CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

Introduction: the two problems of social order

Hume wrote, in the abstract to the *Treatise of Human Nature*, that causality is the 'cement of the universe'. What ensures order in the physical world is that events of one type invariably follow upon events of another type. In this book, I discuss the conditions for order in the social world. What is it that glues societies together and prevents them from disintegrating into chaos and war? It is a big problem, second to none in importance. I do not claim to provide a complete answer, nor are the partial answers I offer very deep ones. At the present time, the social sciences cannot aspire to be more than social chemistry: inductive generalizations that stick closely to the phenomena. The time for social physics is not yet here, and may never come.[1]

I shall discuss two concepts of social order: that of stable, regular, predictable patterns of behaviour and that of cooperative behaviour. Correspondingly, there are two concepts of disorder. The first, disorder as lack of predictability, is expressed in *Macbeth*'s vision of life as 'sound and fury, a tale told by an idiot, signifying nothing'.[2] The second, disorder as absence of cooperation, is expressed in Hobbes's vision of life in the state of nature as 'solitary, poor, nasty, brutish, and short'. Instead of referring to predictability and cooperation, economists talk about equilibrium and Pareto optimality. For reasons that will emerge later, I do not adopt this

[1] 'Physics is parsimonious. A few basic ideas have a validity that extends across nature from the smallness of the atom to the vastness of the galaxy. Furthermore, these basic ideas capture a variety of factual information in the network of logical connections between them. The person who sees charm and beauty in the ideas of physics may see no enchantment whatsoever in chemistry. Lacking the simple predictive principles that are the stock in trade of physics, chemists are marvelous in their ability to hold in their heads at all times a vast array of information. Physicists, on the other hand, work from a base formed by a few remembered ideas' (Rigden 1987, pp. 36–7).

[2] This kind of disorder may, but need not, imply that the agents are uncertain about what to do. An agent may have a dominant strategy that leaves him in no doubt about what to do, but he may still be ignorant or entertain false beliefs about what others will do. In that case, he may feel *surprise* when the outcome materializes, but never *regret*.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

terminology. In Chapter 3 I argue that social norms ensure predictability outside equilibrium and in Chapter 5 that cooperation can lead to Pareto-inferior outcomes.

Disorder as failure to predict is dramatically illustrated by the stock market plunge of October 1987. Even though some analysts could, after the fact, truthfully say, 'I told you so', and even prove that they have put their money where their mouth was, similar Cassandras could probably have been found had the crash occurred six or twelve or twenty-four months earlier. Several writers have suggested[3] that the stock market may be a *chaotic* regime, in the technical sense of a system which is

> characterized by three attributes that can have extremely disturbing implications for the use of econometric forecasting procedures: a) Even though a time series is generated entirely deterministically its behavior is statistically very simliar to that of a system subject to severe random shocks; b) chaotic time series may proceed for substantial intervals of time manifesting patterns of behavior which seem extremely orderly, when a totally new pattern appears without warning, only to disappear just as unexpectedly; c) the presence and location of such abrupt transitions are extremely sensitive to parameter values in the underlying model, appearing and disappearing with changes in the third or higher decimal places, which are beyond anything econometrics may be able to aspire to discover.[4]

I am not sure, however, that this is the right direction in which to look for the sources of unpredictability. The nonlinear difference or differential equations that generate chaos rarely have good microfoundational credentials.[5] The fact that the analyst's model implies a chaotic regime is of little interest if there are no prior theoretical reasons to believe in the model in the first place. If, in addition, one implication of the theory is that it cannot be econometrically tested, there are no posterior reasons to take it seriously either.

To understand the problem of unpredictability we should look instead at the structure of social action and interaction.[6] Sometimes people have too little knowledge about others to anticipate what they will do and hence to

[3] See the interviews in Gleick (1987). [4] Baumol and Quandt (1985), p. 3.
[5] This is true, e.g., of the models in Day (1983) and Bhaduri and Harris (1987).
[6] See also Sen (1986).

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

predict the outcome. Sometimes they have too much knowledge. Sometimes they fail to use the knowledge they have. And sometimes no amount of knowledge, however ingeniously used, can help them. Let me illustrate these four cases, *assuming throughout that people are rational*. In Chapter 3 I discuss how nonrational behaviour may lend stability and regularity to situations that would otherwise be in hopeless flux.

Often, people have seriously incomplete information about each other's rationality, preferences and information. This need not be a destabilizing element. The president of the United States does not know whether the leaders of the Soviet Union would behave rationally in a crisis – for example, whether they would refrain from executing a threat which, at that time, it would not be in their interest to carry out. He does not know the preferences of his Soviet counterparts. (Do they really want world hegemony? Or do they simply want to be left in peace in their own backyard?) And he does not know what their beliefs are concerning his rationality, preferences and information. If, nevertheless, the balance of terror has been fairly stable, it is probably because both parties have made worst-case assumptions about each other, acting on their knowledge about the other party's objective capabilities rather than on any assumptions about subjective states of mind. There is order and predictability in spite of uncertainty and ignorance. (Indeed, more knowledge could make the situation less stable, as explained later.) By contrast, East–West relations have not been known to be orderly in the second sense, that of cooperative behaviour. Although the superpowers have avoided mutual destruction, the agreements on mutually beneficial arms reduction are recent and quite limited in scope.

Often, however, mutual ignorance *is* destabilizing. Consider, for instance, the cobweb cycle generated by people acting on the assumption that current prices will remain in force in the next period. If current prices exceed the equilibrium price, producers will market an above-equilibrium volume in the next period, thus forcing prices down. Assuming that prices will remain low, they will market little in the third period, forcing prices up again, and so on in a cycle that may converge to the equilibrium or diverge from it. Expectations are never fulfilled, and plans never realized. The culprit here is the producer's ignorance about consumers and about other producers. He does not know the full demand schedule. Nor does he know how other producers would react if they knew the schedule, because he does not know what they assume about each other. Under these circum-

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

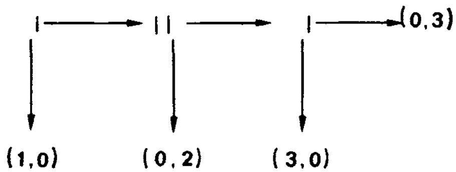

Fig. 1

stances, he might as well assume that prices will remain constant, without being under the illusion that this is any kind of mathematical expectation.[7] The status quo serves as a focal point for belief formation. Like the worst-case hypothesis, assuming perpetuation of the status quo is neither rational nor irrational. It is a maxim for decision making under uncertainty which, in this case, happens to undermine itself when adopted by many people simultaneously.

A surfeit of information can be destabilizing if it is beyond the processing capacities of the agent or organization receiving it. In a complex world, this problem arises frequently and can be quite important from a practical point of view. Theoretically, however, there is not much to say about it. It is more interesting and surprising that apparently simple strategic situations can be indeterminate if the agents are fully informed about each other, yet acquire determinacy if the information falls short of completeness. I shall illustrate this case with an analysis of games in extensive form, the putatively rational outcomes of which are determined by the method of backward induction.[8] Since this method is at the core of the modern approach to iterated games and bargaining, both of which are discussed extensively in later chapters, any doubts about its validity will have important repercussions.

Rather than define backward induction explicitly, I shall illustrate it by means of an example. Consider the game between two players depicted in Fig. 1.[9] There is *common knowledge* that both are rational: both are rational, and know each other to be rational, and know each other to know each other to be rational, and so on. The nature of the game is also common knowledge.

One player moves at a time, beginning with player I. Either he can move down and terminate the game, in which case he gets 1 and player II gets

[7] Keynes (1936), p. 151.
[8] According to Ståhl (1988) this method is due to Zermelo (1912).
[9] The example is taken from Bicchieri (1987).

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

nothing, or he can decide to continue the game to the second node, leaving the next move to II. She can similarly terminate the game, ensuring 2 for herself and nothing for I, or continue the game to the third node. There, I has a choice between two ways of terminating the game. One ensures 3 to himself and nothing to the other, while the other has the opposite payoffs.

It seems clear enough what will happen: I will move down on his first play and the game will end right there. To justify this conclusion, one traditionally invokes the principle of backward induction, reasoning from the last stage of the game back to the first. At the beginning of the game, I contemplates what he would do if he were at the last node of the game tree. He would, obviously, move down rather than across. Knowing that II knows him to be rational, I anticipates that at the second node II will play down, to get 2 rather than 0, which is what she would get if she played across. But that anticipation forces I to move down in the first move, to get 1 rather than 0, which is what he would get if he played across.

This reasoning seems compelling. But it harbours a problem: *why would I contemplate being at the third node?* How could he ever find himself there, if rational players would terminate the game at the first node? The issue turns on the use of counterfactual arguments: under what circumstances can we draw conclusions from premises known to be false? I have argued elsewhere[10] that counterfactuals are assertable when the additional premises used to draw conclusions from the counterfactual antecedent are consistent with that antecedent itself. We cannot, for instance, assert that in the absence of the railroad, economic growth would have been much the same because the automobile would have been invented earlier, if that assertion rests on a theory of technical change that would also allow us to predict the invention of the railroad. Similarly, we cannot assert that I will play down if the third node is reached if that claim is based on an assumption of I's rationality which is inconsistent with that node being reached.

The question, therefore, can be restated as follows. Does there exist a set of assumptions about the players which are strong enough to allow us to infer what the players would do at various nodes in the game while also weak enough to be consistent with these nodes being reached? I have argued that the assumption of rationality and common knowledge is too strong to satisfy the second requirement. Other assumptions may, in special cases, satisfy both requirements.[11]

[10] Elster (1978), ch. 6.

[11] The following draws on Binmore (1987a), summarizing the work of Selten, Kreps and Wilson and others. See also Rubinstein (1988a) for perceptive comments.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

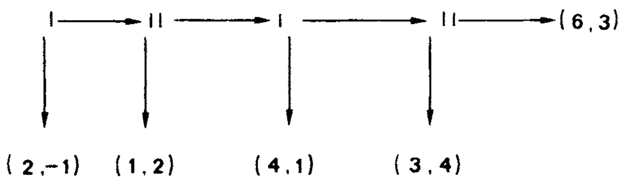

Fig. 2

1. Player I could assume that the later nodes have been reached as a result of mistakes (the 'trembling hand' assumption). The third node might be reached, that is, if there were some probability that a player might, as it were, push the wrong button. If we consider a game similar to the one in Fig. 1, but extended to a hundred successive moves, this assumption becomes extremely implausible, because it would require each player to make fifty uncorrelated errors before arriving at the final node.

2. A second assumption, therefore, is that the errors are correlated, for example, that II is irrational. From the point of view of player I, asking himself how he might have reached the final mode, it makes more sense to assume that II always moves across (the 'automaton' assumption) and that he has reached the final node by exploiting II's irrationality.

3. Player I might consider a more elaborate possibility, namely that II is either an irrational automaton or a rational player who is deliberately pretending to be irrational in order to induce I to play across rather than down. In the game portrayed in Fig. 1 this assumption has no purchase, but in the game depicted in Fig. 2 it could provide a plausible explanation of why later nodes in the game tree might be reached.

To explain why he might find himself at the third node, I may assign probability p to II being an automaton who always moves across and probability $1-p$ to II being a rational agent who fakes automatic behaviour in the earlier stages of the game in order to induce I to move across so that she can move down in later stages.

4. We may weaken the assumption that the rationality of the players is common knowledge.[12] In the game portrayed in Fig. 1, we now assume only that I believes that II is rational, that II believes that I is rational and that I believes that II believes that I is rational. When I contemplates being at the third node, he must ask himself whether this assumption is consistent

[12] This argument is due to Bicchieri (1987).

© Cambridge University Press www.cambridge.org

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

with the belief set. In particular, could II have played across at the second node? To answer this question, he must first explain how she could make sense out of being at the second node. By virtue of the stipulated belief structure, he cannot explain it by assuming that she believes him to be irrational, but he may assume that she believes him to believe her to be irrational, for in that case she might reason that he has played across at the first node in the hope that she might do so at the second. Now, he knows that if the second node is reached, she will in fact play down, as he believes her to believe him to be rational and hence will expect him to play down at the third node. Accordingly, he plays down at the first node. The backward induction argument works, but only because the players' rationality is *not* common knowledge. If the players have less initial knowledge about each other, they can form stable beliefs about each other's behaviour.

Assumptions 1 through 4 differ as follows. Assumption 1 stipulates that I entertains a certain subjective probability that II, while rational, is fallible. Assumption 2 stipulates that I entertains a certain subjective probability that II is an irrational automaton. Assumption 3 stipulates that if, in fact, II is rational she will be aware of I's probability assessment and try to exploit it to her advantage, and that I knows that if II is rational she will be aware of it, and so on. Assumption 4 stipulates that the players are rational and infallible, but that their knowledge about each other's rationality, and knowledge about this knowledge and so on, has an upper limit. Each assumption may, in special cases and under special conditions, provide what we are looking for: a theory which is strong enough to allow us to infer what rational players will do at the various nodes and weak enough to be consistent with their being at those nodes. But none of the assumptions seems to have the simplicity and generality that could support backward induction in a more general context. As a consequence, that principle itself is more vulnerable than appears at first glance.

I suspect that the last word on backward induction has not been said. In later chapters I shall, with some qualms, retain the principle, partly because I am not sure my understanding of these matters is sufficiently deep to allow me to discard it altogether, and partly because the principle may be behaviourally adequate when agents are less than perfectly rational.[13] The fallacy in the backward induction argument, like the fallacies in many

[13] 'My experience suggests that mathematically trained persons recognize the logical validity of the [backward induction] argument, but they refuse to accept it as a guide to practical behavior' (Selten 1978b, p. 133). See also Rubinstein (1988b).

© Cambridge University Press www.cambridge.org

Cambridge

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

other counterfactual arguments, is so subtle that it may not be perceived by ordinary agents going about their business. The problem awaits further theoretical and empirical clarification.

Predictive failures may also occur because people fail to make good use of information which they have. This is irrationality rather than indeterminacy of rationality. In the 1972 presidential campaign, for instance,

> on election eve a large group of the reporters following the McGovern campaign sagely agreed that McGovern could not lose by more than 10 points. These people were wire service reporters, network television reporters, and major newspaper and newsmagazine reporters. They knew that all the major polls had McGovern trailing by 20 points, and they knew that in 24 years not a single major poll had been wrong by more than 3%. However, they had seen with their own eyes wildly enthusiastic crowds of tens of thousands of people acclaim McGovern.[14]

Securities and futures markets also attach excessive importance to current information and insufficient importance to information about the past.[15] A converse fallacy – trying too hard to understand the past – can also lead to predictive failure. For any given set of events in the past, it is usually possible, by looking around for some time, to find some other event set that correlates highly with it. If one requires a 5 per cent significance level, twenty attempts will on the average be sufficient. The chances are, however, that correlations obtained in this way will be spurious and useless for predictive purposes. 'The price that investment analysts pay for overfitting is their long-run failure to predict any better than market averages'.[16]

Finally, some situations are inherently unpredictable. No matter how much or how little information the agents have, and no matter how ingeniously they use it, they will not be able to predict what others will do. I assume that for prediction among and by rational agents to be possible, the predicted outcome must be an equilibrium, that is, a state in which no agent has an incentive to behave differently. Failures of predictability may then occur for three reasons: some situations have no equilibrium; some have multiple equilibria; and some have equilibria which are too unstable to serve as the basis for prediction. The first category is not, perhaps, more than a curiosum,[17] but the other two are quite important.

[14] Nisbett et al. (1982), p. 116. [15] Arrow (1982). [16] Fischhoff (1982), p. 345.

[17] An example is the game in which each person writes down a number and the person who has written down the largest number gains the difference between his number and the

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

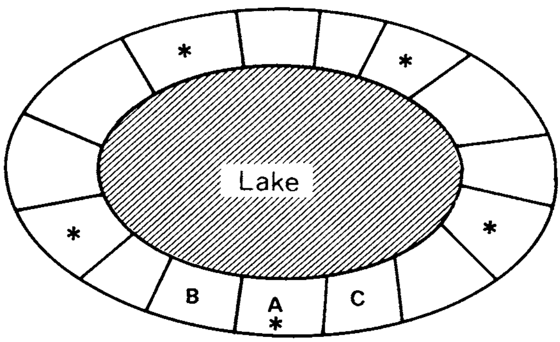

Fig. 3

Games with multiple equilibria that have different winners and losers can be unfathomable. Consider a number of peasant families with plots arranged alongside each other around a lake.[18] The peasants do not have enough cultivated land and would like to fell the trees on their plot, but know that deforestation may bring erosion. Specifically, erosion will occur on any given plot (*A* in Fig. 3) if and only if trees are felled on that plot and on the two adjoining ones (*B* and *C*). Here there are three equilibria, in each of which trees remain standing on every third plot around the lake. One equilibrium, in which trees are felled on the starred plots, is shown in Fig. 3. There is no tacit coordination mechanism, however, by which one of the three equilibria could emerge as the predictable outcome.

Voting – discussed more extensively in Chapter 5 – is another situation with multiple equilibria. The value of voting to the individual depends on how many other people vote. If everybody else votes, the individual has no incentive to do so, since the chance of his being pivotal is negligible. If nobody else votes, he has a strong incentive to do so, since he can decide the outcome by himself. An equilibrium has an intermediate number of voters, each of whom prefers voting to not voting but would prefer not voting to voting if one additional person voted. If there are n voters altogether and m voters in equilibrium, the number of equilibria equals the number of ways one can select m people from a set of n. Since m is usually small, elections with large electorates will have very many equilibria.

average of all numbers. Decentralized wage bargaining with no holds barred could provide a real-life illustration. Assuming that firms can shift any cost increases onto consumers, workers in each industry or firm have an incentive to ask for higher wage increases than those obtained by other workers. Something like this may happen in hyperinflation.

[18] For a fuller discussion of this example, see Elster (1989b), chs. 9 and 10.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-37607-5 - The Cement of Society: A Study of Social Order
Jon Elster
Excerpt
More information

A very general weakness of rational expectation models in economics is that they tend to have multiple equilibria.[19] For an individual, the possibility that there are several predictions he might make, each of which would be self-fulfilling, poses no problem.[20] He can simply make the prediction whose outcome he prefers. A group of individuals who know that if they all act on one set of expectations these will come out true, and if all act on another set these, too, will be verified, are in a more difficult predicament. In special cases they may be able to coordinate their actions. If one set of expectations yields an outcome preferred by everybody, it will be chosen. Also, asymmetries of power can stabilize the situation and ensure coordination around a cooperative equlibrium. The peasants around the lake might achieve coordination through bargaining, if some of them have other plots not threatened by erosion. Because they can survive without their plot by the lake, they can credibly announce that they are felling trees on their land and that others will have to adjust to their actions.[21] But there is no general mechanism for ensuring coordination when there are several equilibria with different winners and losers.[22]

Many games have no equilibrium in pure strategies. Investment in research and development is a plausible example.[23] The only equilibria such games admit consist of mixed strategies, defined as a probability distribution over a subset of the pure strategies. Now, we do not often observe people using lotteries to make decisions in non-zero-sum interactions,[24] and for good reasons. It can be shown that in equilibria with mixed strategies an individual can do no worse for himself – although by definition no better – by using any other probabilistic combination of the pure strategies that enter into his equilibrium strategy, as long as others stick to their equilibrium strategies. The tiniest flicker of uncertainty or ignorance could then induce a shift to his maximin strategy: it will protect him if others

[19] See, e.g., Begg (1982). [20] Elster (1984), pp. 48, 106.

[21] This principle of 'justice according to Saint Matthew' – to him that hath shall be given – is further discussed in Chapter 2.

[22] Harsanyi and Selten (1988) is a book-length attempt to confront and solve this problem. Yet as observed by Robert Aumann in his foreword to their book, 'Although the theory selects a unique equilibrium, as a theory it need not be unique'. Indeed, their solution concept may be seen as representing one of several *reflective equilibria* in the sense of Rawls (1971). Hence it is not clear that their theory will have much predictive and explanatory power, unless or until it is adopted by economic agents in their decision making.

[23] Dasgupta and Stiglitz (1980).

[24] Elster (1989a), ch. 2, has a survey of the use of lotteries in decision making. See also Rubinstein (1988a) for some further critical comments on the explanatory value of mixed strategies.

© Cambridge University Press www.cambridge.org