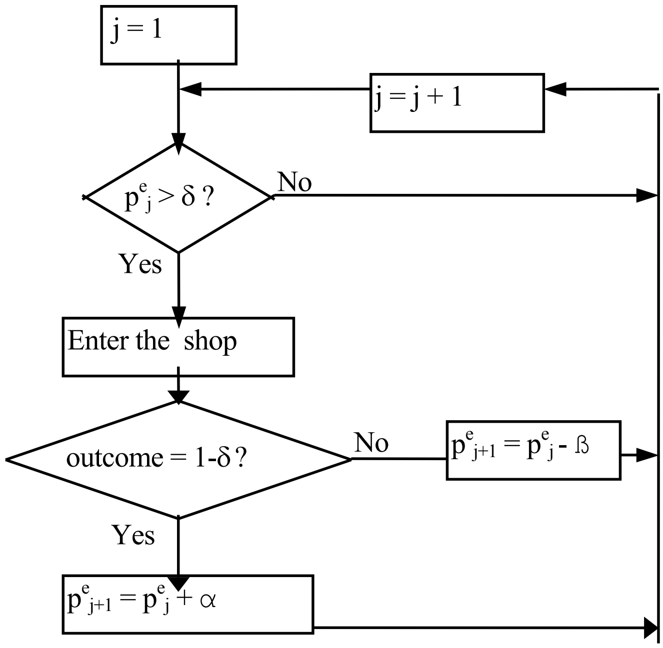
(1) the *effective* proportion of honest sellers p exceeds indeed δ,

(2) the initial *expected* p^e is (subjectively) set high enough to engage in a first transaction (thus initial $p^e>\delta$),

(3) the first transactions are predominantly with honest dealers, so that p^e is allowed to increase to the point where eventual "bad" encounters can be tolerated (thus where bad experiences will not cause p^e to decrease below δ), and

(4) α will be high enough, and ß low enough, so that the probability that, in some round, $p^e < \delta$, will be fairly low.

[10]For an overview of similar approaches, see Platteau (1994:757-764)



Legend:
j = number of round
α = factor by which p_j is augmented after positive experience
ß = factor by which p_j is decreased after negative experience

-- **Figure 3.4.** An algorithm to experience trust --

Hence, if there is a positive probability that the free market can emerge spontaneously, it will do so only within some socially determined bounds, which can be specified as follows:

First, players generalize knowledge about previous encounters with specific sellers to a given population, the boundaries of which are socially predetermined. The redefinition of a society's boundaries will, of course, have an immediate impact on the opportunities to establish a more or less complex division of labour. According to Platteau, Europe's history is unique in this sense, because it was the first continent in which a *generalized morality* emerged: "If limited-group morality is understood as morality restricted to concrete people with whom one has close identification while generalized morality is morals applicable to abstract people (to whom one is not necessarily tied through personal, family, or ethnic links), there is good sense in arguing that the western world has a somewhat unique history rooted in a culture of individualism pervaded by norms of generalized morality" (Platteau 1994:770).

Second, it is not sufficient that *some* people are applying moral norms to abstract people: a significant *majority* must do so[11]. At a minimum, the objective proportion p of honest sellers in the population must exceed δ, but, given condition (3), the chances of spontaneous market emergence will be significant only if p-δ is significant.

A *Third* social condition for the emergence of a free market is that buyers must have a (subjective) estimate of p exceeding δ to begin with, if not, they will not engage in any transaction at all (condition 2). In other words, it is not sufficient that there *is* a significant majority of players who act honestly in general, everyone must at least initially *believe* there is.

Fourth and finally, players are not easily discouraged by bad experiences and positively encouraged by good experiences (condition 4). In other words, buyers' estimates should be *biased* towards confirmation of their initial beliefs, and against disconfirming evidence.

Note that, while "trust" (expected p) is supposed to be highly dynamic within this setup, the objective proportion of honest sellers doesn't change during the (iterated) game, and neither do a society's boundaries or the players' initial beliefs; they have been fixed ex ante, if not, the buyer's attempt to estimate objective p based on former experiences is a naive excercise. Players are still maximizing actors, though they do so within some cognitive or normative bounds which are imposed upon them by a society's *cultural endowment*[12]. According to Platteau, cultural transmission of what he calls *moral norms*[13] takes place during early childhood, during the process of primary socialisation[14] since primary socialisation "takes place under circumstances that are highly charged emotionally" (Berger and Luckmann 1967:131). Moreover, as moral norms are subject to erosion, they need to be reinforced, "especially so if the cost of honesty is high or increasing" (Platteau 1994:771-773). Examples of norm-reinforcers range from teachers, the Church, the law and role models, to people one can communicate with.

Two remarks are due here.

---

[11]It can be added here that even a very low (subjective) probability p that there are SOME "honest" players in the total population is sufficient to significantly broaden the force of egoistic micro-motives; indeed, e.g. Kreps (1990:536-43) has shown convincingly that this situation suffices to predict cooperative outcomes for games which are finite (but which are iterated at least some 22 rounds). As we argued above, however, iterated games cannot provide for a full answer to the problem at hand.

[12]North defines culture as "transmission from one generation to the next, via teaching and imitation, of knowledge, values and other factors that influence behavior" (1990:37).

[13]Platteau defines them as follows: "An individual actually conforms [to the rule] because of a personal attitude about the act itself, that is, when conformity becomes a motive of its own because it is *intrinsically* rewarding or because deviation is *intrinsically* costly... As a result, moral norms are followed even when violation would be undetected, and therefore unsanctioned, because the moral act -which appears to be in conflict with the immediate or direct interests of the actor himself -is valued for its own sake" (Platteau 1994:766). Note the similarity with Elster's concept of *social norms* (1989).

[14]"Primary socialization is the first socialization an individual undergoes in childhood, through which he becomes a member of society" (Berger and Luckmann 1967:130)

First, by putting such emphasis on the role of primary socialization in the reproduction of a society's cultural endowment, Platteau implicitly underscores the necessity of the **non-egoistic micro-motives** on which a family is constructed; without an emotional tie between parents (or more generally: primary socializers) and children there would indeed be no place for socialization of moral norms.

This makes his account somewhat problematic: how could these ties ever develop in a truly *anonymous* society ? Robert Putnam goes even further; he asserts that the norm of what he calls "generalized reciprocity"[15] is "likely to be associated with dense networks of social exchange" (1993:172). In other words, it is not sufficient that individuals are brought up within a "thick trust" network, they must also be surrounded by dense networks as adults[16]. This is why Putnam uses the prevalence of all sorts of associations (from soccer clubs to trade-unions) as an indicator of "civicness". The non-economist Putnam has of course no problem with thick trust networks. Economist Platteau, however, recognizes the tension between such networks and a developing society, which is essentially anonymous. In the end, Platteau's analysis amounts to a pessimistic forecast: with increasing division of labour, those non-economic institutions that reproduce (the family) and reactivate (church, state, ...) moral norms will be gradually undermined: "capitalist development being based upon a high degree of division of labour and upon continuous processes of competitive innovation that call for ever-increasing consumption, it tends to cause a gradual erosion of the moral norms on which the effective functioning of the market rests" (Platteau 1994:794) [17].

We think, however, that the assumption that developing societies grow ever more anonymous is too strong to be realistic. As Platteau mentions elsewhere in the same text, organized hierarchies (firms), personalized employer-employee relationships, imperfect competitive markets and strategic interaction are *also* essential elements of a complex economy. Hence, in more complex societies, there is considerable space for "personalized relationships", even if *some* essential transactions are realized between anonymous partners.

The same argument can be made, besides, against North's analysis: While both North and Platteau see the necessity to broaden economic analysis beyond mere rational choice, they don't question the axiom that economic development cannot but be sustained by almost perfect –and hence anonymous- markets in at least some essential goods. Adhering to this axiom, North is able to downgrade the economic insights provided by the dependency school as a rationalization and a

[15]defined as the expectation "that a benefit granted now should be repaid in the future" (Putnam 1993:171)
[16]A similar (but still different) argument is made by Berger and Luckman (1967): human beings are always in need of "significant others", if not, they are not able to carve out a meaningful existence in this world.
[17]for a similar statement on the erosion of the family, see Frank (1988). According to Frank, however, the withering away of the family as a norm-producting entity can (and should) be counterbalanced by the introduction of some "moral courses" in the educational system. Frank's belief in the impact of education is also testified by his research on the differences of belief in others' trust between students of economics and students of psychology (Frank, Gilovitch and Regan 1993, 1996). The experimental results are controversial, however (Yezer, Goldfarb and Poppen 1996, Laband and Beil 1999).

reinforcement of the economic backwardness of Latin-America. This is not the place to dispute this particular dispute about dependency theory. Instead, we may point to a growing body of literature which disputes the aforementioned axiom itself (Simon 1992, Amsden 1989, Piore and Sabel 1984). Some point to the growing importance of *organizations* and trust networks, but, more interestingly, others point to the fact that *real* market transactions are not to be confused with *ideal* market transactions:

> "the multiple acts of buying and selling, characteristic of advanced market societies forge all sorts of social ties of trust, friendliness [and] sociability... Under perfect competition there is no room for bargaining, negotiation, remonstration or mutual adjustment and the various operators that contract together need not enter into recurrent or continuing relationships as a result of which they would get to know eachother well. Clearly this latter tie-forming effect of markets can be important only when there are substantial departures or "lapses" from the ideal competitive model. But the fact is that such lapses are exceedingly frequent and important" (Hirschman 1982:1473).

Hirschman's last sentence implies that it can still be (and we have no reasons to doubt) that economic development (defined, in this literature, as wealth) goes hand in hand with increasing allocation of resources by *real* markets. Hence, unless we suppose that such "lapses" from the ideal competitive model are rather dysfunctional perversions, we have to assume that the axiom is at least incomplete, and hence that the study of non-anonymous transactions (varying from clearly non-market interactions to *real-market* transactions) is at least as necessary to include in a theory of economic development.

A **second** problem with the approach is some theoretical inconsistency, which has been formulated as follows by Robert Frank: "Why don't [parents] just say '*Be a cheerful member of the group, deliver on your commitments where failure to do so can be observed, but otherwise avail yourself of every opportunity for gain'*?" (Frank 1988:48). In other words, it is not rational for parents to inculcate their children to *be* honest as long as there a trade-off between honesty and material benefits. One could of course always assume that norms are passed "unthinkingly" from one generation to the next, e.g. via simple imitation. A reduction of the socialization process to blind imitation effects is not very convincing, however. Moreover, what would explain the *emergence* of the norm of generalized morality in the first place, if, as is implied by the model above, this norm will only have some survival value after a critical mass (p) has adopted it ?

3.3.2. Signals and emerging honesty

Part of the argument of what follows is reflected in the story of the stupid ass:
"There was a very stupid donkey -so stupid that given the choice between a large carrot and a small one, it always chose the small carrot. This stupidity made the donkey very famous and people would come from far and near to try out the experiment on the donkey, which always chose the smaller carrot. One day another donkey asked the silly ass why he behaved so stupidly. 'Do you think', answered the stupid donkey, 'that so many people would come from everywhere and constantly offer me a choice of carrots if I did not always choose the smaller one ?' " (Sen 1983:13)

It is clear that the "very stupid donkey" was in fact rather brilliant, as his systematic refusal to behave as predicted given the incentive structure gave an incentive to those on whose behaviour he depended to choose *him* rather than another ass to feed carrots. While the average weight per carrot offered was smaller than in case he would have acted like his neighbour, the *amount* of carrots offered largely compensated for this disadvantage, resulting in a higher average weight of carrots consumed per time period. In fact, the "very stupid donkey" realized that his entitlement to food not only depended on his own capacity to distinguish between the bigger and the smaller carrot, but also on the behavior of the owners of carrots. By always choosing the smaller carrot, the brilliant donkey performed poorly in terms of the first allocation mechanism, but he also activated a second one.

To be sure, the second allocating device is not only activated by brilliant donkeys. An alternative would be a donkey conditioned to be polite (and instilled never to take the biggest piece), or a compulsive donkey, one that believes that even eating only one large carrot would eternally curse him. We can assume, however, that knowledge of the donkey's true motives won't change its success among carrot owners: compulsive, brilliant and polite donkeys are all interesting enough to pay a visit.

The story can be formalized and extended by translating it into the language of evolutionary game theory. Evolutionary models simulate how different types of a species evolve over time. Conventional models determine a type's success by its average payoff[18]. Eventually, successful types will raise a larger offspring than the unsuccessful ones, causing their weight in the total population to grow. In this way, the "inefficient" types are weeded out by natural selection, or an equilibrium is reached where the average payoff of each type is equalized.

The success of the translation of evolutionary models from biology to the social sciences depends, of course, on the quality of the fundamental assumptions we make about human beings. More

[18] For a more complex, and realistic view, see Bowles (1998:82-5), who suggest instruments to incorporate "non-random mating" and where the "success" of the type is not only determined by its payoffs but also by other types of trait-transmission .

specifically, any model should make explicit (1) the rationale for differentiating at least two "types" of human beings, (2) how these "types" can "breed" and/or "die" , and (3) whether and how these "types" can be differentiated by others -or *signalled* to others. The possibility to signal one's type is crucial since it allows to activate the second allocation mechanism.

In the case of Sen's donkey story, there are three possible answers to question (1): Different types of donkeys can be assumed to exist on the basis of (i) differences in intelligence (some donkeys are brilliant), (ii) differences in the boss's intelligence (a boss training his donkey to be polite has to give less food himself) or (iii) differences in psychological make-up (some donkeys are compulsive). The typical answer to question (2) is that more success leads to higher replication propensity, which makes sense, in our example, only as long as we assume that (a) carrots are the most important determinant of an ass's survival and (b) food is the most important determinant of his ability to replicate. The causal connection between average payoff and replication propensity can also be somewhat more complicated, though: It can e.g. be expected that, eventually, polite donkeys will grow in importance as other bosses notice the profitability of owning a polite donkey. Note, however, that we have to make explicit, in this case, how donkey owners can learn from eachother. Further, learning can also occur between the donkeys themselves; if, in Sen's story, the neighbouring donkey was curious to know the other ass's motives, we can expect that he was planning to use this knowledge for his own benefit[19]. Finally, concerning question (3), In Sen's story, *information* about the donkeys' type was freely available. The ass's behaviour was so surprising that the mere pleasure of being able to tell the story caused this spread in information. Note that the way in which the donkey was *known* was not a good description of his type: very stupid asses would probably not be able to see or acknowledge the difference between large and small carrots. Still, the important thing was that the signal worked. It can of course be assumed that, in the case of Sen's story, the spread in information would be a steeply rising function of the relative uniqueness of one type of donkey, and in this respect the story is clearly an inappropriate for our purposes. We will ignore this problem, and suppose indeed that the signal works.

The step from donkeys to second hand car sellers is sometimes very small. Going back to the extended game as presented in figure 3.3., we can see that (1) we specified, indeed, different 'types' of sellers, though we found the rationale behind it, as well as the implicit suppositions about (2) the reproduction of each type, unconvincing. Moreover, unless we specify a signaling device, there is no way in which buyers can make a distinction between the honest and the opportunist sellers, which will cause the honest types to die out, the buyers to stay out of second hand car shops and the market to cease to exist.

[19] In the latter case it might be suggested, however, that the real causal factor is *expected* rather than *average* payoff. More importantly, in case we assume the link between average payoff and replication propensity is mediated through a non-biological process, we should make additional assumptions about the *intensity* of transactions relative to the *speed* of learning. If the donkeys would immediately adjust to the carrot owners' reactions by changing type, the concept of "average payoff" would become a senseless measure of success. Conventional evolutionary models assume that the weight of different types can only change at the end of each generation, and that many interactions (in this case between donkeys and carrot owners) occur within one generation.

Assume now, for the sake of argument, that the sellers' type can be *signaled* to the buyers, provided the latter pay s (scrutiny costs). It is the possibility to signal one's "type" which facilitates the activation of another allocation mechanism. Note that "signals" are more encompassing than "reputation". Whereas the latter concept refers to the track record of previous behaviour, the former is more general, it comprises every clue buyers have at their disposition to decide about the sellers' type.

Suppose p, the proportion of "honest" sellers in the population of salesmen, approaches 1 in the initial stage of the evolutionary game. As a consequence, no potential buyer will need to scrutinize the seller's character: the probability of buying a lemon is too low to take costly precautions. Since no-one scrutinizes, however, opportunistic sellers will have opportunities to sell their lemons. Consequently, their demographic importance will grow, and p will decrease. If we start at the other end (p approaching 0), buyers will *always* check the seller's character before buying his car, at least in markets where the net expected result is still positive. This means there must be a point where

$$p(1-\delta-s) + (1-p)(-s) = 0$$

or
$$p(1-\delta) = s$$

or
$$p = s/(1-\delta) \qquad (3.2)$$

In other words, markets where scrutiny costs are positive don't necessarily exist: if, in the initial situation, $p<s/(1-\delta)$, the buyers' profits realized in transactions with honest dealers cannot compensate for the losses they incur each time they scrutinize a seller which happens to be dishonest. Once p rises above $s/(1-\delta)$, however, average profits become positive. The equilibrium p* will be reached at

$$p^*(1-\delta-s) + (1-p^*)(-s) = p^*(1-\delta) + (1-p^*)(-\delta)$$

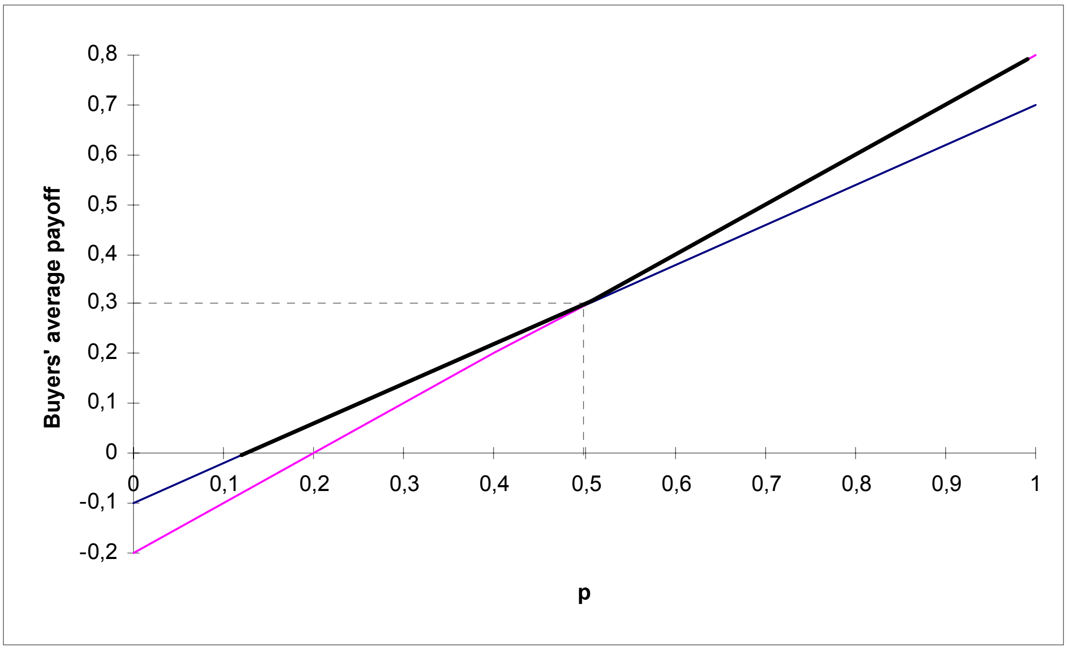
or
$$p^*d + s = \delta$$

or when
$$p^* = 1 - s/\delta\ . \qquad (3.3)$$

In the following figure, we present the evolution of buyers' average payoffs in function of the proportion of honest sellers in the population for s=.1 and δ=.2. Note that, with perfect information, all dishonest sellers would be driven out of the market. Once information gets costly, however, we'll need a minimal initial proportion of honest sellers ($p>s/(1-\delta)$), and there will always be a proportion (s/δ) of free-riders. The thick line in the figure presents the set of values of p for which a market exists. All elements in this set are unstable, with the exception of the equilibrium point p* with coordinates (.5,.3).

Figure 3.5.

Relationship between buyers' average payoffs and proportion of honest sellers



Buyers' average payoffs are calculated as p(1-δ-s)+p(-s) for p<(1-s/δ) and as p(1-δ)+(1-p)(-δ) for p>(1-s/δ). In the example, s=.1 and δ=.2

The costliness of scrutiny to verify a seller's type is crucial in this model. At relatively[20] low scrutiny cost, the payoffs to the different types of sellers are not only allocated in function of their own choice but also in function of the selection made by the buyers. But what is meant in reality by the metaphors of "signal", "scrutiny cost" and, especially, "type" ? An answer to these questions must necessarily make explicit how people believe, perceive and behave.

A first path worth reviewing has been defended by Robert H. Frank (1988). Frank argues that honesty is a **moral norm**. The mere thinking about transgressing the norm will produce not only emotions of guilt (which will cause the payoff structure to be changed), but also physical signs like eye movements, a ruddy complexion, posture, the rate of respiration, etc. Opportunists could try to mimic these outer signs, but their problem is that these signs are by-products, "at least partially insulated from direct control" (1987:594). Precisely because they are by-products, it is *by definition* impossible to "invest" in them (Elster 1983). As a result, in order to *appear* honest, it may be necessary to *be* (or at least to be able to say to oneself one *is*) honest. We can now understand why parents choose to educate their children to be honest (and why teachers and economy professors should do the same): someone having successfully internalized these teachings is observably different from others. "That he might receive a material payoff for such behavior is completely beyond his concern. And it is precisely because he has this attitude that he can be trusted in situations where his behavior cannot be monitored" (Frank 1988:69). In terms of Sen's donkey story, we would suppose that the donkey chose the small carrot simply because his boss has trained him to be polite.

[20] Relative to the loss if buying a lemon (-δ) and the profit if buying a good car (1-δ)

Our difficulty with Frank's account is, however, that his solution necessarily assumes that one's "type", i.e. the moral sentiments one has developed, is the product of cultural conditioning: During childhood, children are conditioned as their parents would wish them to be. Contrary to Platteau, norms are not anymore passed "unthinkingly" from one generation to the next, but, in any case, properly conditioned children become polite donkeys once they pass into adulthood, and children who lacked proper conditioning during socialization will face the dim prospect of dying as opportunists. The emphasis on cultural conditioning enables Frank to evade the problem that you cannot feel guilty merely because you *want* to feel guilty: Indeed, in Franks view, people cannot feel guilty at will, but he assumes that parents can make their children feel guilty at will. A similar position is taken by Akerlof (1983) and Becker (1996). It is also contradicted by empirical evidence (Elster 1996:763, Bowles 1998:82-3): Educating children is definitely different from conditioning donkeys.

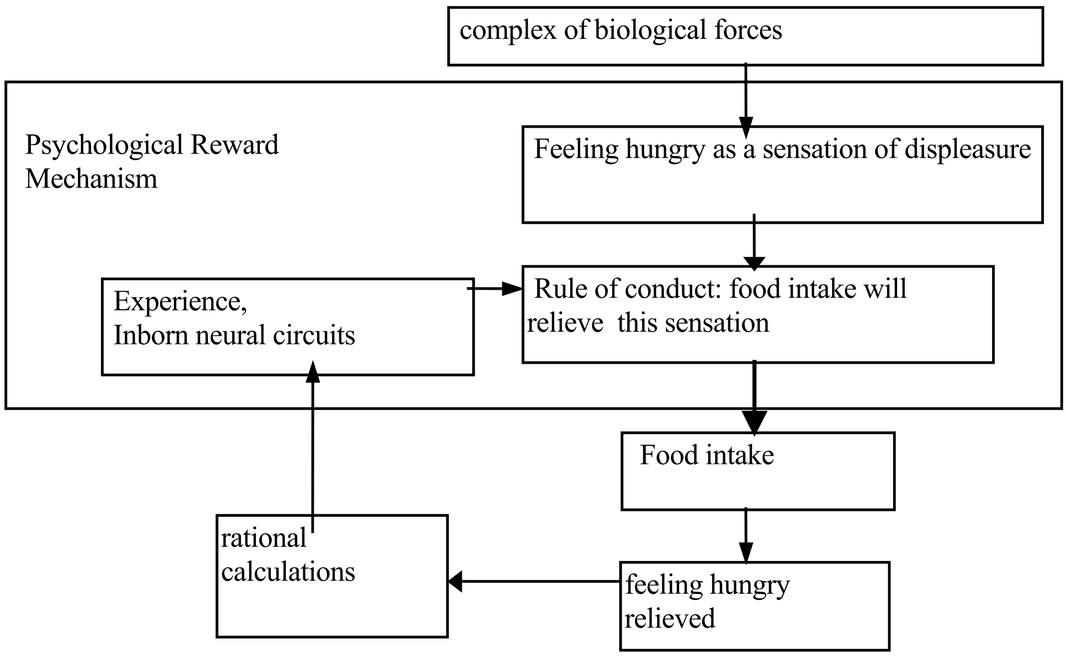
Besides verifying a seller's physical appearance supposedly revealing his emotional make-up, buyers can also verify his **reputation.** In the donkey-story, the "signal" triggering the higher-order allocation device was indeed the donkey's past behaviour. Reputation is also a rather problematic signal however: True, many buyers do, in reality, scrutinize a seller's reputation before they enter his shop, but this information can never be complete, nor representative. People must make a judgment on the basis of partial knowledge, knowledge derived from occasions where the seller's behaviour has been observed. But why would dishonest sellers sell lemons in these occasions, if they knew that it would destroy their reputation ? In other words, wouldn't a track record of past dealings be a very bad signal precisely because the seller can anticipate the buyers' action and manipulate it ? Honesty, says Frank, is not to be confused with intelligence or prudence. Prudent and intelligent sellers will only cheat where they are sure no-one will detect them. So, when buyers are checking a seller's reputation, they are only verifying his prudence, intelligence, rationality, not his honesty, and, as a result, the information they have about a seller's reputation must be a bad predictor of his behaviour in circumstances where his behaviour is not observable by other potential customers (Frank 1988:72-75). Are buyers who are checking a seller's reputation irrational, then ?

3.4. nonwealth-maximizing motives.

According to Frank, they are not. His answer to this question, and to the related question of signals capable to differentiate the "honest" and the "dishonest" types, is grounded in a more complex "picoeconomic" theory of how human beings function and reason. A presentation of this theory and some of its variants will allow us to discuss various alternative solutions to the problem of exchange in the anonymous market. In the first sections (3.4.1. and 3.4.2), we will present the essence of the theory, by making use of Frank's (1988) writings but mainly of the contributions of George Ainslie (1992). They can be skipped by readers acquainted with it. Subsequent sections discuss its relevance for our purposes.

3.4.1. Temporary-preference theory

We have tried to present Frank's model of man (and woman) in the figure below, drawing on his account of why we eat. The essence of the model is that human behaviour is *directly* guided by a complex psychological reward mechanism: we eat because we want to relieve the sensation of feeling hungry. Experience, and perhaps inborn neural circuits, tell us that food intake will relieve this sensation.



-- **Figure 3.6.** Why we eat --

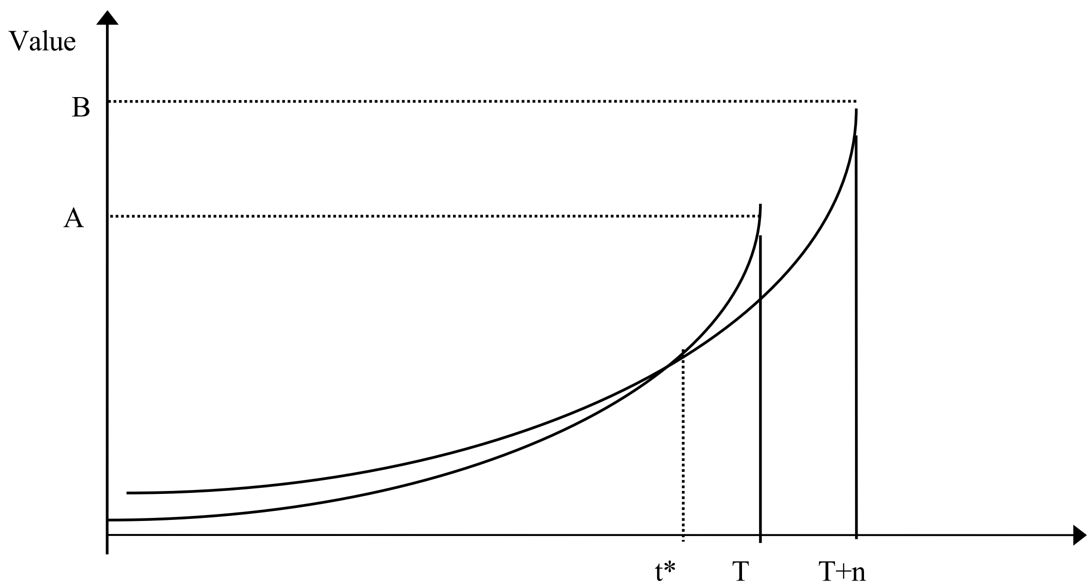
Rational calculations (e.g. theories about caloric intake, experience with food prices, trade-offs between food consumption now and other types of consumption later etc.) play only an indirect role in this process: the direct cause of food intake is a pico-economic motive to eliminate a sensation of displeasure.

Food intake can also be related to other sensations, however. People may have an impulse to eat simply because food is available[21], an impulse that may be have emerged during evolution. Conversely, people may *abstain* from eating because they would consider it impolite or inappropriate. Abstaining from eating is, in this case, *directly* caused by some negative emotions (e.g. feelings of guilt and shame), and *in*directly by previous experiences, other types of learning processes and/or rational calculations. As a result, some internal conflict may arise between several contradictory impulses, the "strongest" motive being able to direct behaviour.

So far nothing new: pico-economic motives could be compared to "tastes", and, provided preferences can be ranked in a consistent way, the usual micro-economic apparatus can be applied to predict people's behaviour. However, picoeconomic preferences can*not* be ranked in a consistent way. More precisely, picoeconomic motives differ from "ordinary" preferences *inter alia* because they are discounted in an hyperbolic rather than in an exponential way (Ainslie 1992). More specifically, we will adopt Ainslie's proposition that the value of a good,

$$V = A/(1+(T-t)) \tag{3.4}$$

with
A = the psychological reward it brings at the time T it will be consumed, and
T-t (delay) = the time between the moment of valuation t and the moment of consumption T.



-- **Figure 3.7.** Temporary preference--

Before proceeding, it must be stressed that this alternative formula to discount delayed rewards is consistent with (1) the results of numerous experiments with humans as well as animals (Ainslie

[21] Even the totally satisfied character in Monty Python's *the meaning of life* did not refuse the tiny chocolate that eventually caused him to explode.

1992:58-60), and with (2) theoretical developments of the so-called “matching law” in psychology (Ainslie 1992:63-76, Rachlin 1995).

The crucial consequence of hyperbolic discount curves is that the same person will choose differently between two alternatives at different moments t. Take two rewards, A to be consumed at time T, and B, significantly greater, to be consumed at time T+n. In the figure, it appears that, while individuals will prefer good B over good A at almost any time, they will prefer A over B during the time-segment $]t^*,T]$, and hence, if they must decide between A or B when $t \in\ ]t^*,T]$, they will opt for A, even if they would have opted for B at all other moments $t \notin\ ]t^*,T]$.

3.4.2. Impulses and personal rules

The most well-known example of a temporary urge that disables us to obtain a far better, but more delayed outcome is probably given by the story of Ulysses and the sirens; the sirens symbolise an irresistible temptation that is even able to make the most experienced sailor blind to the risks he takes when trying to come nearby, so that “he has no prospect of coming home and delighting his wife and little children as they stand about him in greeting” (Homer, cited in Schelling 1980:532).

The same story also suggests a first solution to the problem: Ulysses deliberately chose to be bound to the mast of his ship, the mast symbolising, as it were, third party enforcement. But, in this case, third party enforcement is of a very radical kind: as Ulysses knew beforehand that he would give in to the temptation, he chose to bind his hands even before he had committed the error.

A second solution to the temporary preference problem is suggested by Ulysses’ companions, who put “sweet wax of honey” in their ears: they were now in the position to sail safely past the rocks, unable to hear neither the Sirens’ singing nor Ulysses’ lamentations. At the infra-personal level, this solution has been described by Freud as *repression*; it could also be described as “distorting information about environmental opportunities for impulses” (Ainslie 1992:134). Such distortion may be realized at a pre-conscious level.

Two other solutions require a bit more attention. To begin with, human beings might decide to follow a *personal rule* in a particular situation, i.e. they might decide to act systematically in a particular choice contect, irrespective of the immediate rewards they will obtain. The *need* to follow a rule is clear once one has the intuitive understanding of what an impulse is all about: After having chosen the “wrong” alternative A an individual not only knows he has made the wrong choice, he also knows he will make the “wrong” choice again in a similar situation, unless he decides to intervene in the decision mechanism.

Consider the problem of overeating: The eye is larger than the belly. As the capacity of my stomach bears no proportion to the immensity of my desires, I am tempted to eat as much as I can (reaping reward A), even if I will subsequently suffer from an aching stomach (eating moderately would bring me a larger, but delayed reward B, reflecting absence of an aching stomach). Moreover, after having experienced this, I will be able to predict the same pattern of conflicting rewards in similar circumstances in the future, *unless* I follow a personal rule. Some problems remain, however:

Why would I follow the rule at the moment of choice ? Indeed, I can clearly feel the need to follow a rule, but precisely at the moment of choice (the interval]t*,T]) following a rule clearly contradicts my short-run interests. However, following a rule implies that the payoff-structure I face will also be altered: I should add the (actualized) rewards to the current rewards from either systematically over-eating or systematically eating moderately. Suppose that eating more than necessary gives me a reward of 1 at time T=10, but a feeling of displeasure of -3.5 at time T+5 which I could evade by eating moderately. At the moment I have to make the choice (T=10), it is easy to see that I will prefer to eat more than necessary, if I *only* consider the actual choice situation: Applying the discounting formula presented above (3.4.), we get

$$1/(1+0) > 3.5/(1+5), \text{ or}$$
$$1 > .58$$

If, however, I see my choice at T=10 as one example of similar choices I will have to make at T+10, T+20,...,T+70, I will have to compare the sum of actualized payoffs from systematically eating moderately in those situations against the actualized payoffs from eating systematically more than necessary. The payoffs are likewise altered, the more I add future situations to the list of "similar" situations:

Table 3.1.

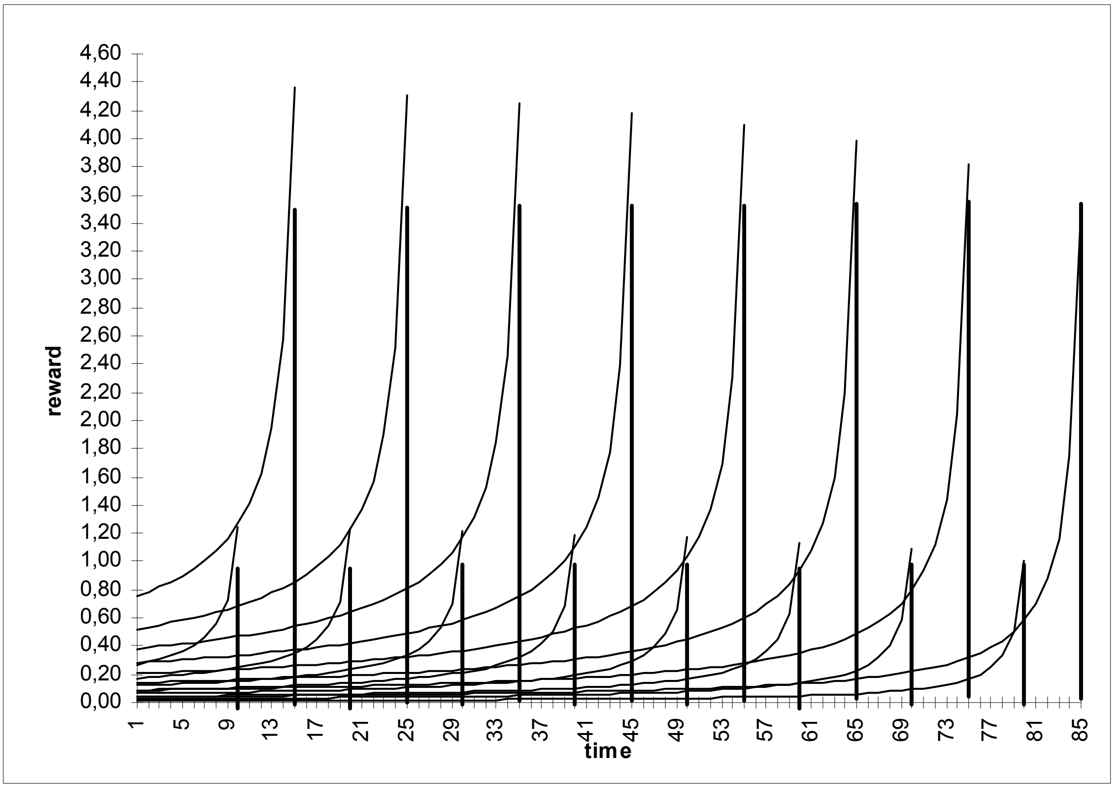
Aggregate payoffs at T=10 when discounting similar rewards, consumable at later times T+(n-1)*m, 2≤n≤8

N	1	2	3	4	5	6	7	8
Overeating	1	1,09	1,14	1,17	1,20	1,21	1,23	1,25
Eating moderately	,58	,80	0,94	1,03	1,11	1,17	1,23	1,27

Source: own calculations, with A (overeating) =1, B (eating moderately)=3.5 and m=10

Figure 3.8.

Comparing families of similar choices



Source: own calculations, same assumptions as those of table 3.1.

Eventually, the balance shifts in favour of eating moderately. The graphical result can be observed in the figure 3.8.: when I can expect that I will follow the personal rule for at least 8 consecutive periods, the discount curve of discounted rewards will eventually exceed the payoffs from systematic over-eating at all periods, or the time interval]t*,T] will have shrinked to nothing.

Why will my future selves follow the rule ? Indeed, the strategy to follow the rule all the time is clearly Pareto-optimal whenever there is a perspective that at least 7 consecutive future selves will follow it as well, but this is hardly a Nash-equilibrium: at each point in time, I can reach a better short-run outcome by defecting (and overeating). The game I play with my future selves is an ordinary one-shot n-person prisoner's dilemma: each of us have to abide by the rule, if not we will not realize the collective optimum. How can I trust my future selves to co-operate, then ? The answer may be hidden in the fundamental difference between an interpersonal and an infra-personal n-person game: In the infra-personal case my future selves have more information: they will know what I have presently chosen. In other words, the game I play with my future selves is both an n-person prisoners' dilemma and an n-times iterated supergame. This opens at least two different perspectives to a solution.

The intuition of the first perspective was noted by Schelling (1996 [1980]:553-4), and Kreps' already mentioned contribution may serve to give it a more formal reading: when $t<t^*$ (when my thoughts are still not heated by the temptation) I may fix the following strategy and counter-strategy: whenever I refrain from overeating at T,T+m, T+2m,... I can reward myself at T+2, T+2+m, T+2+2m,... (e.g. smoke a cigarette after a moderate meal). Conversely, when I cannot refrain from overeating, I will inflict a punishment (e.g. forbid myself to smoke a cigarette after a copious meal). The folk-theorem suggests that such ex-ante fixation of strategies and counter-strategies may well lead to the co-operative outcome. But how can I be sure that my future self will punish me ? Why would the threat be credible ? This was a difficulty Schelling worried about, but if we accept temporary preference theory, worries are not needed: At T+2, T+2+m,... I will indeed regret what I've done or be relieved if I really refrained at T,T+m,..., and hence my self will not feel any difficulty either to punish or reward myself at those moments. I think the empirical relevance of this strategy is testified by the fact that many of the people who quit smoking increase in weight: they reward themselves with sweets (or beer ?) each time they feel themselves attracted to a cigarette.

However, relevant as this perspective may be, the empirical fact that many non-smokers with a taste for cigarettes do not need sweets suggests that another perspective may prove far more relevant: The fact that I don't overeat at T will make me feel confident that my future selves will co-operate as well at T+m, T+2m,... : Assuming the other actors are like me, I may regard my own behaviour as an indicator of the other actors' behaviour, and *therefore* co-operate myself. If, in the *inter*personal case, such a reasoning amounts to magical thinking (indeed, I am in fact manipulating the indicator, not the others' behaviour itself) the assumption that the others are indeed like me makes quite a lot of sense in the *infra*-personal case, my future selves and I *are* after all one and the same *person* -at least this is what I would like to think. In other words, when I see my current choice as an indicator of my future selves' choices, I bet on the fact that I am, indeed, a personality, endowed with a free will, a person being in control, not someone being "triggered" by the most immediate impulse or changing his mind from one period to the next. My current co-operation is a *precedent*, a signal that reveals my *self-image*[22]. Suppose I decide to co-operate (follow the rule and eat moderately). My future selves are then not only confronted with the choice to abstain or not, but also with the option whether or not to subscribe to my previous behaviour, whether or not to honour the image I have of myself.

According to Ainslie, a feeling of *guilt* arises when it becomes clear that I have not been able to follow the rule I set to myself. Guilt makes me wonder what kind of person I am, and even if I still *am* a person, i.e. someone who can think "about oneself as having an enduring self, over and above habits and inclinations" (Elster 1989:202). Within temporary-preference theory, a feeling of guilt can hence not be formalised as a reward we can incorporate in a utility function: it

[22] See also Sen's definition of "identity" : "how the person sees himself or herself" (Sen 1985:348). See also Knight who, in the preface to the 1933-edition of *Risk, uncertainty and profit*, states that "no distinctively human interest is at bottom economic. They all centre, as a matter of social-psychological fact, in a wish to be a person of some kind in a society of some kind" (1946:xxxiv).

reflects, as it were, the unresolved tension *after* we have made a choice. It is *intrinsically* unaccountable.

Will I always follow the rule ? Supposing I am involved in this game *as a person*, the answer to this question must be: I will not co-operate if I can convince myself that my self-image will not change with what I am currently choosing. When I have good reasons to suppose that what I'm currently choosing will not be interpreted by my future selves as a precedent, I may as well give in to the temporary impulse. These "good reasons" are called *loopholes*[23] by Ainslie. If T=0 is Christmas evening, some persons could consider this feature so exceptional that they can deviate from the rule they set for themselves at T=0 without rendering it implausible to deviate at times T+m (when it is not Christmas evening). Finding a loophole results in the sum of the actual reward to overeating (1) and the future discounted rewards to abstaining (1,27 -0,58 ≈ 0,69). This is obviously more rewarding than following the rule *also* at Christmas evening (1,69 > 1,27).

Will my future selves always follow the rule ? The least interesting option will be to eat moderately today but to overeat tomorrow (0,58 + 0,25 ≈ 0,83): this is the option chosen by the persons who create a false hope today that their current choices will determine their future choices. My current choice is only a precedent for future choices, my future selves retain the freedom to subscribe or to invalidate the image I currently have of myself.

The reasoning makes clear that the degree to which personal rules will be followed is to a large extent determined by the way in which individuals have "mapped" reality.

First, a personal rule can only be constituted when an individual perceives different types of situations as similar compared to others, and, beyond a certain point, this cannot be but somewhat arbitrary: drinking during the coffee break could or could not be considered as part of the game I play against my identity as a moderate eater. To give a less trivial example, someone living in Congo might see the analogy between hunting and diamond digging (De Boeck 1998), while to us, foreigners, these activities seem to be of an entirely different kind.

Second, rules must provide for clear criteria (Ainslie adopts Schelling's concept of "bright lines") to distinguish the "good" from the "bad" decision. The rule not to eat "more than necessary" has some obvious arbitrariness, except perhaps for experts in caloric requirements. The rule e.g. to eat only one time a day seems much more appealing for non-experts, the rule to eat nothing at all would be the most perfect one -if feasible. Similarly, ex-smokers do better adopting the rule "not to smoke at all" than "to smoke only when their host offers them a cigarette". The last rule might serve his long-term interests better, but it is also more ambiguous: you can e.g. pay more visits to cigarette-smoking hosts, or try to seduce the host to offer you a cigarette.

---

[23] Or, what Elster calls *adjunct norms*: "whenever there is a norm, there are often a set of adjunct norms defining legitimate exceptions" (Elster 1989:100-1).

Third, loopholes must be clearly distinguishable as well. Christmas evening might be defined as a loophole, but there are other exceptional days as well, and the criterion for defining a day as an exception may be very variable. In situations of ambiguity, it can be predicted that the present self will interpret reality in the way that suits him or her best.

The interest of the above theory of norm formation is, I think, that it argues for a different link between behaviour and the constraints a person faces, a link which identifies, moreover, some ways to induce behavioral change. Actual behaviour in a particular situation is not so much determined by the type of constraints actually faced, but rather by one's previous experiences and by the constraints faced at previous moments. Moreover, if choices depend on one's perception, one may alter choice by altering perception, by trying to convince an actor that his "mapping" is somewhat erroneous, that a particular situation might be dissimilar from others he had mapped as similar, or that the constraints are not (anymore) as he thinks they are.

3.4.3. Is honesty the best policy ?

These insights can now be used to interpret the problem of the trust game. To begin with, the problem that a>b (figure 3.1) does not have to mean that sellers will always cheat, even if they have a clear incentive to do so at the moment of choice. Sellers will cheat if (1) they are not able to control themselves, (2) if the rule they set for themselves allows them to cheat or (3) if they have found a loophole, a good reason not to follow the rule forbidding them to cheat in this particular case.

Case (1) is the case of the systematic opportunist. I think it is also a case of minor importance - though perhaps not for psychiatrists. Normal people do have at least the *ability* to constitute personal rules and to live up to them. Ainslie attributes the inability to constitute rules that free the actor from immediate impulses to a childhood past of either too much or too few discipline (1992:213-6). Too *much* discipline eliminates the need to constitute rules: if behaviour is totally regulated from outside, one doesn't have to make the rules oneself[24]. Too much discipline breeds psychopaths, and most people aren't. As the making of personal rules is advantageous in itself, there is no further need to explain why children do this. It can also be supposed that the ability to make rules will be strengthened by experiencing the advantages of such a strategy. Parents may be thought to influence this process in two ways. First, they may use their "power to affect, through reward and punishment, which of his behaviours will be especially costly" (1992:215). Their most important role, however is to provide for a model, "showing him both effective rules and skilful rule-making" (1992:215). By identifying itself with its parents, the child will be able to use their behaviour as sources of criteria for personal rules. People deprived too early from parental assistance in rule-making will develop compulsions, they will rigidly adhere to a rule, not knowing how to adapt it to changing needs (1992:212).

[24] For an experiment that nicely illustrates this statement, see Akerlof and Dickens (1982:318)

Case (3) is of minor importance, I think, for the problem at hand: except if we have to deal with fly-by-night operators, no seller will classify his transactions as exceptional happenings. To be sure, this argument is not very convincing if applied to the selling of second-hand cars: many sellers *are* fly-by-night operators in this market. At the other hand, not only sellers, but also buyers are supposed to be human beings, and buyers, especially final consumers, can sometimes better be seen as fly-by-night operators. Ainslie explains the "selling trick" to present transactions as "exceptional offers" by his model (Ainslie 1992:240). If a sale is presented as exceptional, or if it is presented in a way that makes comparison with competing offers difficult, the buyer will not apply the "usual" rule he had developed himself (this rule might be e.g. to scrutinise the seller's reputation) and take a myopic decision.

Case (2) is more interesting. Suppose the rule one follows is "honesty is the best policy". This supposition is rather problematic in our case: the whole point of a fully *anonymous* market is that *there are no (significant) delayed and bigger material rewards by abstaining from selling lemons*, so how could this rule have emerged in the way Ainslie described it ? To be sure, being always honest is clearly not a rule that applies only to selling behaviour: Being honest may make one a good spouse, a preferred member of the club of card players, etc., and hence it may *on average* be a good rule. Moreover, if buyers are able scrutinise a seller's reputation by interviewing the seller's wife or by visiting his club, the rule may be very appropriate. Still, selling cars seems an activity sufficiently different from playing cards, so that it can be supposed that people will apply different rules in these different spheres of activity. It can even be argued that different rules will be sustaining eachother: The rule to make good profits (by selling lemons) is clearly complementary to the rule to be a good spouse.

However, the rule to be honest will become more interesting when we introduce a slight departure from the purely anonymous market (and towards a more realistic market): suppose that there are varying chances that actual giving-in to immediate impulses may be signalled to future customers. Faced with this environment, the most interesting rule the seller may adopt is: *deliver on your commitments where failure to do so can be observed, but otherwise avail yourself of every opportunity for gain*. Such a rule would be followed by what Frank, paying tribute to Adam Smith, called a *prudent* seller. However, the problem with it is that it is dependent on a very ambiguous variable: "where failure can be observed" is indeed a continuous and noisy variable, not a clear bright line, and hence the merely prudent seller risks to sell a lemon in situations where he will in fact be observed. Just as an ex-smoker should distrust flexible rules, sellers operating in quasi-anonymous markets will only survive if they follow the rule *never to sell lemons.*

Sellers adhering to such an absolute rule will indeed be the favourite targets of scrutinising buyers. However, the paradox of this solution is that the belief that one's will is weak causes the will to be strong and to resign from selling lemons. Eventually, the seller may come to realise that this belief is in fact not consistent with experienced reality: not giving in to the temptation may indeed be considered evidence of self-control. Once having realised this inconsistency[25], nothing will deter the seller from relaxing the rule[26] -as a result of which he would lose his good reputation. In other words, this kind of solution would not be very robust. Most sellers are not like ex-smokers, they seem to be able to forego temptations without having to adopt a compulsive attitude towards them.

To be sure, mere *prudence* seems to be all that is needed whenever everyone's behaviour can be perfectly monitored (in the evolutionary game model, s=0); in that case, there *are* indeed bigger and delayed material rewards to be gained by abstaining from selling a lemon.

Robert Frank solves the problem of the merely *prudent* seller by re-introducing the role of moral sentiments:

"If the psychological reward mechanism is constrained to emphasize rewards in the present moment, the simplest counter to a specious reward from cheating is to have a current feeling that tugs in precisely the opposite direction. Guilt is just such a feeling. *And because it too coincides with the moment of choice, the matching law does not discount it relative to the competing material reward. If it is felt strongly enough, it can negate the spurious attraction of the imminent material reward... Thus, when we see that a person has never been caught cheating, we have reason to believe that his behaviour is motivated, at least in part, by nonmaterial rewards*" (Frank 1988:82-4, italics as in original).

Sellers endowed with a conscience will both (1) express physical signs if faced with the possibility of cheating and (2) be enabled to overcome the problem of temporary preferences. Both characteristics will enable and stimulate buyers to select *them* and to discriminate against cheaters and merely prudent sellers. Their material success will eventually convince the "honest" sellers to raise honest children too.

Note, however, that the way in which Frank uses "guilt" is different from the use of it by Ainslie, according to whom "guilt primarily reflects a decline in the expectation of self-control, rather than being a force that strengthens self-control" (Ainslie 1992:194). Note also that while Frank uses reputation as an equivalent to physical signals of one's "type", his account remains unconvincing to the extent that the "type" itself is determined ex-ante, viz. through a process of "cultural conditioning" (1988:92-3).

At the other hand, I concur with Frank's argument that moral sentiments, i.e. the qualification of selling a lemon as "opportunism" and selling a good car as "honesty", may overcome the temporary-preference problem, though I'd like to add that the problem with prudence is *not* that it is not appropriate, but only that it is not *sufficient* to explain honest behaviour. To make this argument, a short visit to the original Adam Smith seems necessary.

---

[25] "It is part of the notion of a belief that to believe something implies the belief that one holds the belief for a reason, and not merely because of the utility of holding that belief" (Elster 1983:51).

[26] Until I read Ainslie I've been unable to quit smoking precisely because I had no difficulty to stop smoking for some time.

3.4.4. respect and dignity.

Essentially, human beings as Adam Smith describes them in his Theory of Moral Sentiments are primarily *social subjects*. The initial state of the world as he knows it is vastly different from an Hobbesian State of nature, where men are assumed to have "sprung out of the earth, and suddenly, like mushrooms, come to full maturity, without of any kind of engagement with eachother" (Hobbes 1966 [1651]:109). The people Smith had in mind were very different from mushrooms, as is testified by his description of "that great purpose of human life which we call bettering our condition":

"what are the advantages which we propose by that great purpose of human life which we call bettering our condition ? To be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation, are all the advantages we can propose to derive from it. It is the vanity, not the ease, or the pleasure, which interests us. But vanity is always founded upon the belief of our being the object of attention and approbation. The rich man glories in his riches, because he feels that they naturally draw upon him the attention of the world, and that mankind are disposed to go along with him in all those agreeable emotions with which the advantages of his situation so readily inspire him. The poor man, on the contrary, is ashamed of his poverty. He feels that it either places him out of the sight of mankind, or, that if they take any notice of him, they have, however, scarce any fellow-feeling with the misery and distress which he suffers. He is mortified upon both accounts; for though to be overlooked, and to be disapproved of, are things entirely different, yet as obscurity covers us from the daylight of honour and approbation, to feel that we are taken no notice of, necessarily dampens the most agreeable hope, and disappoints the most ardent desire, of human nature" (TMS:50-1).

Contemporary economists could read this statement as a somewhat complex version of the assumption that people are rational egoists, since even if "bettering our condition" breaths social considerations rather than intrinsic utility, the assumption that people seek material enrichment is not disqualified by this nuance. However, some elements suggest a more complicated reading.

First, in the *Wealth of Nations*, "bettering our condition" is contrasted with "the principle which prompts to expence" in the following terms:

"The principle which prompts to expence is the passion for present enjoyment; which, though sometimes violent and very difficult to be restrained, is in general only momentary and occasional. But the principle which prompts to save, is the desire of bettering our condition, a desire which, though generally calm and dispassionate, comes with us from the womb, and never leaves us until we go into the grave" (WN 341)[27] .

In other words, actors seem to be far from cool utility calculators sometimes, as their calculations seem to be distorted "from within", by "violent" but "momentary" passions. The framework of temporary-preference theory seems perfectly able to account for what Smith had in mind.

[27] Though not referring to Smith, Ainslie invokes a very similar argument to explain why people save, i.e. why they seem to discount the future much less than originally assumed by temporary preference theory (Ainslie 1992:228-41). Ainslie regards savings rather as rule-confirming exceptions, however; many other spending patterns can only be explained by assuming a much higher discount rate.

Elaborating on this, elsewhere in the *Wealth of Nations,* Smith argues that "some particular men may sometimes increase their expence very considerably though their revenue does not increase at all", but that "though the principles of common *prudence* do not always govern the conduct of every individual, they always influence that of the majority of every class or order" (Smith, WN:295). In other words, *prudence* seems to describe the behaviour of *most*. Smith defines it as follows:

> "The qualities most useful to ourselves are, first of all, **superior reason and understanding**, by which we are capable of discerning the remote consequences of all our actions, and of foreseeing the advantage or detriment which is likely to result from them: and secondly, **self-command**, by which we are enabled to abstain from present pleasure or to endure present pain, in order to obtain a greater pleasure or to avoid a greater pain in some future time. In the union of those two qualities consists the virtue of prudence, of all the virtues that which is most useful to the individual" (TMS:189, our emphasis).

Analogous to Ainslie, Smith sees no need for "cultural conditioning" in order to overcome the impulse-control problem; the mere capability to discern the remote consequences of all our actions is already sufficient to signal the *necessity* of not responding to the immediate impulses. Further, self-command seems to be Smith's analogue for Ainslie's concept of free will. Again, a concept like "self-command", the ability to abstain from present pleasure in order to obtain a greater pleasure in the future, is not needed within a conventional rational choice framework, where temporary preference reversals are per definition outlawed.

Second, according to Smith, the "generally calm and dispassionate" desire to better our condition is essentially *merely instrumental* in triggering the others' approval and attention. Though Smith uses the word "vanity" in the above-mentioned citation, he does not imply by this that by enrichment one seeks to distinguish oneself from one's neighbour or to be envied by him; quite to the contrary, the rich man seeks to be "attended", "approved", he wishes that "mankind is disposed to go along with him". Enrichment is supposed to trigger a feeling of **dignity**, caused by the others' expressions of **respect,** rather than a feeling of distinction caused by the others' expression of envy.

In sharp contrast, the poor man finds himself in "obscurity", covered from "the daylight of honour and approbation, which disappoints the most ardent desire of human nature". Elsewhere, Smith defines shame as "that consciousness that his misery is felt by himself only, which is of all sentiments the most unsupportable. There is no sympathy in the other" (TMS:61).

Of course, personal enrichment is not always considered sympathetic. In fact, the trust game precisely exemplifies that the seller's enrichment will most probably trigger the buyer's indignation and contempt. At the other hand, selling a lemon seems to be consistent not only with "bettering one's condition", but probably also with receiving sympathy, complacency and approbation from those to whom the profits will be distributed. How can the seller decide, then, (and how can *we* predict) whose approbation he should give priority to, and whose contempt he should ignore ?

Smith solves the question by submitting that we desire not only to be praised, but also to be praise-*worthy*. The *immediate* bystanders will judge us in function of what they see, know and experience themselves. Their praise (or their blame) can very **partial**, in the double sense of based on incomplete knowledge and biased by their own interests. An *ideal* (impartial) spectator, at the other hand, is supposed both to know everything and not to be affected by (the consequences of) our actions on ourselves nor on others. He shares the latter quality with *real* observers we could qualify as strangers. If an actor can suppose an indifferent observer would praise his action, he can qualify this action as "praise-worthy". If we want to know whether we are really worth what they say, we have no other option than to imagine ourselves in an indifferent spectator's position, to use the spectator, as it were, as a

"looking-glass by which we can, in some measure, with the eyes of other people, scrutinize the propriety of our conduct... We can be more indifferent about the applause, and, in some measure, despise the censure of the world; secure that, *however misunderstood or misrepresented*, we are the natural and proper objects of approbation" (TMS:112, our italics).

It is by regular observation of the others as judges that we eventually come to define so-called *rules of conduct*, i.e. "general rules concerning what is fit and proper either to be done or to be avoided" (TMS:159).

3.4.5. Rules of conduct

Applying this to the trust game, it can be supposed that an impartial observer would tell the seller neither to gain approval of the buyer, nor to better his condition *inconditionally*:

"In the race for wealth, and honours, and preferments, he may run as hard as he can, and strain every nerve and every muscle, in order to outstrip all his competitors. But if he should justle, or throw down any of them, the indulgence of the spectators is entirely at an end. It is a violation of fair play, which they cannot admit of. This man is to them, in every respect, as good as he: they do not enter into that self-love by which he prefers himself so much to this other, and cannot go along with the motive from which he hurt him. They readily, therefore, sympathize with the natural resentment of the injured, and the offender becomes the object of hatred and indignation" (Smith, TMS:83).

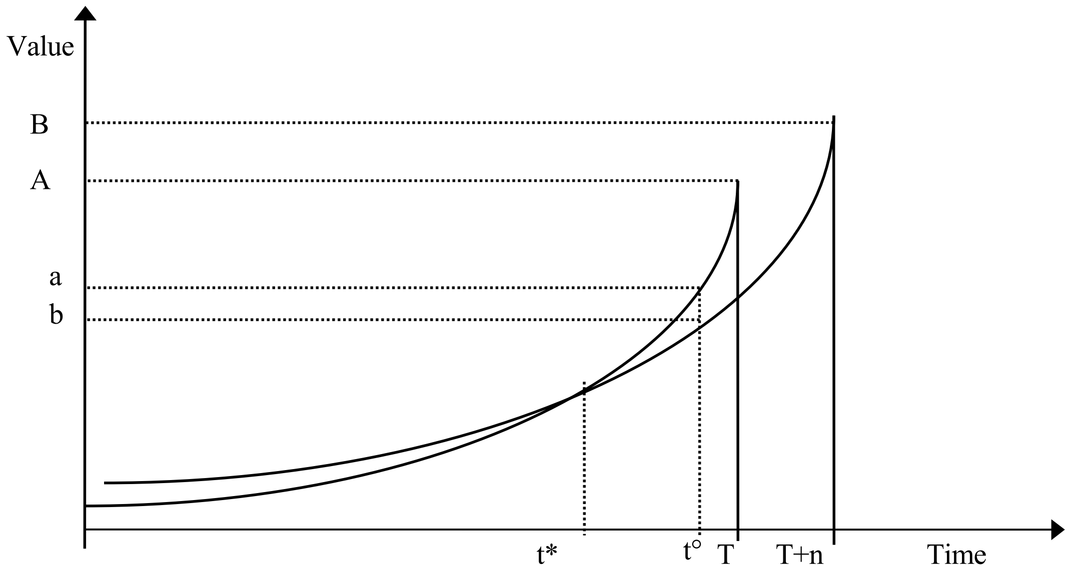
Hence, we can predict that, if a seller is indeed guided by the desire to be prais-worthy, he will *in principle* not act opportunistically. This being said, some caveats must be noted.

Opportunism as a loophole

First, even people who are well *acquainted* with the "correct" standards of right and wrong may still have a difficulty to apply them. The problem of temporary preferences remains:

"When we are about to act, the eagerness of passion will seldom allow us to consider what we are doing, with the candour of an indifferent person. The violent emotions which at that time agitate us, discolour our views of things; even when we are endeavouring to place ourselves in the situation of another, and to regard the objects that interest us in the light in which they will naturally appear to him, the fury of our own passions constantly calls us back to our own place, where everything appears magnified and misrepresented by self-love...” (Smith TMS:157).

Just before acting, the voice of the impartial bystander is overruled, “misrepresented by self-love”, and only “when the paroxysm of emotion...is fairly over, we can identify ourselves, as it were, with the ideal man within the breast” (TMS:158). Precisely “when it is of most importance that [our views] should be otherwise”, we lack the capacity to imagine ourselves in the place of the judge rather than in that of the actor. This is only another formulation of Ainslie’s temporary-preference problem: though we’d like to choose B in principle, we choose the alternative A in practice, for precisely at the moment we have to make the choice, A appears more rewarding than B. B is what an indifferent person would approve.



-- Figure 3.9. The seller’s temporary preference to cheat**--**

Figure 3.9. builds the bridge between temporary-preference theory and the trust game: If reward A represents “self-love” or the direct effect of “bettering our condition” and reward a is the discounted value of it at the moment of choice (t°), reward B stands for the total effect of our behaviour on an (imaginary) impartial spectator, and b is the discounted value of B at t°. Supposing that “when he is about to act” a seller will not be able to imagine himself in the other’s position, we can also assume that the full consequences of his actions will be realized only with a delay, and that the short-term preference at t° close to T will still be to sell a lemon.

Confronted with such a choice situation, and with the perspective of being confronted with similar situations in the future, the development of a personal rule may then be one of the mechanisms to overcome the temporary impulse to sell a lemon:

“It is reason, principle, conscience, the inhabitant of the breast, the man within, the great judge and arbiter of our conduct... who, whenever we are about to act so as to affect the happiness of others, calls to us, with a voice capable of astonishing the most presumptuous of our passions, that we are but one of the multitude, in no respect better than any other in it” (Smith, TMS:137)

In short, simple prudence may allow to overcome this first caveat, since we have changed, this time, the definition of the rewards: In Frank's mind, prudence was not enough since delayed material rewards were largely absent. The rule to "cheat whenever the probability of being detected is insignificant" is indeed not sustainable: prudent cheaters will eventually be detected, as the rule is not clearly specified and erroneous. However, if we suppose that people are ultimately not so much interested in material rewards as in the opinion of their imagined "indifferent spectator", prudent behaviour becomes, again, sustainable[28]. In Ainslie's words "the voice of the man within" is a much better *bright line* than "the probability of being detected", and precisely for this reason it becomes interesting: « Another person's example is potentially a robust boundary between one's long- and short-range interests just because it is not arbitrary; it is especially compelling insofar as his circumstances are like one's own » (1992:297).

In contemporary language, we can say that we model the seller's situation as two overlapping games, one being visible, another invisible. The game we see is the transaction between the seller and the buyer. The invisible game in which the visible game is embedded is the multiple-self-game we detailed above. The reason he gives for his action must be credible -to himself no less than to others. The first game is a one-shot game where selling lemons is clearly more profitable than selling good cars. The second game is a repeated full-information game where the player commits himself to co-operate with his future selves. His identity, simulating the profile of "the man within", is the stake in this second game. Both the visible game and the invisible game are overlapping, meaning that one's choice in the visible game is at the same time an identity-test.

Note that a person's identity is itself fundamentally an internalization, derived from daily experience, or the image we have of what indifferent others would summon us to do and not to do. We are far removed here from "conscience" as a product of social conditioning. Our standards of right and wrong are much more the product of experience than of what our parents have been trying to instill to us, or, in Smith's words,

"We do not originally approve or condemn particular actions; because, upon examina-tion, they appear to be agreeable or inconsistent with a certain general rule. The general rule, on the contrary, is formed, by finding from experience, that all actions of a certain kind, or circumstanced in a certain manner, are approved or disapproved of" (Smith, TMS:159).

[28] Remark that Smith's claim that the majority of people are "prudent" has been given the neo-classical reading that most people try to maximise their self-interest (e.g. Stigler 1971:265). Other economists (mis)used Smith to legitimate the assumption that economic actors are self-interested, or, in Smith's words, moved by *self love* or to *better their condition* (see also Hirschman 1976). As we know by now, these partial readings should be nuanced considerably: First, "bettering our condition" is a *socially informed* and *reasoned* objective rather than a mere innate impulse. Second, *prudent* behaviour involves a deliberate intervention precisely to curb impulses and, also, to restrain self-love.

Legitimate opportunism.

A **second** caveat is related to this "inductive" way in which people come to circumscribe moral rules to follow: To the extent an "impartial spectator" remains a product of social interaction, his judgments are inherently relative, they will differ between different social environments and between the different individuals as these draw on a memory of different experiences with social (dis)approval. However, an impartial spectator would not be worth that name if the individuals themselves didn't believe he was indeed objective, universal, *natural* in Smith's words. Accordingly, standards of right and wrong do not differ as much as the heterogeneity of people's (social) backgrounds.

Yet, the mere *belief* in the objectivity of concrete standards of right and wrong conduct does not contradict their historical relativity. Smith mentioned e.g. the defense of infanticide even by the doctrine of Greek philosophers, who were "led away by the established custom, and upon this, as upon many other occasions, instead of censuring, supported the horrible abuse, by far-fetched considerations of public utility" (TMS:210). Analogously, it may still be possible that a seller who sets out to realize praise-worthiness will sell a lemon: He may have found a convincing argument to do so, and this argument will allow him to "despise the censure of the world", and more particularly the buyer's contempt.

In casu, one possible argument a seller may use to deem a bystander's opinion less relevant is to consider him *misinformed*: 'he doesn't know the *whole* story', or worse, 'he is manipulated'. Another type of argument may be to attribute the seller's action to the buyer's fault: 'he should have known better', 'he should have paid attention'. A third argument may be that the other's well-being is "just" less important in comparison with that layer of humanity to which one belongs oneself.

The interesting inference is that guilt and shame not only sustain *honest* behaviour, they also sustain *opportunistic* behaviour: Since a *person* will not readily admit his erroneous behaviour, he will develop rationalizations and consequently persist behaving opportunistically.[29]

If so, we ought to redefine the types of sellers we characterized as "honest" resp. "dishonest" from figure 3.3. onwards. Just as the carrot-owners were wrongly categorizing the donkey who always chose the smaller carrot as "very stupid", the "dishonest" seller may indeed wrongly have been categorized that way by previous and scrutinizing buyers: Though some sellers may indeed be real opportunists, and others unable to control themselves, we believe that most see themselves as genuinely "honest" and "moral", abhorring opportunism like anyone else, and believing themselves to act *rightly* at least most of the time.

[29] These insights draw on Elster (1998), who emphasizes both "the dual role of emotions in decision making" and the two-way relationship between cognitions and emotions. See also Dimaggio and Powell (1991), who emphasize cognition as the main source of path dependency.

"it is so disagreeable to think ill of ourselves, that we often purposely turn our view away from those circumstances which might render that judgment unfavourable. He is a bold surgeon, they say, whose hand does not tremble whe he performs an operation upon his own person; and he is often equally bold who does not hesitate to pull off the mysterious veil of self-delusion, which covers from his view the deformities of his own conduct" (Smith, TMS:158).

Paradoxically, this very belief in their own fairness is what makes them relatively impervious to evidence that this is not the case. Their problem is not so much that they lack a conscience, but rather that the world as they know it is different from those of the buyers (and, eventually, outside observers), and that they are not constrained to reconsider why they believe their beliefs to be true.

Opportunism as self-deception

This 18th century argument has an actual descendant in cognitive dissonance theory, which, in the version of Akerlof and Dickens, was formulated as follows: People like to think of "themselves as 'smart, nice people'. Information that conflicts with this image tends to be ignored, rejected or accomodated by changes in other beliefs" (1982:308)[30].

However, if one looks at the problem of preference reversals from cognitive dissonance theory, a complication must be noted. Suppose someone has opted for an A which pre-empts his choice for a higher, delayed reward B. In order not to have to revise the illusion of being "smart and nice", such a person will then "rationalize" his choice, by making his belief system consistent with his past behaviour, rather than adjust his behaviour:

"Rather than see our own behaviour under so disagreeable and aspect, we too often, foolishly and weekly, endeavour to exasperate anew those unjust passions which had formerly misled us... and thus persevere in injustice, merely because we once were unjust, and because we are ashamed and afraid to see that we were so" (TMS: 158)

Note that Ainslie's account of personal rules is only one of the applications of cognitive dissonance theory: As persons, people cannot live with the idea that they believe B should be desired, whereas in practice they happened to opt for A (figure 3.9): Cognitive dissonance theory suggests, then, that they try to adjust either their behavior, or their desire. The first possibility (adjustment of behaviour) has been discussed at length in 3.4.2.. However, in our case the delayed reward is rather *imaginary*: the "indifferent spectator" is of course only a product of man's imagination, and hence it could be tempting to change one's belief system rather than to change one's behaviour.

[30] A more general definition is given by Tversky and Quattrone: "an individual who holds two or more cognitions (i.e. attitudes and beliefs) that are psychologically inconsistent will experience an uncomfortable state of tension, called dissonance. The individual will then be "driven" to reduce dissonance by changing one or more of the cognitions so that they are no longer inconsistent" (Quattrone and Tversky 1986:39).

However, even though it is possible that such cognitions can be selected by some sellers some of the time, temporary preference theory provides for some arguments to reduce the arbitrariness as concerns this selection process.

To begin with, "there are many fictions that can be chosen, but there is only one reality. If a person abandons that bright line as a criterion for self-reward, he is in danger of autism. The set of beliefs which the person qualifies as real must be protected from inflation by another set of private rules, governing what psychiatrists call 'reality testing' " (Ainslie 1986:155). The problem with a preference to consume untrue fantasies is, according to Ainslie, not so much that there is no real reward tied to them; if this was the problem, untrue images of an "impartial spectator" would indeed not be threatened with extinction. The problem with untrue fantasies is *an inflation of* the reward tied to them: In Ainslie's words, "they suffer from a shortage of scarcity" (1986:156). Rules to test the reality-value of a fantasy render it scarce, and thereby worthwile to consume[31].

This is, ultimately, why the voice of the *real* spectator (in casu the buyer) is so important: his opinion is the reality-test of the "man within". Though this reality-test can exceptionally be ignored, in which case the seller would succumb to self-deception, the general prediction must be that it won't.

Note that, to the extent the real spectator has some say in the pacing of the seller's self-reward for being praise-worthy, the concept of reputation enters again into the picture: it is by no means the *real* (discounted) effect of loss of reputation which will move the seller, however, this effect being (per definition) too small to motivate him. But the risk of losing one's good reputation is not necessarily unimportant as a pacer of self-reward: Precisely because buyers are free to enter the shop or to stay out of it, the trust they display by entering is an outstanding proof of the seller's praise-worthiness.

To sum up, supposing that, as Smith, as well as Akerlof and Dickens suggested, people value the belief that they are "smart and nice" people more than any other reward, they will let themselves guide by the impartial spectator rather than by some more imminent reward. However, this statement would not be very worthwile if we would not be able to determine how this impartial

[31] A Smithian idea which comes closest to this statement is the desire to persuade. "Men always endeavour to persuade others to be of their opinion even when the matter is of no consequence to them. If one advances any thing concerning China or the *more distant moon* which contradicts what you imagine to be true, you immediately try to persuade him to alter his opinion", (Smith LJ:352) Elsewhere in the Lectures, he sets this desire at an equal footing with hunger, thirst, and the passion for sex (LJ:527). Finally, in the Theory of Moral Sentiments, he equates the ability to persuade with the ability to lead and direct other people. Further, "It is, perhaps, the instinct upon which is founded the faculty of speech, the characteristical faculty of human nature. No other animal possesses this faculty, and we cannot discover in any other animal any desire to lead and direct the judgment and conduct of its fellows. Great ambition, the desire of real superiority, of leading and directing, seems to be altogether peculiar to man, and speech is the great instrument of ambition, of real superiority, of leading and directing the judgments and conduct of other people" (TMS:336). Like in Ainslie's view, the point of persuading is not really the drive to arrive at an accurate view of reality, but rather, it seems, to acquire a measure of control over what is in principle uncontrollable.

spectator is construed. We argued that the "man within" is a product of "objective" reality as well as of imagination. The fact that the buyers have some power over allocation of payoffs doesn't result in a "survival of the fittest seller" as in Frank's model, but it results in the "survival of the fittest cognition": the buyer's opinion is the most rewarding test of the seller's praise-worthiness, the most important represenant of an "impartial spectator" with an opinion which happens to be consistent with the seller's long-range interest, i.e. his reputation.

As a consequence, we have defined two types of opportunists. First, there are those sellers who have found a loophole to defect, i.e. an argument to sell a lemon in the expectation that "in normal times" they won't admit themselves to do so. Second, there are those sellers who have developed a rationale for selling lemons. As long as there will be no good motive to question this rationale, the reluctance to question one's dignity will enable one to ignore much available evidence and perhaps even to "harden" one's view. In such a situation self-deception may be of a rather eternal kind. It can be expected, however, that cognitions about the impartial observer which are much less compatible with opportunistic behaviour will dominate the population of sellers.

* * *

We conclude our rereading of Smith through the spectacles provided by temporary-preference theory by making four comments on Frank's "solution" to the trust game:

First, we agree with Frank that, since there are no bigger, delayed rewards from selling good cars in an anonymous market, we should complement material incentives with "moral sentiments".

Second, we think Smith would disagree with Frank that a person endowed with a conscience would be affected by counteracting moral sentiments precisely at the moment of choice; according to Smith, we are *incapable* of taking sufficient distance from ourselves at the moment of choice,. Our moral sentiments, embodied in a so-called "indifferent spectator", will exert their influence only *after* having sold the lemon.

Third, whereas Frank's account revolves around *guilt*, Smith's account revolves around *shame*. The seller of a lemon feels shame at the thought of what the buyer must think of him, and if he adopts the rule never to sell a lemon, he adopts it in order *not* to feel ashamed.

Fourth, in Frank's eyes, the "second allocation mechanism", viz. the buyers scrutinizing the sellers, incites parents to educate their kids "with a conscience", and, as a result, the demographic importance of unconscienscious, merely prudent sellers, will decrease. We argued to the contrary that most people do have (or act as if they think they have) a conscience, but this would *add*, rather than reduce the problem of obtaining a cooperative outcome as long as sellers weren't motivated to test whether the buyers' definition of opportunism would be part of that belief system they accept as true. The buyers' ability to select a particular seller is precisely what motivates them to test this.

Precisely because the second allocation mechanism is active, sellers will "work up" the suitable cognitions which make use of it, and these cognitions provide for the direct reasons why they discipline their behaviour, even in situations where it is, everything considered (and properly discounted), more advantageous to sell a lemon.

3.5. illustrations

3.5.1. Hirschman's maintenance habit.

The argument can be illustrated by reviewing Albert O. Hirschman's well-known comparison between different types of transport: Just as honesty (Gambetta 1988), maintenance of capital goods is not directly (or sufficiently) beneficial for the economic actors, once consumers have entered the shop, resp. capital goods are acquired. Hirschman observed that at least some capital goods

"*must* have maintenance because its absence carries with it a very high penalty, that is, leads to accidents or immediate breakdown rather than to slow deterioration in the quantity and quality of output... The fact that the performance of the airlines in Columbia is excellent, that of the railroads mediocre, and that of the roads outright poor can be explained in terms in this [above mentioned] criterion" (1971:58).

A decade later, he specifies the "high penalty": "narrow-lattitude tasks will, if performed poorly and (ex hypothesi) disastrously, give rise to strong public concern and outcry -voice" (1984a:100). Yet, Hirschman presented the subject already in his *Strategy of Economic Development* (1958), mainly as part of an argument to favour high-tech industrialization. More in particular, he argued that high-tech contrasts with more labour-intensive industry not only because it demands relatively much costly capital, but also because of a much less visible effect:

"The basic difficulty about maintenance of capital... is that it is a *preventive* activity which must be performed at fairly long intervals that are neither known with precision nor signaled by the capital itself. For maintenance to be effective, people must be ... made to act *as though* it had to be undertaken at precise intervals, suppressing their better knowledge that deferment by a day, a week, or a month may not matter; they must organize this fiction, submit to it, and set up a signalling system to enforce it... underdeveloped countries may well make a surprising success of ventures with a complicated technology which must be maintained in top working order. It is in these industries that the maintenance habit can be acquired and from there spread to the rest of the economy" (1958:142)" (Hirschman 1958:141).

Hirschman's general "blessing in disguise"-argument is well-taken. However, his particular argument about the disguised blessings of transport by air is doubtful. We would like to give three comments.

First, supposing that mere material incentives don't suffice to solve the maintenance problem, you should try to generate an extra-economic motive (e.g. maintenance as a *habit*). However, it is somewhat problematic to suppose, as he does, that such an extra-economic motive would be generated by an industry in which maintenance is a clear economic necessity, and hence, where no extra-economic motive is required.

The only particularity of airlines is that economic incentives are not only affected by prices but also by other types of signals. Indeed, the lack of maintenance will be clearly (and cheaply) signalled in the case of airlines, though not in a standard way. Accidents are as spectacular as donkeys always choosing the smaller carrot. They become a news headline, and even if the press is all but interested in a good functioning airline (it may indeed only be interested in selling newspapers) the result is that, if the airline managers want to continue to attract customers, they must prevent the potential customers to be shocked, and hence they must prevent the occurence of accidents, and hence they must invest resources in maintenance. In terms of ordinary welfare economics, one could say that a free press obliges the airline management to internalize externalities. It could be argued that, in case of externalities, the free press (or, as Hirschman wold have it, *voice*) complements the price signal.

Second, a press-reported aircrash does not only incentivate to invest in maintenance. A safe trip by air is undoubtedly the product of a cooperative effort by many people and organizations. Hence, if man's first preoccupation is with *dignity*, it may be suggested that the first type of reaction of airline management to a crash will be to develop convincing arguments that, in case of an accident, *they* should not be blamed. As a result, an airline accident will become an occasion to discuss, in public, each party's responsibility, how to prevent future accidents, etc.. If car accidents would be as spectacular as airplane crashes, every car would carry a black box, we would have a much more detailed idea of how to prevent them and, consequently, dispose of a much safer traffic system. In other words, transport by air is much safer because we simply *know* much more about aircrashes: Aircrashes are so spectacular that it is difficult to be as fatalist about them as about car accidents. Motivated by the risk to lose their dignity through a public scandal, the airline management as well as other parties will try to do not only their level best but also clearly outline each party's responsibility. The unintended result of this dynamic will be better knowledge, better prevention and safer airtrips.

Third, it may be, as Hirschman seems to suppose, that "habits" can be conceived as behavioural patterns which may "unthinkingly" reproduce and even spread over the economy, whether or not there is an underlying motive to *enact* them. Such a supposition is not sustained by temporary-preference theory, however. Though this theory provides an explanation for the rationality of behavioural patterns, it also suggests that behavioural patterns need (long-term) incentives to emerge. People can make themselves act as though maintenance "had to be undertaken at precise intervals, suppressing their better knowledge that deferment by a day, a week, or a month may not matter", if there is an appropriate incentive to re-organize their cognitions in this way.

More specifically, maintenance as a pattern will most probably emerge in industries where the incentive to maintain is neither too low, nor too high: If it would it be too high, as in the case of airlines, one doesn't need to reorganize one's mind, every single act of negligence will indeed almost certainly be penalized. Would it be too low, there wouldn't be a long-run advantage in maintenance, and hence no need to develop such a routine. Only in the intermediate case can one expect that people will spontaneously develop additional cognitions in order be able to consume the long-run reward even in the face of many competing but lower imminent rewards. Whether this "intermediate case" is transport by train is an empirical matter, of course.

In any case, the probability that a pattern, supposedly having emerged in industry x, will spread to industry y, must depend on there being a sufficiently big (but not too big a) motive in y too, to copy the maintenance routine developed in x. While this statement would suggest that we'd better forget about maintenance as a routine behaviour which can be generated in one industry and "from there spread to the rest of the economy", this is not entirely true, as we will try to argue on the basis of a more relevant illustration.

3.5.2. Famines, democracy and poverty.

"famines seem to have never occurred in a democratic country, which permits opposition parties and adversarial politics and does not suppress and independent press. If the government has to go to the polls to secure re-election and has to withstand criticism in the media and in the legislatures, it would have good reasons to act quickly and effectively in preventing threatening famines" (Sen 1997:29)

Sen's argument about the relationship between political regime and the occurence of famines is well-known, but not much disputed. The argument Sen makes appears to be consistent with "sound" economic thinking, as it explains voters' as well politicians' behaviour in terms of incentives. However, the relationship between democracy and famine prevention is likely to involve a considerable amount of nonwealth-maximizing elements. This argument can be made for political incumbents as well as for voters. We will discuss each in turn[32].

Famines and democracy

To begin with, if no famine has *ever* occured in a democratic country, this implies that *not any* democratic government has really *tested* whether there is indeed a relationship between allowing a famine to occur and electoral defeat[33]. This suggests that prospect of electoral defeat is an indirect rather than a direct negative reward. More remarkably still is that the relationship between democracy and famine prevention is independent of whether the government was indeed the cause or can be accused of having *caused* the famine or the crash: democratic governments appear to take their responsibility in both cases. Apparently, democratization changes the perception of reality and how to change it, rather than (and besides changing) the structure of real rewards[34].

---

[32] Another strategy to discuss Sen's thesis is to test it empirically; such a strategy would confront us with the methodological problem of how to "measure" democracy (or the degree of democracy); this seems to be an non-trivial issue to test the truth-value of the thesis, as it appears that three of the four African success-stories in famine prevention (as exposed in Drèze 1995) were practically one-party systems (at least according to the criteria used in Ferree, Singh and Bates (1996). Sen's thesis does *not* imply, of course, that authoritarian regimes are *not* performing well in terms of hunger prevention, only that democracies cannot not perform well.

[33] It could, besides, be argued that the political market does not, in this aspect, differ from e.g. the airline market.

[34] To be sure, this statement is easily to criticize, the non-occurence of famines in democracies might simply be seen as evidence that a simple rational-choice explanation has good predictive value. Yet, while we don't want to ignore the direct linkage, this does not eliminate the possibility of an indirect linkage.

One reason to account for this change in perception is that democracies differ from authoritarian regimes not only as concerns the regular organization of elections: Another distinguishing characteristic is undoubtedly the existence of a free press and free speech (albeit that "freedom" is undoubtedly a matter of degree rather than an absolute quality). It is established fact now that an independent press is one of the most effective "early warning"-devices of famines (Drèze 1995, Ravallion 1997). Conversely, in non-democratic regimes, the information may be delivered by people whose destiny is too closely dependent on the superiors they must inform (e.g. China during the 1959-61 famine).

Another is that only democratic regimes are *motivated* to take their responsibility so as to prevent famines. An electoral victory is a long-term objective and, moreover, the short-run advantages of being in power may conflict with the long-run advantage of staying in power. Just as "reputation", "electoral victory" may be desired, but direct attempts to realize them are bound to be unsuccessful. The long-term objective will, then, only indirecly influence action, *viz.* by influencing one's reality-testing rules and, subsequently, one's belief system[35]. Note that this argument suggests that voters' convictions matter in a different way than as suggested by North: Voters seem to select not so much the political candidates or the political parties, but rather the way in which the political actors organize their belief systems –though it is because citizens can vote that politicians will think differently. Conversely, famines occur in authoritarian regimes not merely because they don't threaten political stability –some of them in fact did- but rather because those regimes' belief systems are not constrained by the voters' opinions.

Some evidence for this hypothesis can be found in the grim history of famines.

Compare a democratic ruler's position with that of the colonial administrators in pre-independence India:

"Warren Hastings, who had tackled a famine in Bengal in 1783-4 by using public channels for moving food into the region, was rapped on the knuckels by Colonel Baird-Smith for not having understood his Adam Smith, adding that Hastings could 'scarcely have been expected' to have absorbed Adam Smith so soon (1783) after the publication (1776) of the Wealth of Nations. The basically non-interventionist famine policy in India lasted late into the nineteenth century, changing only around the last quarter of it" (Sen 1981:160)

We may assume that Baird-Smith was, if not himself endowed with a well-functioning conscience, at least making a conscientious argument, as he probably referred to Adam Smith's apology of a food market freed from any political intervention. Free trade was a sound principle, it was advocated by sound thinkers, and Hastings must have been looked at as rather impulsive and *weak* for not having been able to withstand the local pressure for public intervention in the food market. It should be noted, to be sure, that Adam Smith's ideas about famines were read too

[35] This is how we interpret Sen's statement that "our conceptualization of economic needs depends on open public debates and discussions, and the guaranteeing of those debates and those discussions requires an insistence on political rights" (1994:32).

selectively by Baird-Smith to do him justice (Sen 1995)[36] . Nevertheless, this does not imply that the famine would not have occured if the moral or the intellectual capacities of Baird-Smith would have been of a higher level. Rather, he seems to have lacked sufficient incentive to consider the plight of the famine deaths as an ultimate reality-test of his own (and Smiths) insights about food markets.

Almost two centuries later, the Colonial administration was equipped with a “Famine Code”, which stipulated an alternative (interventionist) food market policy to be applied in case of a famine. Unfortunately, the Great Bengal Famine of 1942-3, eventually causing “around 3 million” casualties, was never officially declared a famine. Part of the explanation for this astonishing fact is that the authorities attributed famines to Food Availability Decline (FAD), and since, at the time, the shortage in production was not deemed catastrophic, nothing signalled disaster. Remarkably, after the severity of the famine finally did surface, the available data on production were reconsidered and revised downwards, not FAD itself. The Great Bengal famine was now ready to become known, in the history of food crises, as being caused “when floods destroyed the rice crop, costing some 2 million to 4 million lives” (Sen 1981:83). The most important problem was, again, not so much the moral or intellectual capacity of the expatriates, but rather a lack of incentives to question one’s own mental models.

Famines and poverty

For a famine to become an important political issue, we need to assume that voters are also concerned with issues which don’t affect their immediate self-interest: Why would the majority of voters bother about the problem of a minority ? Indeed, famines rarely affect 5% and at most 10% of the population (Sen 1994:36) and are anything but random phenomena, affecting only some specific occupational groups, notably typically those who lack political voice. In this way, Sen’s example of famines seems to (partially) contradict North’s prediction that the laws enforced by a democratic state will be biased towards the private interests defended by organized interest groups who value the net benefits of political action higher than the net benefits of private action.

The assumption that people may be moved by another’s sorrow is of course not very controversial, at least as long as we stick to the subject of famines and similar disasters “that are easy to understand, in which sympathy can take an especially immediate form” (Sen 1994:35). Aircrashes belong to the same category. To push this idea further, it could be supposed that rational democratic politicians would not only respond as quickly as possible to such disasters, they could also be expected to be as reluctant as possible to fight acute and lasting famine *vulnerability*: A situation of mass poverty is indeed the ideal breeding ground for famines, and hence for mediatized opportunities to show one’s political qualities. Though this grim prediction seems to be confirmed by the sharp contrast between success in preventing famines and failure in

[36] But perhaps he could “scarcely have been expected” to have absorbed Adam Smith so soon (1783) after the publication (1776) of the Wealth of Nations...

reducing famine vulnerability in India (Drèze 1995:155), other elements suggest a much less perverse relationship between acute crises and chronic problems.

First, as most famine deaths are *directly* caused by "ordinary" diseases, the performance of the health care sector is part of a famine prevention program. In other words, successful famine prevention not only depends on highly visible "crisis management" and special famine-prevention programs, it also depends on activities which require systematic (unmediatized) attention even in ordinary times, if they are to perform as expected if a famine should show up. As a result, unmediatized poverty is partly tackled merely as a by-product of the effort to tackle mediatized famines.

Second, a threatening disaster is not only an opportunity for political success, it is also a major moment of information gathering and public debate, and as such an imminent crisis may generate inputs to resolve chronic problems -even in other domains. To illustrate, the English war-government only realized to what extent pre-war poverty reports suffered from overoptimism after it had organized some evacuations (to prevent the town-dwellers from being bombed). This information was later used, not only to design an effective war-economy, but also to establish the post-war system of social security:

"...the evacuation of mothers and children and the bombing of homes during 1939-40 stimulated inquiry and proposals for reform long before victory was even thought possible. This was an important experience, for it meant that for five years of war the pressures for a higher standard of welfare and a deeper comprehension of social justice steadily gained in strength. And during this period, despite all the handicaps of limited resources in men and materials, a big expansion took place in the responsibilities accepted by the state for those in need" (Titmuss 1950:507-8).

Third, as mentioned above, disasters evoke sympathy for the victims (and antipathy for the perceived cause of their distress), day-to-day differences are set aside and people with whom one felt perhaps the faintest commonalty suddenly are regarded as equals. Such a re-formulation of identity and commitment may have a long-lasting character. Again, we may refer to war-time England here:

"The reality of military disaster and the threat of invasion in the summer of 1940 urged on these tendencies in social policy. The mood of the people changed and, in sympathetic response, values changed as well. If dangers were to be shared, then resources should also be shared. Dunkirk, and all that the name evokes, was an important event in the war-time history of the social services. It summoned forth a note of self-criticism, of national introspection, and it set in motion ideas and talk of principles and plans." (Titmuss 1950:508)

In England, the name of "Dunkirk" evokes probably not so much the defeat, as French historians would have it, but rather the moment at which Winston Churchill summoned his compatriots to write world history[37] and prevent the whole world to "sink into the abyss of a new Dark Age,

[37] "If we can stand up to [Hitler], all Europe may be free and the life of the world may move forward into broad, sunlit uplands. But if we fail, then the whole world, including the United States, including all that we have known and cared for, will sink into the abyss of a new Dark Age, made more sinister, and perhaps more protracted, by the lights of perverted science. Let us therefore brace ourselves to our duties, and so

made more sinister, and perhaps more protracted, by the lights of perverted science". As we know by now, not only was this war won, but also, and in the mean time, did life expectancy increase to a degree unequalled by other decennia of this century (Drèze and Sen 1989:182). The same values which united war-time England and which caused the English to withstand "barbary" and to increase wartime life-expectancy also underscored the post-war construction of a full-blown system of social security, one of its results being a further increase in life-expectancy.

Some evidence points to similar dynamics from the disaster of a famine over an emergency program to prevent it, to a permanent system of poverty relief: Ravallion (1997:1236)) mentions the Indian "Employment Guarantee system" (Maharashtra), Drèze (1995:589) invokes the African experiences of Cape Verde and Botswana.

To sum up, besides being rational, people are also sympathizing with others' plight and capable to learn and to re-assess their commitments. These characteristics sustain Sen's argument for the primacy of political rights over economic ones: giving an individual political rights will not only be advantageous for this person but also for all others this individual feels acquainted with (Sen 1994). Note that this argument can also be found in North. More controversially, we argued that exceptional crises seem to inhere exceptional opportunities for long-term development, because they incite people to test alternative definitions of identity. Note, however, that this relationship depends in a crucial way on the degree to which a society may be called democratic. To the extent this is not the case, political leaders don't feel the necessity to outwit in excellence during disaster, people possessing relevant information may not be heard and a spirit of fatalism instead of activism may take leaders as well as citizens in its grip.

Note, finally, that the Colonial administrators serving during the Bengal famine in 1942 and the administrators of the poverty programs at home probably shared the same education and background, both groups had undoubtedly good reasons to see themselves as belonging to the civilized part of humanity. Their widely different response to poverty can, then, only be explained by the widely different structure of incentives they faced.

This wartime example is comparable to the different weight actually given to the well-being of (voting) nationals and nonnationals. The very same political leaders can consistently preserve the Belgian Social Security System at the one hand, and do nothing to prevent an upsurge of violence which *directly* caused already 1,5 million people to loose their lives, but which has not ended up to now. This arrogance can partly be explained by a lack of information, but the belief in their own integrity undoubtedly also caused Belgian politicians to *ignore* the information available. The direct reason to withdraw the Belgian troups from Rwanda in 1994 was the killing of 10 of our *own* "national" voters. Had we not had these casualties, there wouldn't even have been a parliamentary commission to discuss the Rwandan genocide, its precedents and its consequences.

---

bear ourselves that if the British Empire and its Commonwealth last for a thousand years, men will still say, 'This was their finest hour'" (Churchill 1940:37).

3.6. Conclusion

We started our inquiry into “human nature as we know it” by specifying the “trust game” as an archetypical social dilemma, i.e. a situation where actors would not reach the cooperative outcome if they would act as assumed by conventional rational choice theorists. The trust game “as such” is of course not solvable, we need to introduce more context.

A first way to do this is to assume the existence of a third party, possessing the means to *enforce* honesty. This solution, as worked out by North, *inter alia* served to point out that the issue of nonwealth-maxizimizing motives is not dissipated by invoking a third party: this third party cannot be assumed to work effectively without it being considered as “neutral” by both parties, which implies that they consider this third party (1) not to be motivated by self-interest, and (2) to have a “true” image of reality. While (1) could eventually be guaranteed by democratic structures, (2) was more problematic.

A second route worth exploring has been that of “generalized morality”, as worked out *inter alia* by Platteau (1994): in contexts where individuals’ preference structures are partly determined by the “right” moral norms instilled to them during childhood, the trust game ceases to be a major impediment to economic growth. The problem with this conception is, however, that “culture” is conceptualized, in this route as well as in the former one, as a somewhat “blind” and unwieldy force; people are not insensitive to “moral considerations”, but acquiring the “right” moral norms seems to be a matter of historical accident and, by inference, market-based economic development a matter of good luck (Gambetta 1988b, Sabel 1993).

A third route explores a slight departure from the anonymous market: Suppose “honesty”, or “having a conscience” can be reliably signalled to customers. This implies at least that the “honest” are not weeded out in an evolutionary “survival of the fittest” game. Such a situation is especially relevant if we assume, at the pico-economic level, that (1) people cannot choose to become honest, but (2) their parents can choose to raise them with or without a conscience. These assumptions are sufficiently weird to explore a fourth, but related route.

This fourth route emphasizes that most people at least want to cultivate the belief of being not only smart but also *nice*. Conscientiousness is not the issue, the problem is rather that the “right” thing to do will also depend on one’s conception of reality. We concur with rational-choice theorist’s claims that one’s definition of “opportunism” and “honesty” is part of one’s belief system (Boudon 1997) and that what one believes is necessarily a by-product, i.e. something which cannot be believed simply because it would be interesting to believe it; beliefs must be able to withstand validity-tests.

The argument we developed boils down to the statement that people will “work up” cognitions sustaining the moral need to cooperate when the incentive to do so is neither too high, nor too low. Would the incentive be too high, cooperation would be the result without any need for moralizing. Would it be too low, there wouldn't be any incentive to take the other's opinion serious. But whenever a buyer is deemed to have some incidence on the seller's success, it will be interesting to consider his trust as evidence of the seller's praise-worthiness. The evolutionary game is not "realized" in the real world, it is "realized" only in the seller's head.

Finally, we explored the dynamics of nonwealth maximizing motives, first by commenting Hischman's argument about the maintenance habit, and second by looking at Sen's contentions about the relationship between democracy and famines. The basic theoretical idea is that, complementary to North's view, change occurs not (only) at the margin, but precisely at moments of crisis, moments which incite and enable people to explore different ways of relating to eachother[38]. To the degree such moments occur in reality, they will threaten the path-dependent character of institutional change, and by inference the development-pessimism which goes with it.

[38] Note that our argument does not contradict the idea that cooperation can be "engineered" by formal institutions. Rather, it says that these formal institutions are oftentimes given the freedom to engineer only in times of crises.

Bibliography

Ainslie, G. 1986 “Beyond microeconomics. Conflict among interests in a multiple self as a determinant of value” in: Elster, J. (ed.) *The multiple self* Cambridge:Cambridge University Press, pp. 133-176.

Ainslie, G. 1992 *picoeconomics; the strategic interaction of successive motivational states within the person* Cambridge:Cambridge University Press.

Akerlof, G.A. 1986 [1970] “the market for “lemons”: quality uncertainty and the market mechanism” in: J.B. Barney and W.G. Ouchi (eds.) *Organizational economics* San Fransisco:Jossey-Bass Inc. pp. 27-39.

Akerlof, G. A. Dickens, W.T. 1982 “The economic consequences of cognitive dissonance” in: *American Economic Review* 72(3) (June), pp. 307-19.

Akerlof, G. A. 1983 “Loyalty filters” in: *American economic review*, 73(1) (March), pp. 54-63.

Amsden, A.H. 1989 *Asia's Next Giant: South Korea and Late Industrialization.* New York: Oxford University Press.

Basu, K. 1995 “Civil institutions and evolution: concepts, critique and models” in: *Journal of Development Economics* 46, pp. 19-33.

Bates, R.H. 1996 “Institutions as investments” in: *Journal of African Economies* 6 (3), pp. 272-87.

Becker, G. S. 1996 *Accounting for tastes* Cambridge MA/London England:Harvard University Press.

Berger, P.L. Luckmann, T. 1967 *The social construction of reality; a treatise in the sociology of knowledge* Anchor Books, Doubleday.

Boudon, R. 1996 “the ‘cognitivist model’: a generalized ‘rational-choice model’” in: *Rationality and society* 8(2): 123-150.

Boudon, R. 1997 “The moral sense” in: *Rationality and Society* 12,1 (March), pp. 5-24.

Bowles, S. 1998 “Endogenous preferences: the cultural consequences of markets and other economic institutions” in: *Journal of Economic Literature* XXXVI (March), pp. 75-111.

Churchill, W. *Mr.Churchill Speaks* London:Cassell and Company Ltd.

Coleman, J.S. 1990 *Foundations of social theory* Cambridge MA:Harvard University Press.

Collier, P. Gunning, J.W. 1999 “Explaining African Economic Performance” in: *Journal of Economic literature* XXXVII (March), pp. 64-111.

Dasgupta, P. 1988 “Trust as a commodity” in: Gambetta, D. (ed) *Trust; making and breaking cooperative relations* pp. 49-72.

De Boeck, F. 1998 “Domesticating diamonds and dollars: identity, expenditure and sharing in southwestern Zaire” (paper written for colloquium “gifts and interests”, Faculty of Philosophy, Louvain (3-4 April).

DiMaggio, P.J. Powell, W.W. 1991 "introduction", in: Powell, W.W. DiMaggio, P.J. *The new institutionalism in organizational analysis*, Chicago/London:University of Chicago Press, pp. 1-38.

Drèze, J. 1995 "Famine prevention in India" in: J. Drèze, A. Sen and A. Hussain (eds.) *The political Economy of Hunger; selected essays* Oxford:Clarendon Press, pp. 69-177.

Drèze, J. 1995 "Famine prevention in Africa: some experiences and lessons" in: J. Drèze, A. Sen and A. Hussain (eds.) *The political Economy of Hunger; selected essays* Oxford:Clarendon Press, pp. 554-604.

Drèze, J. Srinivasan, P.V. 1997 "Widowhood and poverty in rural India: some inferences from household survey data" in: *Journal of Development Economics* 54, pp. 217-34.

Dupuy, J.-P. 1992 *Introduction aux sciences sociales; logique des phénomènes collectifs* Paris:Editions marketing.

Elster, J. 1983 *Sour grapes; studies in the subversion of rationality* Cambridge, Cambridge U.P..

Elster, J. (ed.) 1986 *The multiple self* Cambridge:Cambridge University Press.

Elster, J. 1989 *The cement of society; a study of social order* Cambridge U.P., Cambridge.

Elster, J. 1994 "rationality, emotions and social norms" in: *Synthese* 98:21-49.

Elster, J. 1996 "rationality and the emotions" in: *Economic Journal* 106 (September), pp. 1386-97.

Elster, J. 1996b "More than enough" in: *University of Chicago Law Review* 64(749), pp. 749-64.
Elster, J. 1998 "Emotions and economic theory" in: *Journal of economic literature* XXXVI (March), pp. 47-74.

Ferree, K. Singh, S. Bates, R.H. 1996 "Political institutions and economic growth in Africa" Harvard University Department of Government.

Frank, R.H. 1987 "If homo economicus could choose his own utility function, would he want one with a conscience ?" in: *American economic Review* 77(4), pp. 593-604.

Frank, R.H. 1988 *Passions within reason; the strategic role of the emotions* New York - London: W.W. Norton&Company

Frank, R.H. Gilovich, T. Regan, D.T. 1993 "Does studying economics inhibit cooperation ?" in: *Journal of economic perspectives* 7 n° 2 (Spring), pp. 159-171.

Frank, R.H. Gilovich, T. Regan, D.T. 1996 "Do economists make bad citizens ?" in: *Journal of economic perspectives* 10 n° 1 (Winter), pp. 187-192.

Gambetta, D.(ed.) 1988a *Trust; making and breaking cooperative relations* New York/Oxford:Basil Blackwell.

Gambetta 1988b "can we trust trust ?" in: Gambetta, D.(ed.) 1988a *Trust; making and breaking cooperative relations* New York/Oxford:Basil Blackwell, pp. 213-237.

Granovetter 1985 "Economic action and social structure: a theory of embeddedness" *American Journal of Sociology* 91:481-510.

Granovetter, M. 1992 "Economic institutions as social constructions: a framework for analysis" in: *Acta sociologica* 35:3-11.

Harsanyi, J.C. 1977 *Rational behavior and bargaining equilibrium in games and social situations* Cambridge, Cambridge U.P.

Hirschman, A.O. 1958 *The strategy of economic development* New Haven, Yale University Press.

Hirschman, A.O. 1971 *A bias for hope: essays on development and Latin America* New Haven:Yale University Press.

Hirschman, A.O. 1977 *The passions and the interests: political arguments for capitalism before its triumph* Princeton:Princeton University Press.

Hirschman, A.O. 1982 "Rival interpretations of market society: civilizing, destructive, or feeble ?" in: *Journal of Economic Literature* XX(December), pp. 1463-84.

Hirschman, A.O. 1984a "A dissenter's confession: 'the strategy of economic development' reconsidered" in: Meier, G.M. Seers, D. *Pioneers in development* World Bank:Washington, pp. 87-111.

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

Hobbes, T. 1966 [1651] *philosophical rudiments concerning government and society* in: William Molesworth (ed.) *The English works of Thomas Hobbes of Malmesbury* vol. II. Germany:Scientia Verlag Aalen.

Knight, F.H. 1946 [1921] *Risk, uncertainty and profit* Boston/New York: Houghton Miflin Cy.

Kreps, D.M. 1990 *A course in microeconomic theory* New York, Harvester Wheatsheaf.

Kreps, D.M. 1996 "Corporate culture and economic theory" in: P.J. Buckley and J. Michie (eds.) *Firms, organizations and contracts; a reader in industrial organization* Oxford:Oxford University Press, pp. 221-75.

Laband, D.N. Beil, R.O. 1999 "Are economists more selfish than the other 'social' scientists ?" in: *Public choice* 100(2), pp. 85-101.

Margolis 1982 *Selfishness, altruism and rationality* Cambridge:Cambridge University Press.

North, D.C. 1990 *Institutions, institutional change and economic performance* New York:Cambridge University Press.

Ostrom, E. 1998 "A behavioral approach to the rational choice theory of collective action" in: *American Political Science Review* 92/1 (March), pp. 1-22.

Piore, M.J. Sabel, C.F. 1984 *The second industrial divide; possibilities for prosperity* New York:Basic Books.

Pollack, R.A. 1985 "A transaction cost approach to families and households" in: *Journal of Economic Literature* XXIII (June) pp. 581-608.

Platteau, J.-Ph. 1994 *Behind the market stage where real societies exist* Journal of Development Studies 30:3-4, pp. 533-577, 753-817.

Putnam, R.D. 1993 *Making Democracy Work, Civic Traditions in Modern Italy* Princeton, Princeton University Press

Quattrone, G.A. Tversky, A. 1986 "Self-deception and the voter's illusion" in: Elster, J. (ed) *the multiple self* pp. 35-58.
Rachlin, H. 1995 "Self-control: beyond commitment" in: *Behavioral and Brain Sciences* 18/1, pp. 109-59.

Ravallion, M. 1997 "Famines and economics" in: *Journal of Economic Literature* XXXV (September), pp. 2105-42.

Sabel, C.F. 1991 "constitutional ordering in historical context" in: F.W. Scharpf (ed) *games in hierarchies and networks* Boulder CO:Westview Press, p. 65-124.

Sabel, C.F. 1993 "Studied trust: building new forms of cooperation in a volatile economy" in: Swedberg, R. (ed.) *Explorations in economic sociology* New York:Russell Sage Foundation, pp. 42-64.

Sally, D. 1995 "Conversation and cooperation in social dilemmas; a meta-analysis of experiments from 1958 to 1992" in: *Rationality and Society* 7:1 (January), pp. 58-92

Scharpf, F. W. 1989 "Decision rules, decision styles and policy choices" in: *Journal of Theoretical politics* 1(2), pp. 149-176.

Schelling, T. 1996 [1980] "The intimate contest for self-command" in: A.P. Hamlin (ed) *Ethics and economics* vol.I, Cheltenham:Edward Elgar publishing ltd. pp. 532-56.

Sen, A.K. 1979 [1976] "Rational fools: a critique of the behavioural foundations of economic theory" in: Harris, H. (ed.) *Scientific models of man; the Herbert Spencer lectures 1976,* Oxford:Clarendon Press, pp. 1-25.

Sen, A.K. 1981 *Poverty and famines; an essay on entitlement and deprivation* Oxford:Clarendon Press.

Sen, A.K. 1983 "Carrots, sticks and economics: perception problems in incentives" in: *Indian Economic Review* XVIII(1), pp. 1-16.

Sen, A.K. 1985 "Goals, commitment, and identity" in: *Journal of Law, economics and organization* 1(2) (Fall), pp. 341-55.

Sen, A.K. 1994 "Freedom and needs; an argument for the primacy of political rights" in: *The New Republic* (January 10&17), pp. 31-38.

Sen, A.K. 1995 "Food, economics, and entitlements" in: J. Drèze, A. Sen and A. Hussain (eds.) *The political Economy of Hunger; selected essays* Oxford:Clarendon Press, pp. 50-68.

Sen, A.K. 1997 "Wars and famines: on divisions and incentives" in: *Disarmament* 19,3, pp. 18-37.

Simon, H.A. 1981 *The sciences of the artificial* Cambridge MA/London England:MIT Press.

Simon, H.A. 1991 "Organizations and markets" in: *Journal of Economic Perspectives* 5(2) (Spring), pp. 25-44.

Smith, A. 1976 [1766] *Lectures on Jurisprudence* ed. R.L. Meek, D.D. Raphael, P.G. Stein, Oxford:Clarendon Press.

Smith, A. 1976 [1790] *The theory of moral sentiments* ed. D.D. Raphael, A.L. Macfie, Oxford:Clarendon Press.

Smith, A. 1979 [1991] *An Inquiry into the Nature and Causes of the Wealth of Nations* ed. R.H. Campbell A.S. Skinner, Oxford:Clarendon Press (2 vols).

Stanfield, J.R. "Institutions and the significance of relative prices" in: *Journal of Economic Issues* XXIX(2) (June), pp. 459-466.

Stigler, G.J. 1971 "Smith's travel on the ship of state" in: *History of political economy* vol. 3(2), pp. 265-77.

Titmuss, R.M. *History of the second world war; problems of social policy* London: HMSO/Longmans.

Vanberg, V. 1988 "rules and choice in economics and sociology" in: *Jahrbuch für neue politische Ökonomie* 7, pp. 146-67.

Williams, B. 1988 "Formal structures and social reality" in: Gambetta, D. (ed) *Trust; making and breaking cooperative relations* pp. 31-48.

Williamson, O.E. 1975 *Markets and hierarchies; analysis and antitrust implications: a study in the economics of internal organization* New York: The Free Press.

Williamson, O.E. [1985] 1987 *The economic institutions of capitalism; firms, markets, relational contracting* The free Press, New York.

Williamson O.E. 1989 "The economics of governance: framework and implications" in: Langlois, R.N. (ed.) *Economics as a process: essays in the new institutional economics* Cambridge: Cambridge University Press.

Williamson, O.E. 1993 *The economic analysis of institutions and organizations -in general and with respect to country studies* Paris:OECD ED-Working Paper n° 133.

Williamson, O.E. 1994 "Transaction cost economics and organization theory" in: Smelser, N.J. Swedberg, R. (eds) *The handbook of economic sociology* Princeton U.P., Rusell Sage Foundation, Princeton/New York, pp. 77-107

Yezer, A.M. Goldfarb, R.S. Poppen, P.J. 1996 "Does studying economics discourage cooperation ? Watch what we do, not what we say or how we play" in: *Journal of Economic Perspectives* 10(1) (Winter), pp. 177-186.

Young, J.T. 1992 "Natural morality and the ideal impartial spectator in Adam Smith" in: *International journal of social economics* 19:10-12, pp. 71-82