

## **Maria Bianchi, How to Learn Sociality: True and False Solutions to Mandeville's Problem**

Marina Bianchi in her article on Nicholas Barbon:

*"Mandeville repeated Barbon's theory, even more strongly and provocatively, but thereafter it was lost, first in renewed emphasis on saving and productivity, then in the primacy of place given to static allocative efficiency. It can be recovered only with difficulty, one successful method being to stress the links between pleasure and novelty, on the one side, and novelty and learning, on the other. Isolated eighteenth-century authors did this. Modern behavioural psychology has also taken up the theme. But for economists to enter novelty into the utility function' may seem to subvert an act for them to contemplate seriously. That resistance may in part explain why this theme in Barbon (and in Mandeville) has been so greatly neglected."*  
(Marina Bianchi, 'The infinity of human desires and the advantages of trade: Nicholas Barbon and the wants of mind', in Peter Groenewegen (ed.), *Physicians and political economy* (2001), p. 60.)

### **How to Learn Sociality: True and False Solutions to Mandeville's Problem**

Source: History of Political Economy 25 (2): 209-40.

What do we know about cooperation? Or of the coordinating role of institutions, such as market, money, and competition? Mainstream economic theory has encouraged us to take institutions as given and to think in terms of preordered outcomes, such as market-clearing prices. The process of coordinating different interests remains unexplained in this tradition. The question of how cooperative norms emerge among uncooperative individuals has a long history, however, and a growing recent literature on institutions stresses the importance of explaining them as the endogenous and unplanned result of the separate pursuit of individual interests. The idea of institutions as spontaneous, unintended order may be traced to Bernard Mandeville, and the idea of economic coordination as a complex process of acquiring knowledge through trial and error belongs to Friedrich Hayek. More recently analysts have tried to formalize these insights in a game-theoretic approach.

Hayek does more to pinpoint the problem of arriving at social order than he does to solve it, while in the game-theoretic approach the structure of games does not allow any learning beyond the point where the cooperative solution has been identified through a modification in the rules of the game. If the basis for the original differences between agents still remains, however, there will be incentives to learn new ways of breaching any tentative cooperative solution, to one player's own advantage. This is especially to be expected in institutions such as competitive markets, where the whole point is to create and exploit differential advantages. This was clearly seen by Mandeville, though his insight has been obscured by a tendency to privilege the orderly outcomes supposedly produced by the invisible hand. In this connection it is interesting that modern treatments of "competition" invariably refer to Smith (not Mandeville) but that recent historical research stresses *balance*, not *creative change*, as the hallmark of late eighteenth-century thinking (see, for example, Wise 1989). This paper examines the various facets and traps of invisible hand explanations and offers a framework, using game-theoretic ideas, within which the fullness of Mandeville's intuition can be understood. Two points in particular will be emphasized. The first is the link which Mandeville establishes between individual passions and the institutional framework; the second is that this link is at the basis of his original theory of the evolution of social norms and institutions.

I should note before proceeding that I will be offering an interpretation, illustrating the main points with references to Mandeville, not undertaking a complete analysis of Mandeville's writings. The textual basis for this paper is the two-volume Kaye edition of *The Fable of the Bees* ([1924] 1988) containing, apart from "The Grumbling Hive" [1705]

with its associated Remarks and "An Inquiry into the Origin of Moral Virtue" [both 1714], the "Essay on Charity-Schools" and "A Search into the Nature of Society" [both 1723], as well as his six Dialogues [1728]. References will be given in the form of a short title followed by 1 or 2 (for vol. 1 or 2 of Kaye) and a page number.

### **1. Mandeville: Private Vices, Public Virtues**

Many questions lie behind Mandeville's parable of the grumbling hive. What are the conditions for social order and cohesion? What are the behavioural rules, habits, and moral codes that make for social richness and growth? What is the institutional structure of an opulent society?

Mandeville's answer is surgically sharp. For the individuals in society, social benefits do not rely on their "virtuous" commitment to the common good but on their privately motivated desires. In Mandeville's strong terms, private vices enable nations to live in splendour; virtues alone bring them to poverty. As the moral of the fable concludes:

Then leave Complaints: Fools only strive  
To make a Great an Honest Hive  
T' enjoy the World's Conveniences,  
Be fam'd in War, yet live in Ease,  
Without great Vices, is a vain  
*Eutopia* seated in the Brain.  
Fraud, Luxury and Pride must live,  
While we the Benefits receive.

("Grumbling Hive," 1 :36)

But what are the social conditions which allow this paradox to arise and to challenge the more common sense view, then as now, that common good must be commonly desired and founded?

Through the fable of the bees that live in luxury and ease in the fruitful hive, Mandeville compares two different images of society. One is the image of a small, pacific, frugal society, closed to external exchanges, with low consumption and no money. The other is the image of a large, commercial and military society, based on mutual exchange of goods and services.

Mandeville's conviction is that there exists an unavoidable link between the dimensions of a small, closed community and the possibility, on the part of agents, to have direct control over it. Only in this society, in Mandeville's view, can the social links among its members be based on their direct and voluntary agreement and cooperation. In this limited and easily controllable society private moral principles may become also the conditions for public benefits: benevolence, altruism, frugality, and temperance constitute the social amalgam of a homogeneous and compact community. At the same time, for Mandeville, this golden age society, small and virtuous, is closed to development and prosperity, as well as to sciences and political supremacy.

In the second kind of society, the large and commercial one, the social links are not based on a moral sense of community among its members but exclusively on their private drives and self-interested behaviour. In this society the social coordination of the different individual desires does not depend upon the control of anyone of its members. There is no direct connection between private virtues and social benefits. On the contrary, morally unacceptable actions may benefit society, while private morality and individual virtues may be dangerous for social welfare.

But how can a self-interested, unsocial collection of individuals become a "society"? What is the process through which harmful inclinations and desires become the original drives of social cohesion? For Mandeville the answer is to be found in the development, by slow degrees, of a specific framework of socially created norms and institutions: exchange, free trade, and enterprise, combined with a system of laws and government, represent that complex system of social interrelations which connect and "socialize" self-motivated individuals.<sup>1</sup> On one hand, the division of land and the expansion of trades and handicrafts and of navigation and commerce are for Mandeville the social processes that, by touching human passions, stimulate unending innovations and profitable changes (Remark [Q], 1:184). Through these processes individual drives find themselves operating socially and beneficially: fraud and deception stimulate creative ways to increase profits; self-love, pride, and luxury excite the desire for new wants and the

search for new ways of satisfying them. Avarice, envy, and greed elicit accumulation and growth as well as emulation and the diffusion of innovations. Art and science, rooted in and stimulated by these selfish inclinations, enlarge and nourish the process.<sup>2</sup>

On the other hand, the set of institutions represented by markets, trade, and private property is combined with a legal and political framework whose task is to shape and discipline individual drives and help convert them into a cohesive behaviour. "So Vice is beneficial found, When it's by Justice lopt and bound" ("Grumbling Hive," 1 :3637). The system of law and rules is essential to contrive sociality and structure from individually unsocial behaviour: "If by Society we only mean a Number of People, that without Rule or Governement should keep together out of natural Affection to their Species or Love of Company, as a Herd of Cows or a Flock of Sheep, then there is not in the World a more unfit Creature for Society than Man" ("A Search," I :347). Indeed man lives in a "Body Politick," and it is laws and letters that "make him fond of Society" (Sixth Dialogue, 2:3(0)). As well, private vices may be turned into social benefits by "the dextrous Management of a skilful Politician" ("A Search," 1:169; see also "An Inquiry," 1:51).

Mandeville's insistence on the role of wise government and law has led some readers to think of him as a defender of interventionism in economics (see Viner 1958), in contrast with the more common interpretation of Mandeville as a forerunner of laissez-faire. But this set of opposites is itself inappropriate. As Rosenberg (1963, 191) correctly points out, Mandeville sees institutions along with laws as evolving slowly and experimentally. Governments, arts, and sciences, as well as the rules of morality and language, are "the joint Labour of Many Ages" (Dialogues, 2: 128, 186-87,236, 287). What we ascribe to "the Excellency of Man's Genius ... is in Reality owing to the length of time and the Experience of many Generations" (Third Dialogue, 2: 142).<sup>3</sup>

Mandeville's solution to the problem of social coordination among self-motivated individuals relies therefore on the discovery of the role of the institutional complex represented by the market, extensive division of labour, and competition in "utilizing" for social ends private and uncoordinated drives. Between social institutions and individual desires a kind of feedback loop is established by Mandeville. Institutions, while shaping and channelling human passions, stimulate and enlarge them into a cohesive community. Individual passions, for their part, are the inner motor that induces and (unwittingly) promotes those social orderly rules.

In Mandeville's discussion of the evolution of language, for example, he stresses that we use language mainly "to persuade" and "triumph over" others (Sixth Dialogue, 2:289, 291, 293). This is an expression of a private passion and potentially conflictual. In other words, communication, an aspect of sociality, is generated by private vices. A similar argument applies in the case of good manners. One does not yawn or stretch in mixed company (287); but why not? Self-esteem, a vice, dictates such conventions. In politeness, as in dressing, "It is Our Fondness of that Self . . . that could first make us think of embellishing our Persons" (304, 306).

This view of the emergence and role of social institutions is surely new. It is deeply different from the traditional answers to the problem of the constitution of societal rules. The theory of "social contract," in which the wise statesman directly creates the rules of rational government, leaves unexplained how the complex process which coordinates separate individuals can be discovered and reproduced. On the other hand, the "utilitarian" solution in which private interests translate in an orderly way into social welfare simply assumes away any possible discrepancy between individual and social needs and drives. Both these solutions refer to the large society and its social rules as if they had the features of the small and controllable group, as if the rules and goals which belong to one belong to the other as well. In contractarianism, the social authority can combine the dispersed and differentiated individuals of the large society as if the rules of coordination were as easily knowable as in small groups; utilitarianism attributes to individuals a homogeneity and uniformity of interests, as if they belonged to that small society that had shaped and formed their interests and goals in sameness and cohesion.

The rest of this essay elaborates on the newness of Mandeville's vision. To give a brief overview, Mandeville's "precocious feeling for evolution", and for the historical process of growth (Kaye [1924] 1988, I :lxiv-lxv), has been strongly emphasized by Hayek, who recognizes in Mandeville the first discoverer of the tradition that, through Hume, Ferguson, and Smith, suggests an "invisible hand explanation" of institutions (as

more recently these kinds of explanation have been called; see Ullmann-Margalit 1977; Vanberg 1986; Vanberg 1989). Hayek closely follows Mandeville's idea of social order and welfare as the unintended result of human efforts and desires and opposes it to the classical equilibrium solution of economic theory. In section 2 I take up this theme.

Evolution is often invoked to support the idea of an automatic and spontaneous selection of better forms of institutions. In social situations involving either the existence of a common interest or the existence of conflict, however, evolution does not necessarily guarantee the selection of the unique and efficient social outcome (Sugden 1989). New rules which reinforce or replace the existing ones thus need to be created. Mandeville's insistence on the formation of institutions such as the legal system which binds and bends individual drives seems to be in line with this latter emphasis. This will be the topic of sections 3 and 4.

Even if bent and socially shaped, individual passions and drives do not disappear. They remain the inner source of human ingenuity. Mandeville's main claim to novelty lies exactly in answering the question, How can these new changing rules be created? Mandeville's suggestion is simple: new rules arise as men find new ways of breaking or of getting around the old rules when these become an obstacle. Defection may be the general name for this "selfish" ability to alter or circumvent the established set of institutions when they clash with individual interests. Mandeville's view of market competition, as well as of the changing set of laws, as tending toward better social norms rests upon individuals' ability to innovate and to exploit or to create comparative individual gains. The final sections, which compare different conceptions of competition, explain this further.

In short, Mandeville views sociality as emerging mainly from two things: the presence of obstacles and human ingenuity. The nature of *obstacles* is either natural, such as the original danger from beasts, or social, such as the threats which men pose to each other (see the steps toward society as narrated in Dialogues 2:230-31 and 266-67). Human *ingenuity*, the ability innovatively to overcome obstacles, is rooted in and nursed by private passions (vices). Mandeville's solution implies that rules evolve and are created because individual drives remain in a sense unsocial; that is, they are not bound to obey a predesigned plan but are able to create differentiation and novelty. This is what gives substance to Mandeville's view of unsocial sociableness: society arises not despite but *because* of individual conflicting interests. This aspect of Mandeville's analysis, which alone justifies his use of the paradox of individual vices as the origin of social virtues, has been generally overlooked or misconstrued. Interpreters have always dealt in the dichotomies laissez-faire versus interventionism (Viner) or moral versus immoral societies (Smith), thus causing us to lose sight of Mandeville's main contribution as a theorist of institutional change.

## **2. Competition as a Discovery Procedure**

Hayek was the first to utilize Mandeville's idea that sociality is based on individual drives. The shaping of the question in recent discussion owes much to him. Hayek emphasizes the spontaneous and unplanned order of social institutions but gives this more precise form as a process of social learning and discovery (Hayek 1968, 253, 260; Hayek 1978, 269). The process that in Mandeville transforms private vices and inclinations into socially beneficial outcomes is, in Hayek, specified as a process of acquiring and transmitting knowledge through trial and error. In the complex structure of the large society depicted by Mandeville, knowledge exists only individually, dispersed in specific and particular ways. Institutions, such as market, competition, and money, emerge as the informational devices which can help to overcome the limits of individual knowledge. Through these institutions people can learn how to coordinate and combine their different and private information: "Wherever the use of competition can be rationally justified, it is on the ground that we do not know in advance the facts that determine the actions of competitors" (Hayek 1968, 179; see also Hayek 1937, 54).

This view of economic coordination is, for Hayek, something completely different from traditional economic accounts of competitive equilibrium which simply assume the problem of coordination to be already solved. The role of market institutions like competition and the price system, Hayek maintains (1949, 94), is to activate that process through which the "data" on which individuals base their plans are detected, changed, and adapted to each other. The role of prices, in fact, is that of signalling where

changes occur and where to look for better ways to adapt to the continuous changes in the data. These signs alone guide entrepreneurs, for example, to satisfy the wants of a distant and unknown consumer (Hayek 1988, 100).

The concept of spontaneous, extended order that Hayek later introduced and repeatedly elaborated is an attempt to give expression to this process of coordination. The spontaneous order of the market is for Hayek only one expression of a more general principle of evolutionary order, examples of which can be found in physical, biological, and social structures. A common feature of these different kinds of spontaneous order is that they evolve through an experimental process which *selects* better fitted rules and forms of adaptation. A feature specific to the social order, however, is that these selected rules are learned and culturally transmitted and that they dominate the innate rules of genetic evolution. It is through this process of social evolution that institutions have developed, those less effective in reconciling divergent interests being eliminated along the way (Hayek 1967, 100).

Although less stringent than the concept of equilibrium, Hayek's concept of social order is also more vague, and the process of selection and emergence of better forms of institution and how they are learned and transmitted is but sketched by him.

Several specific problems are connected with a simple (Hayekian) evolutionary approach to institution formation, as is now frequently stressed (see Vanberg 1989 and Rutherford 1989).<sup>4</sup> First, it is not certain that the process of evolution will systematically select the efficient set of institutions. As Mandeville notes, we may observe order and supreme wisdom in all the works of nature, "but she is not a Machine, of which every Part more visibly answers the End, for which the whole was form'd" (Fifth Dialogue, 2:233). Current institutions therefore may, for example, preclude the creation of more efficient competitors (see Rowe 1989). For an institution to spread and prevail, there may be required a critical mass of adherents that is not guaranteed to exist (see Witt 1989 and recent contributions to evolutionary games, such as Maynard Smith 1982 and Sugden 1989). Second, the selected set of institutions may be efficient but not stable. For example, price signals of scarcities, which Hayek refers to as a means of coordination, may not fulfil this role if the temptation to mutual cheating is strong enough. This suggests that the evolution of norms may well be spontaneous but is not completely independent of individual planned action.

In addition to these problems, if the coordination solution is not efficiently guaranteed, because conflictual private interests may prevail, it is very difficult, on Hayek's analysis, to understand what rules of action and selective processes are activated. We require a more specifically articulated analysis of the problem situation to which spontaneous order and social coordination are the unplanned solution. As will soon be clearer, many of Mandeville's suggestions become useful here.

Within the framework of game theory some of these problems have recently been structured anew. Utilizing this new perspective, it is possible to explore what set of institutional rules and relationships are likely to evolve from a repeated situation or problem of social interaction described as a game. The literature on this topic is vast and increasing, and I refer to a wide range of contributions in the course of the following discussion.

### **3. Repetitive Games and Institutions**

If we represent social situations of strategic interaction as games, the most interesting games are not those which are played once and only once between a given set of players, but rather, games that are iterated over time (supergames).

If we confine ourselves to the analysis of noncooperative games, where no communication and binding agreements are possible among agents,<sup>5</sup> we have what looks like a sensible and challenging framework to try to explain those institutional forms which emerge without any agent or group of agents consciously designing them. This endogenous approach to the emergence of institutions sees institutions not so much as part of the rules of the game but as modes of behaviour that are part of the solutions to iterated social problems (see Schotter 1981, 28; Schotter 1985).<sup>6</sup> What kind of recurrent social situations are these? Following Ullmann-Margalit (1977), we can distinguish three types of highly stylized, recurring noncooperative problem situations: problems of coordination, problems of Prisoner's Dilemma (PD) type, and problems related to some inequality-preserving rule. All these problems describe situations of interdependent

decisions, that is, games of strategy (as opposed to games of pure chance or skill). They involve two or more players facing different courses of action (strategies) whose outcomes mutually depend on the strategies chosen by the other players. The best course of action for each player, therefore, depends upon what he or she expects the other players will do, knowing that the other players also are trying to anticipate what he or she will do (see Ullmann-Margalit 1977, 78). The differentiating characteristic of this kind of situation is that conventions and norms can be shown to emerge endogenously without invoking any explicit contract or agreement. There will remain, however, as we will see, important differences among the solutions.

### **Coordination Problems**

In line with the early analyses of Schelling (1960) and Lewis (1969),<sup>7</sup> coordination problems describe situations in which the interests of the agents coincide but not in an a priori obvious way. A common example is the case of two persons getting lost in a crowded place or in an unknown city: where will they meet again? In this case each person has to choose according to his or her expectations or what the other will do, and the success or failure of one is the success or failure of both.<sup>8</sup>

When problem situations of this type reappear systematically, we may reasonably suppose that some kind of norm or convention will become prominent (or salient, to use Lewis's term), learned and tacitly transmitted to other players in order to coordinate their choices and to avoid nonequilibrium outcomes in the future. Norms, in this sense, can be thought of as the "noncooperative," unplanned, equilibrium solution to some recurrent coordination problem.<sup>9</sup> The characteristic of this norm is its stability; once learned and repeated, no one has any incentive to violate it.

Repeated games of the coordination type thus seem very well fitted to represent the spontaneous emergence of norms. The evolution of media of exchange from direct barter, the creation of money through the banking system, the emergence of the minimal state depicted by Nozick (1974) as an enforcement system of individual rights (though this last example belongs more to the type of cooperative games) are all good economic examples of invisible hand processes (as shown in Ullmann-Margalit 1978). They seem able also to capture Mandeville's (and Hayek's) intuition of a social, beneficial order emerging from the individual pursuit of private interests rather than from voluntary agreement or by decree. An example of this tentative and experimental process is given by Mandeville's creative account of the emergence of language.<sup>10</sup>

Two caveats are necessary, however. One is that the conventions and norms thus discovered and selected are not necessarily the most efficient, even if they are surely stable. Think, for instance, of the U.S. mixed system of weights and measures in a largely metric world and the popular resistance to moving to a wholly metric system. (For other examples of coordination norms which do not improve collective welfare or, conversely, which do improve it, but at the expense of individual utility, see Elster 1989.) The second caveat is that if a less than efficient solution prevails, the shift to a more efficient one cannot happen spontaneously because of this very stability. Change, if it is to occur, then requires an explicit agreement or decision.

This brief discussion on coordination games is in line with Mandeville's view of sociality as a slow and complex process of experience, adaptation, and learning. For coordination to come about seems to require that the strategy space of individuals must incorporate some search or experimenting activity, some trial and error procedure through which rules can emerge, be imitated, and diffuse themselves. Moreover, as just noted and as Mandeville himself suggests, this learning process involves some active intervention when a shift to more efficient rules is required. A shift is not automatically guaranteed by the simple evolutionary process. Rosenberg correctly stresses that society, in Mandeville, is the intertwined product of historical evolution *and* wise government and laws (Rosenberg 1963, 191). In other words, change in the direction of efficiency involves reinforcing some behaviours and discouraging others. Mandeville notes several times that "the great Business in general of a Politician is to promote, and, if he can, reward all good and useful Actions on the one hand; and on the other, to punish, or at least discourage, everything that is destructive or hurtful to society" (Dialogues, 2:321; see also 300, 309, and "A Search," 1 :347-48, 369). The result of this system of legal restrictions and law enforcement is twofold. Not only are human passions constrained

and bound (as the view of institutions based on models of bounded rationality has led us to think),<sup>11</sup> but laws also expand and enlarge individual capabilities to conceive new options and alternatives. As the vine that, left to itself without trimming and skilled pruning, will yield little fruit, so restrictions and norms, if judiciously applied, yield new opportunities and changes within the flourishing hive (see "Grumbling Hive," 1:36-37).

The discussion on PD games, to which I now turn, reinforces this conclusion and shows us new steps of the learning process. Learning takes place not only in the developing but also in the breaking of cooperative forms of sociality. For Mandeville it is in this vice-driven process - the deliberate breaking of learned norms - that the origin of new and better rules lies.

### PD Problems

PD situations have received widespread attention as a way of modelling social situations of divergent, conflictual interests among the players. Mandeville's paradoxes of "vices" being transformed into social "virtues" appear to rest on situations of the PD sort. Vices *are* vices, for him, because they do not incorporate any intended social good, any original innocence or integrity ("A Search," 1 :345). Similarly defecting would not be the dominant strategy in PD games if the players had the social dimension in mind, if they were "naturally" cooperative. In fact Mandeville's basic question, How can cooperation emerge from this uncooperative behaviour?, may be structured in a PD form. A specific example of the simple case of two players and two strategies, where the strategies are cooperation C and defection D, is:

		Player 2	
		A	B
Player 1	A	2,2	0,3
	B	3,0	1,1

The preference ordering for players 1 and 2 is, respectively,  $DC > CC > DD > CD$  and  $CD > CC > DD > DC$ . With preferences as reflected in these payoffs, the equilibrium solution, as is well known, is DD. PD problems therefore represent situations in which for any equilibrium strategy (DD) with equilibrium payoffs (1,1) there exists at least one nonequilibrium strategy (CC) which is Pareto superior to it (nonequilibrium payoffs 2,2). This type of situation represents a dilemma for the players because each of them, trying to reach their most attractive outcome, DC (for player 1) and CD (for player 2), or to avoid the less attractive CD and DC outcomes, ends up with the mutually harmful result DD. This result could have been avoided had they cooperatively chosen CC. (For further discussion of the existence and nature of the PD paradox, see Campbell 1985.)

The way cooperation is thought to emerge from this dilemma involves essentially the idea of iterating the game. If these situations are repeated, players may learn the type of behaviour they can expect from each other and build up a set of norms of behaviour that would avoid the repeated use of equilibrium but nonefficient strategies (see Schotter 1981, 39, 24; Ullmann-Margalit refers to PD situations as situations which "call for" norms [1977, 22]). In this case, as with coordination situations, cooperative rules and regularities of behaviour may be expected to emerge spontaneously. Before analyzing more closely the problems connected with this solution, an immediate, evident difference with the coordination problem solution should be pointed out. In fact, if a rule always to cooperate should prevail in a group, there still remains for the individual members of the group a strong incentive to defect unilaterally. The norm is not self-enforcing as is the case with coordination norms. Some sanctioning rule that reinforces the power of the norm must be introduced (see Axelrod 1986). Recall that in Mandeville it is the task of the legal system to spell out the rules and prescribe the punishments. As noted earlier, and as the results of the present discussion lead us to think is unavoidable, he insists properly we can now say - on the crucial role of imposed rules and punishments.

### Inequality-Preserving Problems

The third kind of problem is represented by those situations connected with some inequality-preserving institutions, or norms of partiality (for an extensive analysis, see Ullmann-Margalit 1977, esp. 134). They represent a rather interesting case, which,

however, I shall not analyze here. Instead, I concentrate on PD games, focusing on learning as a way to avoid becoming stuck with a mutually unwanted outcome.<sup>12</sup>

#### **4. PD Norms and Institutional Change**

The idea to utilize the analytical structure of supergames (for the formulation of which, see Friedman 1977) allows us to describe social institutions as (non-cooperative) equilibriums of games that involve the infinite iteration of some particular original game which represents the constituent game (as originally defined by Friedman 1977). Institutions can therefore emerge as the equilibrium solution of the supergame on (see Schotter 1981, 24, 54). For example, in the two-person PD game we have referred to, only the solution DD is allowed if the game played just one time. If the game is iterated over time, on the other hand, more equilibrium solutions are possible, solutions that, though irrational in the short run, may be optimal in the long run. Take, for example, the case in which the players adopt a "grim trigger strategy". A grim trigger strategy advocates cooperation both to start with and at an ongoing behaviour until some other player defects, in which case defection is practiced thereafter. In this context, if one player defects while the other is cooperating, he or she will receive the DC (>CC payoff once, but thereafter his or her payoff will be DD <CC). With its strong punishing rule, the grim trigger strategy seems to enforce cooperation (unlike the tit-for-tat strategy, where if a player defects once and then plays tit-for-tat, as does his or her opponent, there will be an alternation of DC and CD forever).<sup>13</sup>

This view of institutions as solutions of supergame problems presents several analytical difficulties which have been extensively discussed in the literature. I refer to them only briefly (for a recent clear survey, see Fudenberg and Tirole 1989). First, cooperation can become a dominant strategy only if the repetition of the game is infinite (or if the players do not know the end of the game; for experiments on this type of PD game, see Roth 1988). In the presence of a finite-horizon supergame the equilibrium solution is the same as the one-shot game.<sup>14</sup> Second, in this new infinite horizon setting cooperation, though a possibility, is not uniquely guaranteed. A great variety of equilibrium solutions is now allowed, including mutual defection. In addition, as noted earlier, in the generalized version of PD norms, where numerous players are involved, the problem of free riding arises; that is, it pays the single player unilaterally to violate the norm (for an analysis of the pervasiveness of this problem not limited simply to PD problems, see Thomela 1991).

All these cases imply an ability to select among the multiplicity of equilibria and an enforcement of cooperation when a selection is made.

These conditions may emerge spontaneously. The "discipline of continuous dealings," as Adam Smith calls the emergence of trust and reciprocity, represents those tacit behavioural rules which reinforce cooperation and without which a market-exchange system would be almost unimaginable.<sup>15</sup>

It is this sort of spontaneous mechanism that Mandeville refers to when he notes that it would be impossible for men to endure social restraints if they "could not be taught to play the Passion against itself" and to exchange the natural symptoms of the passions for "other Symptoms, equally evident with the first, but less offensive, and more beneficial to others" (Third Dialogue, 2: 125-26). Thus pride is turned into honour and fear of shame (125; Remark [C], 1:63-64), and force and violence are turned into politeness and good manners (Sixth Dialogue 2:291, 295). This spontaneous process of social learning is never irreversible, Mandeville always warns. He stresses how social and moral codes, not being founded upon any real principle of virtue or religion but being the result of art and education, can easily be turned into vices again: "The same Fear of Shame, that makes Men sometimes appear so highly virtuous, may at others oblige them to commit the most heinous Crimes" (Third Dialogue, 2: 124; "A Search," 1 :343). In this case, the conditions for stabilizing the cooperative strategy must be articulated and codified in a system of laws, contracts, or mutual agreement.<sup>16</sup> The result of these new incentives is to change the very structure of the game and transform it into a cooperation game. This might happen by rendering defection very costly and/or by rewarding cooperation. The payoff structure is therefore changed in such a way that unilateral defection no longer dominates cooperation.

Mandeville saw this point so clearly that it is worthwhile to refer to it directly. Whilst narrating the various steps of human civilization, Mandeville entertains the



possibility that voluntary cooperation will emerge (Sixth Dialogue, 2:267-70):

Hor[ace]: But would not their Sufferings in time bring them acquainted with the Causes of their Disagreement: and would not that Knowledge put them upon making of Contracts, not to injure one another?

Cleo[menes]: Very probably they would; but among such ill-bred and uncultivated People, no Man would keep a Contract longer than that interest lasted, which made him to submit to it.

Nor will religion be enough, nor ambition. Hence the people invent penalties and prohibitions. But this last resource is not enough until the invention of letters, the third and last step to civilization. No multitudes can in fact live without government; no government can live without laws; and "no Laws can be effectual long, unless they are wrote down" (269).<sup>17</sup>

It is only with the invention of letters and the improvement of laws that the superiority of human understanding becomes a contribution to sociableness: "Whereas the same Endowment before that time, the same Superiority of Understanding in the State of Nature, can only serve to render Man incurably averse to Society" (300).

If rule emergence is the result of a process of learning and selection, how does this process really work? Within a game theoretic approach, how learning proceeds has not been successfully addressed. For example, a punishing strategy (like a grim trigger strategy) leads us to expect that cooperation will emerge and prevail but still leaves unexplained how the strategy itself is selected and learned.

The process of learning involves an ability for actors to engage in trial and error exploration, but errors are in fact forbidden in a game theoretic representation of interacting players. The notion of perfect equilibrium, and subgame perfectness, first introduced by Selten (1975) in order to rule out non-credible threats, when applied to repeated games, excludes nonequilibrium strategies in all the sequential moves (the path) of the game. In our case, if the players start from an equilibrium outcome like DD and observe an out-of-equilibrium move like cooperation, they can only interpret it as it is, a mistake. As a result they will not change their strategy but will continue to defect (on the impossibility for the players, given the rules of the game, to recognize and understand an out-of-equilibrium move, see, in particular, Binmore 1987). If this is true, however, even in repeated games, strategic moves are independent and unaffected by each other, with the result that any real learning is precluded and cooperation can never appear.

Therefore, unless the players are allowed to use the game as a discovery procedure where they gather information from other players' past moves to understand and possibly influence their own future moves, we cannot expect a strategy like a punishing strategy or a reputation-building strategy to emerge. For this to happen the game must be enlarged to make available a wider space of strategic options and possible solutions to the players.<sup>18</sup>

In our case this means that as long as the game is repeated, the players, learning from their errors (the accumulation of losses resulting from DD), develop a new set of rules involving cooperation which, though out-of-equilibrium, may become a new equilibrium once signaled, recognized, and adopted. A deviant move in this case is not simply a mistake but is more like an experiment made by the players (as suggested by Sugden in the case of coordination games [1989, 92]), the success or failure of which will depend on the ability of the players to understand and influence each other's moves. In this way only are players supposed to act strategically rather than pragmatically. By introducing this distinction, Buchanan correctly stresses that the objective of strategic action is precisely to influence the behaviour of the other in such a way as to produce the preferred outcome (a strategy which will normally violate the simple utility maximization rule of a nonstrategic setting [1976, 75]).

In this different framework the resemblance to games of coordination is instructive. There the ability of agents to discover and to learn from each other's past moves was due to the ambiguity or the multiplicity of the possible final outcomes. In PD (super)games, this ambiguity and therefore opportunity is recreated if the players are provided with an ability to introduce new moves, in this case to recognize and select a

cooperative strategy of whatever sort.<sup>19</sup>

Conjecturing and discovering the possible strategies and outcomes thus necessarily becomes a learning experience.<sup>20</sup>

By contrast, introducing a strategy, such as a grim trigger strategy, that under certain conditions is able to generate an equilibrium solution based on cooperation simply implies that the rules of the game have been changed exogenously. No social learning activity is implied which might justify the evolutionary emergence of equilibrium (on the limits of the Bayesian learning rule such as Schotter's, which implies one single-exit outcome, see Mirowski 1988).

On the other hand, through learning - that is, through searching cooperatively for a solution to a problem which remains unsolved within the structure of the original game - an endogenous process of rule changes and new strategies is activated. In this way many options are opened up. We have explored DD and CC, but why should not CD or DC be taken up by participants?

## 5. The Emergence of Competitive Innovation

The problem dogging the super-game structure of PD norms is how to avoid the stable equilibrium solution, DD, of the constituent game. I have suggested as a way around this the activation of a learning procedure. Through the constant replication of the mutual losses as compared to their possible mutual gains, people might learn to cooperate, to develop tacitly or explicitly new sets of rules. To stabilize this solution might mean to change the outcomes associated with the various strategies.

Once a learning procedure by agents is recognized to exist, however, and it is accepted that it might possibly change the *structure* of the game, our interest in the process of rule formation does not cease. The whole strategic structure of PD games is relevant for the representation of the players' moves. For a second, though not alternative, way of escaping the DD outcome is represented by the attempts by agents to stabilize their most attractive outcome, the unilateral defecting strategy, while the other continues to cooperate (CD or DC). How can this result be reached? Cheating, in a world of trust, is a source of comparative advantage for the cheater. Cheating can be detected and punished. This does not preclude, however, that *new* forms of cheating can become a temporary source of comparative gain, or even not so temporary, as where forms of invisible cheating are practiced such as misrepresenting one's preferences (see Hurwicz 1973; Roberts and Postlewaite 1976). Mandeville saw this very clearly, and he, in effect, made fraud an essential ingredient of flourishing trade: "All Trades and Places knew some Cheat, No Calling was without Deceit" ("Grumbling Hive," 1 :20). So multiform was deceit that Mandeville despaired of being able to list the specific ways in which it infected practice: "But who can all their Frauds repeat?" (23). This he accepts fully as the inevitable condition of a rich society: "Fraud, Luxury, and Pride *must* live, While we the benefits receive" (36, emphasis added; see also 1 :61, 185). A typical example of fraud as failure to disclose full information about costs and quality to the buyers is the use by sellers of private marks. For Mandeville this is "a certain Sign that ... [they] are equally careful in concealing the prime Cost of what they sell" (Remark [E], 1 :81, n. O. Similar are Mandeville's discussion of "fair trade," which in fact involves the parties deliberately exploiting asymmetries of information,<sup>21</sup> and his entertaining description of the supreme art of prudent dissimulation and hypocrisy displayed by the retailer toward the young lady, his customer ("A Search," 1:349).

Fraud in Mandeville may be seen as the generic name for all those activities which imply some defection or, more generally, finding new ways to escape the limits and constraints of existing rules. Precisely this ability to circumvent the established rules represents for Mandeville the way change is activated and new rules are developed. In Mandeville's epigraphic words: "vice *nurs'd* ingenuity" ("Grumbling Hive," 1 :26; emphasis added). Nor is this ability confined to trade. Rules of government, too, and laws, follow the same procedure. "It is not Genius, so much as Experience, that helps Men to good Laws .... The wisest laws of human Invention are generally owing to the Evasions of bad Men, whose Cunning had eluded the Force of former Ordinances, that had been made with less Caution" (Sixth Dialogue, 2:319).

The process seems to be that rules and restrictions, while solving some cooperation problems, also stimulate creative ways around them. Following Mandeville, Rosenberg (1963, 190) has stressed that the task of discovering the appropriate laws (rules of the

game) is enormous because men's vices will suggest innumerable ways to profit at others' expense. At the same time, for Mandeville this is the only way for a fruitful hive to expand. For "the Sociableness of Man arises only from these Two things, viz. The multiplicity of his Desires, and the continual Opposition he meets with in his Endeavors to gratify them" ("A Search," 1 :344).<sup>22</sup> Learning here takes the particular form of *breaking* constraints. In Mandeville's flourishing hive, fickleness in matters of diet, furniture, and dress was an aid to industry, but laws too "were equally Objects of Mutability":

For, what was well done for a time,  
in half a Year became a Crime;  
Yet while they alter'd thus their Laws,  
Still finding and correcting Flaws,  
They mended by Inconstancy  
Faults, which no Prudence could foresee.

("Grumbling Hive," I :25; see also Fourth Dialogue, 2: 187)

Fraud as a name for norm breaking also embraces all those activities which characterize competition and its main motor, strategic innovation. It is to this admittedly unconventional way of viewing competition that I turn in the next and final section. As we shall see, competition as innovative behaviour seems to have been Mandeville's perspective too.

Cheating, in fact, and new forms of cheating can be easily imitated, in which case the PD game solution degenerates to a DD outcome, or the cheating is detected and punished by new sets of rules. The development of new cheating devices, though always a temptation, cannot be the permanent, systematic method for stabilizing CD. Matters are different, however, if the agents learn to innovate. Innovation, the introduction of a new move not yet known by the other player, generates a comparative advantage for the innovating player. Market competition represents exactly this kind of behaviour. The introduction of new products, new technologies, and new quality differentiation, and the discovery of new markets and new forms of organization (as first made explicit by Schumpeter) are all unilateral "defecting" moves which allow for a comparative gain. Again, Mandeville's examples of multiplication of needs and new ways of satisfying them are instances of precisely this sort: competition as a process of discovering new ways of creating comparative advantage.

Unlike simple forms of cheating, competitive innovation is less easy to imitate because of the costs of research and discovery, and it thus carries distinct advantages to the first mover. That is why competitive innovativeness represents precisely that systematic institutional form which emerges as an attempt to stabilize CD. From different forms of strategic advantages different forms of enforcing rules have emerged: for example, the rule to punish "bad" competition (cheating, fraud) and the rule to protect "good" competition through the creation of patents. Competitive advantage of course is not eternal: imitation and diffusion eventually occur. They represent forms of sharing and equalizing the initial one-sided gains. But this does not mean the end of competition.<sup>23</sup> New forms of behaviour are always likely to emerge from this process of learning and differentiating activity. The vast literature on innovation, learning, and diffusion, as well as the new game theoretic literature within industrial organization, offers a large variety of examples of this kind of strategic behaviour.

In this different perspective competition represents that institutional form which evolved to temporarily stabilize a situation which is not a game theoretical equilibrium but which is still the most attractive for the one who gains an initial advantage. Unless the payoffs of the game are changed through this process (thus transforming the game into a different one, for example, a game involving an inequality-preserving rule) and unless we imagine a permanent capacity of innovation in the competing firms (both hypotheses not far from some real-world situations), unilateral defection is not stable but gives rise to either mutual competitive losses or collusion.

This way of representing the role of competition as a discovery procedure, to paraphrase Hayek, is very different from the traditional way of addressing the coordination role of competition. Under the hypothesis of negligibility, or of individual insignificance, and the assumption of a central auctioneer, perfectly competitive markets "appear 'rigged' to induce coordination success" (Weintraub 1979, 133). However, within

the framework of cooperative games, a different use of the process of competition has been advanced. Is this different use of competition as a coordination process viable?

## 6. The Two Meanings of Competition

One of the main differences between the theoretical models of competitive equilibrium in cooperative games (characterized by free preplay communication and precommitment among players) and the Walrasian theory of perfect competitive equilibrium is that in the former the equilibrium solution to the game is the result of a process of a multilateral bargaining among players. Firms and consumers are not simply assumed to be price-takers but to have a behavioural flexibility to form groups or coalitions and to contract (make voluntary agreements) among themselves as to how the available gains are to be divided. Perfect markets and the competitive equilibrium solution are not assumed to exist at the outset but emerge endogenously as the trading process is repeated over time.

The game theoretical solution to this problem makes use of the concept of the core which can be shown to be equivalent, in market games (as analyzed by Shapley and Shubik 1969), to Edgeworth's contract curve<sup>24</sup> (the literature on the topic is vast, but some of the central issues may be seen in Arrow and Hahn 1971; Hildenbrand and Kirman 1976). Following an intuition which originally belonged to Edgeworth (as first shown by Shubik 1959), this solution reaches the important result that, as the economy becomes larger, in the limit the only core allocations that remain viable are also competitive equilibria (for a discussion on the non-tatonnement convergence to Pareto allocations, see Weintraub 1985, esp. 153). To give an idea of the process, take a simple example (Stigler 1987, 534, citing Edgeworth). If we start with a market situation with only one seller and two buyers, the seller gains all the benefits of the sale, while each buyer is charged the maximum price he is willing to pay. Now, if a second seller appears, it will be advantageous for him to offer better terms to the two buyers. We can imagine that as the number of traders increases on both sides, prices converge to the competitive equilibrium prices where no trader can influence them. In the limit, the contract curve shrinks to the competitive equilibrium, and the two solutions, the Walrasian and the core, merge (for the measure of difference between the two equilibrium sets, the core and the Walrasian, see the set of limit theorems for large economies or replica economies in Hildenbrand 1974; Hildenbrand 1982).

The crucial element in this explanation of the perfect competitive equilibrium as a cooperative game solution is the assumption of the infinite enlargement of the number of players in the economy. However, this assumption, though important in showing under what conditions the core solution coincides with the Walrasian solution, is exactly what requires explanation. If the purpose of the whole theoretical exercise is the *emergence* of the equilibrium cooperative solution from mutual bargaining, the conditions under which this solution is supposed to prevail have to be explained too (see, on this point, Witt 1985).<sup>25</sup> If we suppose that the number of traders is not yet large enough to make each one a price-taker, is there a process which allows for the number to increase? In other words, is there any process which guarantees the institutional conditions (infinite replica economies) for the perfect competition to emerge? To go back to our example, it is easy to see in that case that if the two buyers join together they are better off than if they compete. But this collusion will destroy the profit opportunity for a potential new seller, who, therefore, will not appear. The same argument applies no matter how large the number of traders. They can still join on one or both sides of the market, thus giving rise to monopoly or to bilateral monopoly rather than to perfect competition. By simply joining together, a dominated strategy (ceasing to compete) is stabilized, and the movement toward perfect competition is prevented (see Schotter and Schwödiauer 1980). Collusion, like cartels and trade unions, may therefore stop the process of recontracting at some equilibrium which is not competitive.<sup>26</sup>

The players can do more, however. Once we allow them to play actively, many alternatives are open to them to prevent new competitors from eroding their monopoly position. All those differentiating activities represented by organizational, reputational, and technological change strategies represent new market barriers as well.<sup>27</sup>

So, as it turns out, the core explanations of perfect competition simply assume the conditions under which a competitive equilibrium may emerge from bargaining.

Recognizing this provides space for the idea of competition as a process which multiplies differences among agents rather than eliminating them through the appeal to a number of players large enough to make each one insignificant.

There is thus a clash between conceptions of competition as a means of unintentionally connecting separate and dispersed individuals. In the traditional, and modern, perfectly competitive equilibrium analysis, competitive behaviour is restricted to optimal responses to fixed parametric prices. No real strategic interaction intervenes. No institution plays any effective role, unless it is the Walrasian auctioneer. But traditional perfect equilibrium is not institutionally free or neutral. The set of assumptions of perfect competition in effect imposes a strict set of rules on players without any attempt to derive or defend these rules. The result is that the players' different interests are harmonized before differences are allowed to manifest themselves.

The role of competition as we have depicted it here is quite different. It invokes creating new ways of making and defending profit opportunities. The central idea can be traced back to Mandeville.

To this Emulation and continual striving to outdo one another it is owing, that after so many various Shiftings and Changings of Modes, in trumping up new ones and renewing of old ones, there is still *a plus ultra* left for the ingenious; it is this, or at least the consequence of it, that sets the Poor to Work, adds Spur to Industry, and encourages the skilful Artificer to search after further Improvements. (Remark [M], I: 130)

The implication of this is startling: coordination relies on *differences* of interests rather than on their (pre)reconciliation. That has been largely lost sight of, although it was the very basis of Mandeville's view of sociality.

### Conclusions

Mandeville's answer to the problem of how societies can be orderly in the presence of multiple and conflictual interests, hence in conditions of potential chaos, has always seemed paradoxical. As I have represented him, he suggests that order can arise not despite but *because* of the presence of conflictual individual interests. We have inherited and learned to live with a weaker and more palatable version, the idea that the simple pursuit of private interest may lead unintentionally to socially beneficial outcomes. This proposition characterizes what recently have come to be called invisible hand explanations. In contrast to some alternatives this weaker form is startling enough. It is clearly opposed to the Hobbesian Leviathan, or central authority, which simply supplants the potential for conflict.

My own interpretation owes much to Hayek and to recent developments within the game theoretic literature. From Hayek we have learned to think in evolutionary terms. Game theory has posed questions solutions to which require that we learn to think in terms of experimenting with new strategies and altering rules.

However, something of the paradoxical aspect of Mandeville's answer is lost even in these developments. Hayek does not explain what *socially beneficial* means nor how such beneficial institutions are reached, maintained, or changed. In the game theoretic approach, on the other hand, especially if one represents the problem of cooperation purely in terms of the equilibrium solution of recurrent PD situations, as is often done, one is forced to restructure the problem as if interests necessarily converge to the commonly beneficial outcome. We may be led to conclude, falsely, that the only solution to PD problems is a world of cooperation, a world where conflicts are suspended and interests reconciled.

In Mandeville, by contrast, trade and commerce are seen as the institutional amalgam which socially and unintentionally connects private conflictual interests, but by exploiting and even enhancing them. This last element may be overlooked if we stress only the order in the final outcome, which is the perspective all too often imputed to Smith and constantly translated into traditional analyses of competition. Market competition as I have depicted it here differs from both the game theoretic constructions of the sort just mentioned and the pricetaking behaviour of competitive equilibrium modelling. Competition, as innovation, as the creation of differential advantage, however, explains and preserves Mandeville's original insight, which was to

predict order from conflictual differences while maintaining openness to change and flexibility to adjust to what "no Prudence could foresee."

**For the references, not included here, please see the original article in *History of Political Economy* 25 (2), pp. 236-40**

Notes

- 1). Self-love, economic interest, and emulation alone do not permit coordination and growth. They require a set of institutional norms to operate socially. "*Fabricando fabri fimus* [by making makers are made]. Men become social by living together in society," says Mandeville (Fourth Dialogue, 2: 189). For the importance of the institutional setting into which passions are embedded, see Rosenberg (1963) and Goldsmith (1985, 125). For the central role that the market in particular plays in Mandeville, see Schneider (1987, 131).
- 2). For Mandeville all the "greatest Excellencies ... are acquired" (Sixth Dialogue, 2:296). Mandeville never stops stressing the role of education and experience in transforming individual inclinations into socially useful outcomes. "There is no Innocence or Integrity that can protect a Man from a Thousand Mischiefs that surround him. On the contrary every thing is Evil, which Art and Experience have not taught to turn into a Blessing" ("A Search," I :345). But, Mandeville adds, man is unwilling to accept his natural weaknesses and tends to ascribe to Nature all the social ornaments that are the result of education and art (Sixth Dialogue, 2:306).
- 3). "It is the Work of Ages to find out the true Use of Passions, and to raise a Politician, that can make every Frailty of the Members add Strength to the whole Body, and by dextrous Management *turn private Vices into publick Benefits*" (Sixth Dialogue. 2:319).
- 4). Recent exchanges have led to a useful airing of the characteristics of the so-called old institutionalism and to comparisons to the "spontaneous emergence" framework of the new institutionalism. See Rutherford (1989), Langlois (1989), and Vanberg (1989), who specifically compares Menger and Commons and views them as complementary, Menger being concerned with the evolutionary emergence of general rules of conduct, and Commons with the deliberate creation of organizational rules.
- 5). The formal game theoretic distinction between cooperative and noncooperative games is that in the former the rules allow the players to communicate and to form binding contracts whereas in the latter they do not.
- 6). The spontaneous order approach has maintained a somewhat ill-founded position on conscious actions, namely, that their explanation is a task which belongs to psychology. For Hayek as well as for Schotter (1981, 21), in social sciences conscious actions are data. This argument does not consider that deliberate actions may actually be endogenously related to a specific problem to whose resolution they contribute. Examples of this kind will be given in the course of this essay.
- 7). Coordination problems, though they have not received special analytical attention because they do not yield neat and general mathematical results, are nonetheless an interesting framework for the representation of the cognitive performances of agents. See Dupuy 1989, 55.
- 8) The classic representation of a coordination game with two players and two strategies is the following:

		Player 2	
		A	B
Player 1	A	0,0	2,2
	B	2,2	0,0

where the preference ordering for both players is  $BA = AB > AA = BB$ . More generally, we can say that coordination problems describe situations in which there are at least two combinations of choices that all players prefer to the other possible combinations. The specification "at least two combinations" is important because in the presence of only one most preferred outcome no ambiguity or problem of coordination arises, the coincidence of interests being immediate. Another specification is that not all problems of coordination represent pure coordination situations. Some players in fact may care about one coordination outcome more than another. In this case the preference

order might become for player 1  $BA > AB > AA = BB$  and for player 2  $AB > BA > AA = BB$ .

9). A (Nash) equilibrium solution is that combination of strategies for which no one has an incentive to depart from his or her strategy so long as the others do not.

10). "For the Things they were more conversant with they would find out Sounds, to stir up in each other the Idea's of such Things, when they were out of sight: these Sounds they would communicate to their young ones; and the longer they lived together the greater the Variety of Sounds they would invent, as well for Actions as the things themselves" (Sixth Dialogue, 2:288). As stressed by Kaye in the footnote to this passage (288), Mandeville was a pioneer in "his insistence on the non-divine origin of language and its halting and indirected evolution."

11). "Persons that are well educated, have learn'd ... to tye themselves up to Rules and Decorums for their own Advantage. and often to submit to small Inconveniences to avoid greater" (Sixth Dialogue, 2:306).

12). Inequality problems arise as a subclass of coordination problems. As in the case of nonpure coordination, the game has two equilibrium solutions, each favouring one of the two players. If, from the repetition of the game, a solution emerges which sanctions a status quo of inequality, the less favoured party may try to improve his or her position at the expense of the other. In this case, too, some punishing rule supporting the status quo must be envisaged if the tacitly agreed-to rule is to be defended.

13). Tit-for-tat is more forgiving. implying the following strategy: cooperate to start with; in all the subsequent periods do what your opponent did in the previous period.

14). As is known. this happens because it generally pays the players to defect in the last period. This shortens the supergame by one period; the same - defect in the last period then applies back through the first period. thus rendering the defect strategy the only equilibrium strategy (see Thompson and Faith 1981). However. the infinite repetition of the game implies an infinite inflexibility and invariance of the game structure itself. no matter what other circumstances change.

15). For the role of these tacit rules, see MacKean (1976, 31), and see Opp (1982) for the role of changing preferences. Tullock (1985, 100) suggests an interesting different line of reasoning as a justification of the emergence of modes of behaviour such as trust and credibility. In particular, he notes that in free competitive markets partners can change and people can voluntarily choose their own partners. This means that if somebody plays uncooperatively, he or she will find it very difficult to get people to play with him or her in the future. With the large numbers assumption, however, the same reasoning might also imply anonymity and therefore impunity in defecting.

16). As a referee correctly observes: "Mandeville's system relied not just on spontaneous, invisible hand processes, but also on deliberative enforcement of rules by government. ... The social system *as a whole* may not have been planned by anyone but it is the outcome of both spontaneous processes (invisible hand) *and* deliberative processes (wise government) that *interact* over time."

17). The invention of letters, says Mandeville, "is an Incouragement to all other Inventions in general; by preserving the Knowledge of every useful Improvement that is made. When Laws begin to be well known, and the execution of them is facilitated by general Approbation, Multitudes may be kept in tolerable Concord among themselves" (Sixth Dialogue 2:299-300). A rich variety of real-world examples of ways of solving (or failing to solve) recurrent social problems of PO type is given by Ostrom (1989). She shows how PD norms, to be effective, often require the creation of rules additional to those allowed by the strict structure of the game. On the same lines, see Musu (1989) and Carraro and Giavazzi (1989).

18). See Shackle (1972, 426): "The most advanced and spectacular secret of success [in a contest] is novelty, and novelty is that which an infallible algorithm must, by definition, exclude." I owe this quotation to Ian Steedman.

19). We can also put the problem in this way, as Sugden correctly observes: if the program of game theory could be carried out (i.e. • rational analysis uniquely prescribes a dominant strategy), then axioms of rationality would be enough for a theory of individual and social behaviour. Conventions would be redundant (1989, 89). By contrast, conventions arise when there is more than one solution to the game, as in coordination games. What I am suggesting here is in line with this position, even if the new solutions emerge endogenously from a process of learning activated by the social

problem or dilemma depicted by the game. Rationality is therefore enlarged in order to include this learning searching activity.

20). The theoretical consequences of this different framework are made clear by Rowe. Contrary to the attempts made by Kreps, Milgrom, Roberts, and Wilson (1982), for example, who drop the assumption of common knowledge in order to allow cooperation to emerge, Rowe rejects the assumption that the game is decidable, that is, that the theorist as well as the players can deduce a priori how rational players will act in all the following moves of the game (Rowe 1989, 47). For the paradox of rationality in game theory and ways to solve it, see also Bianchi and Moulin 1991.

21). Mandeville offers the following example of the "innumerable Artifices, by which Buyers and Sellers out-wit one another" (Remark [B], 1:61). There are two big traders. The first, a West Indian merchant, having received the information that a much larger supply of sugar than had been expected will reach England, is happy to sell at the price offered by the buyer, though this had at first seemed too low. The buyer, however, independently learning that the Barbados fleet has been destroyed by a storm, quickly accepts the deal. Both of course dissemble about the information they possess. Mandeville points out that "all this is called fair dealing" but that "neither of them would have desired to be done by, as they did to each other" (63). This violation of the "Golden Rule," Goldsmith notes, is for Mandeville typical of all trading (1985, 126).

22). No society could have arisen from the "Loving Qualities" of men, but only from their imperfections and appetites: "We shall find likewise that the more their Pride and Vanity are display'd and all their desires enlarg'd, the more capable they must be of being rais'd into large and vastly numerous Societies. Was the Air always as inoffensive to our naked Bodies ... and Man had not been affected with Pride, Luxury and Hypocrisy .... I cannot see what could have put us upon the Invention of Clothes and Houses" ("A Search," 1:346-47).

23). An arms race may be viewed in this light. Here cooperation emerges when all the possible ways of introducing new strategic advantages have been exploited. The case of obsolete industrial sectors is analogous. In these cases cooperation is a better option than the search for relative advantage. However, cooperation might become a more difficult option if there still remains some faith in the relative advantage of defection.

24. The core of an economy represents those states of the economy which no coalition of agents can improve upon, or block through a process of contracting and recontracting.

25). See the disarming acknowledgment by Mas-Colell (1980) cited by Janssen (1990, 93). Janssen gives a full critical discussion that substantially reinforces the point.

26). An analysis of the symmetric solutions reached via agreements on the division of joint profits is in Morgenstern and Schwödiauer (1976) and Schotter (1983). In addition, they show that if the players are rational and know the core solution, that solution is unstable (for a general analysis and overview of the whole discussion, see Schotter and Schwödiauer 1980).

27). As accepted in the recent industrial organization literature; see, for example, Fudenberg and Tirole (1989). For clear discussions of the multiplicity of oligopolistic strategies see Dixit (1982) and Friedman (1983).