Rationality and Coordination, CRISTINA BICCHIERI. Cambridge University Press, 1994, xiii + 270 pages.

In her book *Rationality and Coordination* (Cambridge University Press, 1994) Cristina Bicchieri brings together (and adds to) her own contributions to game theory and the philosophy of economics published in various journals in the period 1987–1992. The book, however, is not a collection of separate articles but rather a homogeneous unit organized around some central themes in the foundations of non-cooperative game theory. Bicchieri's exposition is admirably clear and well organized. Somebody with a good knowledge of game theory would probably benefit mainly from reading the second part of Chapter 3 (from Section 3.6 onward) and Chapter 4. On the other hand, those who have had little exposure to game theory, would certainly benefit from reading the entire book. I shall begin with an overview of the content of the book and then offer some critical comments on what I consider to be the most important part of it.

Chapter 1, entitled 'Rationality and predictability', deals with 'the role individual rationality plays in predicting and explaining economic phenomena' (p. 1). Bicchieri distinguishes between two kinds of rationality: *practical* and *epistemic*. Practical rationality coincides with expected utility maximization, while epistemic rationality means that the agent's beliefs are rational. Bicchieri argues that, with very few exceptions, agents must be credited with both practical and epistemic rationality if individual as well as systemic behavior in interactive contexts is to be explained satisfactorily. In particular, it is necessary to understand the *process* by which agents form their beliefs about the environment, which includes the actions of the other agents. This is the central theme of the entire book. The chapter ends with a discussion and criticism of the rational expectations hypothesis.

Chapter 2, entitled 'Equilibrium', contains a general discussion of games in normal (or strategic) form, introducing the notions of pure and mixed strategy, weakly and strictly dominant strategy, rationalizability and Nash equilibrium. The discussion also covers the concepts of focal point, Pareto dominance, trembling-hand perfect equilibrium and proper equilibrium. As Bicchieri states at the beginning of the chapter (p. 33), 'the general thesis ... is that, in many non-trivial cases, knowledge of the game being played and self-evident principles of rationality are not sufficient to guarantee that an equilibrium will be attained'. For a reader with limited knowledge of game theory Chapter 2 is useful reading. However, Bicchieri's discussion leaves out some new developments in the literature, which are highly relevant to the issues that are central to this book. For example, a mixed strategy is defined on page 44 as a conscious randomization by the player over her pure strategies. On page 61 the 'purification' idea of Harsanyi is briefly mentioned and dismissed. Bicchieri does not discuss a different interpretation of mixed strategies, first suggested by Aumann, which is gaining momentum in the game-theoretic literature. Aumann (1987) suggests a re-interpretation of mixed strategies according to which 'Player 1 plays L with probability $\frac{1}{3}$' is a proposition not about how Player 1 determines what he will do, but about Player 2's beliefs about what Player 1 will do. On this interpretation a *Nash equilibrium is a relation that holds between beliefs and not between choices of plans*. Aumann then shows that if it is common knowledge that each player is rational in the Savage sense, then their beliefs must be in equilibrium (it might be a correlated equilibrium, rather than a Nash equilibrium, since a player may coherently believe that the choices of different opponents are correlated, due, for example, to a common experience or background). This approach and the study of the epistemic conditions for Nash equilibrium have been pursued further by Aumann and Brandenburger (1991) (see also Brandenburger, 1992).

Chapter 3, entitled 'Epistemic rationality', introduces the reader to games in extensive form. The notions of perfect and imperfect information, backward and forward induction, subgame-perfect, sequential, trembling-hand perfect and proper equilibrium are explained and illustrated with examples. As in most of the book, the discussion leaves out the technical details and the precise definitions and concentrates on the intuitive content of such notions. The last part of the chapter puts forward a theory of out-of-equilibrium belief revision based on the principle of minimum loss of informational value. Here Bicchieri applies Gärdenfors theory of minimal belief revision to the problem of justifying moves that are ruled out by equilibrium play. According to this theory, when an unexpected move is observed, players revise their original beliefs (which ruled out the observed move) by eliminating first those beliefs that have lower informational value. I shall come back to this in my critical comments below.

Chapter 4, entitled 'Self-fulfilling theories', deals with the problem of 'how much players need to know about the game and other players in order to complete the reasoning required of them and infer a solution' (p. 128). In this chapter Bicchieri deals only with two-player extensive games with perfect information. Her aim is to justify the backward induction solution. I shall give a more detailed account, as well as some critical comments, of the material of Chapter 4 later on.

Chapter 5, entitled 'Paradoxes of rationality', contains a general discussion of promises, threats, commitment, binding versus non-binding agreements, the possibility of cooperation that arises when the game is repeated over time, and reputation. Bicchieri's treatment of these issues is at an introductory level, with simple illustrative examples. The formalism and the technical details are kept to a minimum. Chapter 5 also contains an extended discussion (and criticism) of Gauthier's concept of constrained optimization as well as a proposed solution to the surprise test paradox, based on the hypothesis that players' information is limited (e.g., the students know that the teacher is rational, but the teacher does not know that the students know that he is rational).

Chapter 6, entitled 'Learning and norms: The case of cooperation' deals with social norms of cooperation. Here Bicchieri adopts a view, first put forward by Lewis and later expanded by Ullman-Margalit, according to which social norms can be analyzed in game theoretic terms and are broadly defined as Nash equilibria. Bicchieri (p. 232) gives the following definition of a social norm: 'let R be a behavioral regularity in population P. Then R is a social norm if and only if: (1) almost every member of P prefers to conform to R on the condition (and only on the condition) that almost everyone else conforms, too; (2) almost every member of P believes that almost every other member of P conforms to R'. Bicchieri points out that an explanation of social norms in terms of Nash equilibria raises the question of how such norms emerge and come to be followed by the population at large. Her thesis is that (p. 222) 'norms emerge in small-group interactions and may subsequently spread to an entire population through an evolutionary mechanism'. Thus social norms are the outcome of learning in a strategic interaction context. To illustrate this point, Bicchieri gives a simple example of an evolutionary process. A potential source of dissatisfaction with this chapter is the lack of an adequate account of the literature. For example, on the topic of social norms and conventions Sugden's work (1986, 1989) seems to be highly relevant, and on the topic of the emergence of Nash equilibria through learning in a strategic interaction context there is a large and fast growing literature (for an excellent survey, see Battigalli, Gilli and Molinari, 1992).

I shall now give a more specific account of, and some critical comments on, the most interesting part of Bicchieri's book, which in my opinion is the second part of Chapter 3 and Chapter 4.

The second part of Chapter 3 deals with the notion of belief revision with minimum loss of informational value. Given an extensive game, Bicchieri assumes that each player i starts with a model of the game denoted by M_i^0. 'This model is a state of belief representable as a set of sentences expressed in a given language L' (p. 11). Bicchieri does not explain what, in general, this language L is, or should be. In the examples she uses, the initial models M_i^0 $(i = 1, 2)$ – which are assumed to be common knowledge – include the following sentences (the particular extensive game she refers to is irrelevant for what I want to say):

(i) The players are rational (e.g., expected utility maximizers).
(ii) The players always play what they choose.

(iii) Player 1 chooses to play c.
(iv) Player 1 plays c.

In order to decide whether the equilibrium considered is plausible, Player 1 will ask herself what would happen if she did *not* play c. Observing a choice different from c requires a player to modify her initial beliefs by giving up her belief in sentence (iv). Since (iv) is implied by the conjunction of (ii) and (iii), she would also have to give up one of them, or both. The principle of minimal belief revision requires that she give up either only (ii) or only (iii). Which of the two should be dropped depends on how they rank in terms of informational value. Depending on the payoff structure, sometimes it can be argued that (ii) has more informational value than (iii) and sometimes that the opposite is true. Bicchieri shows that this theory of minimal belief revision can give a theoretical foundation for such concepts as forward induction. Although I find the notion of minimal belief revision intriguing and potentially very interesting, I would have liked to see a discussion of what general criteria should be used in selecting the language L that is the basis for the initial belief sets. For example, sentence (ii) seems 'natural' just because of its implicit reference to the notion of trembling-hand perfect equilibrium, with which we are all familiar. One could think of a number of equally plausible alternatives to sentences (i)–(iv) above. In any case, one ends up with an explanation which is based on an *ad hoc* selection of the sentences that represent the initial belief sets. Bicchieri is aware of these difficulties, when she writes (p. 123, emphasis added):

> *External considerations* will lend plausibility to the selection of sentences we include in the model of the game, as well as to the arguments favoring some given ordering of informational value, *but obviously these will hardly be the only reasonable possibilities.*

I now turn to Chapter 4. Here Bicchieri argues (p. 134) that 'Backward induction as a reductio proof is a proof given outside the game by an external observer. If we instead want to model how the players themselves reason to an equilibrium, we have to model how they come to decide that a given action is optimal for them'. Bicchieri uses the game of Figure 4.1 (p. 132), reproduced below, to illustrate the players' reasoning. The backward induction solution of this game is $(l_1 l_2, L)$. Let R_1 stand for 'Player 1 is rational', R_2 for 'Player 2 is rational' and, if p is a proposition, K_1p for 'Player 1 knows that p' and K_2p for 'player 2 knows that p'. Bicchieri makes the following assumptions about players' rationality and mutual knowledge of rationality:

(1) at node I^{12}: R_1;
(2) at node I^{21}: R_2 and K_2R_1;
(3) at node I^{11}: R_1, K_1R_2 and $K_1K_2R_1$.

She then shows that each player can use his/her knowledge to justify his/her part of the backward induction solution (pp. 137–8). Rationality of player 1 is sufficient for player 1 to justify his choice of l_2 at I^{12}. Since player 2 knows that player 1 is rational, she knows that if she chooses R at I^{21}, player 1 will follow with l_2. Hence, since she is rational, she will choose L at I^{21}, etc. Bicchieri also argues that, based on the knowledge attributed to them, each player would be able to explain the fact that his/her node is reached (if indeed it is reached) in a way which is consistent with what he/she knows. For example, 'can player 2 observe r_1 and still maintain that 1 is rational? Given what player 2 knows, the answer is yes. Player 2 only knows that 1 is rational, but does not know whether 1 knows that she is rational. The choice of r_1 on 1's part is then explainable as reflecting his uncertainty as to the rationality of player 2' (p. 140).

Bicchieri then states the following theorem (p. 138):

> **Theorem 4.2.** In finite extensive form games of perfect and complete information, the backward induction solution holds if the following conditions are satisfied for each player i at each information set I^{ik}: (α) player i is rational and knows it, and knows his available choices and payoffs, and (β) for every information set I^{jk+1} that immediately follows I^{ik}, player i knows what player j knows at information set I^{jk+1}.

Throughout the book, rationality of a player is taken to mean that the player will always choose the option that gives her the highest (expected) utility:

> We … define a player's rationality in the following way: if a player can choose among several actions a_1, … a_n, and action a_i leads to a payoff of at most x and action a_j leads to a payoff of at least y, and y > x, then if i is rational he will not choose action a_i (p. 147).

Now, consider the game of Figure 4.1 modified as follows: the payoff to player 1 if he chooses l_1 is 15 (rather than 1). Thus l_1 is a strictly dominant choice. The backward induction solution is still (l_1l_2, L) and we should be able to justify it by following the recipe of Theorem 4.2. Thus we should attribute rationality to player 1 at node I^{12} in order to conclude that there he would choose l_2, and at node I^{21} we should attribute to player 2 rationality as well as knowledge that player 1 is rational in order to justify the choice of L there. However, at I^{21} player 2 knows that player 1 chose r_1 at I^{11}, despite the fact that simple rationality of player 1 requires him to choose l_1 at I^{11} [using the symbolism of p. 147: $R_1 \Rightarrow l_1$ or, equivalently,

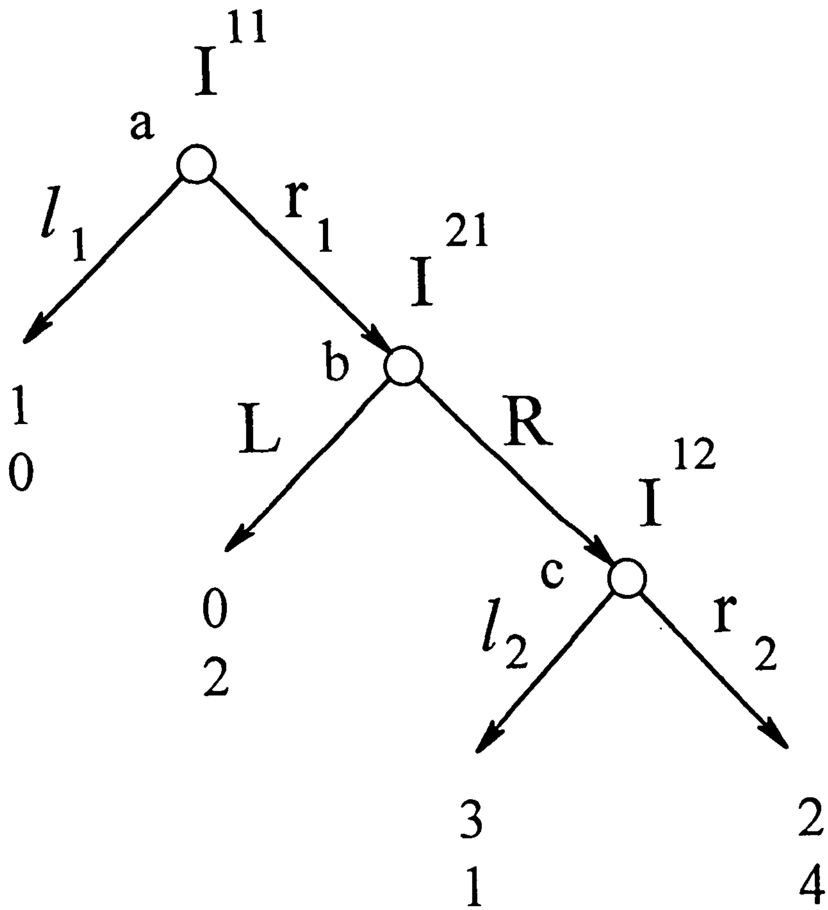

FIGURE 4.1 (page 132)

$\sim l_1 \Rightarrow \sim R_1$; since $r_1 \Rightarrow \sim l_1$ (cf. A_7 on p. 147), it follows that $r_1 \Rightarrow \sim R_1$]. Can player 2 know, at node I^{21}, that player 1 is rational?

One could argue that in this case (game of Figure 4.1 modified so that the payoff to player 1 if he chooses l_1 at I^{11} is 15) there is no need to use the backward induction algorithm to determine that player 1 will choose l_1 at I^{11}, since l_1 is a strictly dominant choice. This objection raises the question of how the backward induction solution ($l_1 l_2$, L) is to be interpreted. There are (at least) two possible interpretations:

(1) The solution is the *play* generated by ($l_1 l_2$, L), that is, the *outcome* associated with the terminal node reached by l_1. According to this interpretation all one can predict is that player 1 will play l_1, but nothing

can be said about what player 2 would choose if node I^{21} were to be reached;

(2) The solution is the entire strategy profile ($l_1 l_2$, L) interpreted as follows: player 1 plays l_1. If, however, node I^{21} happened to be reached, then player 2 would play L and if, despite all this, node I^{12} were to be reached then player 1 would play l_2.

Bicchieri's discussion in Chapter 4 seems to reflect interpretation (2), since her main objective is to construct the reasoning that leads a player to choose her component of the backward induction solution. But then, as argued above, unless one is willing to restrict the class of extensive games considered, one must deal with the following question: can a player be known to be rational at node x even if she is known to have chosen a dominated action at a predecessor of x? Intuition seems to suggest that the answer is negative, unless one adopts a notion of 'imperfect' rationality, such as the notion implicit in Selten's definition of trembling-hand perfect equilibrium. Perhaps 'rationality' is one of those words that need to be defined precisely, because different people are likely to attach different meanings to it. Indeed, one could argue that one of the main conclusions that can be drawn from the recent literature on the foundations of game theory is that it is extremely important, if not essential, to give explicit and precise definitions of all the words used, such as 'rationality', 'knowledge', 'common knowledge', etc. An outside observer might be puzzled by the contrasting claims that can be found in the literature. For example, Bicchieri states (p. x) that 'Chapter 4 proves that common knowledge [of rationality] is neither necessary nor sufficient to obtain a solution in most games belonging to the class discussed [finite, extensive form games of perfect information], and that "too much knowledge" of the theory of the game makes that theory inconsistent'. On the other hand, Aumann (1995) proves that common knowledge of rationality in such games is possible and implies backward induction. At the root of such contrasting claims lie different meanings attributed to terms such as 'rationality' and different modeling choices. Some people believe that a precise formalism sometimes is unnecessary and may even obscure the real issues (see, for example, Binmore, 1994). On my part I believe that precise definitions and complete rigor aid rather than hinder understanding, for then one can concentrate on the question of whether the proposed definitions capture our intuitive understanding of the corresponding concepts, while all ambiguity is removed concerning the other question, namely what can be deduced from those definitions.

Giacomo Bonanno

University of California, Davis

REFERENCES

Aumann, Robert. 1987. 'Correlated equilibrium as an expression of bayesian rationality'. *Econometrica*, 55:1–18

Aumann, Robert. 1995. 'Backward induction and common knowledge of rationality'. *Games and Economic Behavior*, 8:6–19

Aumann, Robert and Adam Brandenburger. 1991. 'Epistemic conditions for nash equilibrium'. Working Paper No. 91–042, Harvard Business School, Boston

Battigalli, P., M. Gilli and C. Molinari. 1992. 'Learning and convergence to equilibrium in repeated strategic interactions: an introductory survey'. *Ricerche Economiche*, 46:335–78

Binmore, Kenneth. 1994. 'Backward induction'. Mimeo, University College London

Brandenburger, Adam. 1992. 'Knowledge and equilibrium in games'. *Journal of Economic Perspectives*, 6:83–102

Sugden, Robert. 1986. *The Economics of Rights, Co-operation and Welfare*. Basil Blackwell

Sugden, Robert. 1989. 'Spontaneous order'. *Journal of Economic Perspectives*, 3:85–97

Economics and Evolution, GEOFFREY HODGSON. University of Michigan Press, 1993, xi + 381 pages.

> 'Nonbiologists who favor the evolutionary conceptualization are often unaware of the Darwinian or neo-Darwinian theory and may, for instance, promote orthogenetic schemes ... A study of such literature demonstrates rather painfully that no one should make sweeping claims concerning evolution in fields outside the biological world without first becoming acquainted with the well-seasoned concepts of organic evolution and, furthermore, without a most rigorous analysis of the concepts he plans to apply.' (Mayr, 1982, p. 627)

Geoffrey Hodgson's new book admirably demonstrates the wisdom of Ernest Mayr's admonition, and for this reason alone it is one of the most important books published in the history of economic thought in the last decade. But its significance extends well beyond salutary warnings about the ineptitude of famous dead white male economists in their quest to domesticate biology. It is also simultaneously a forceful attempt to revive the flagging tradition of American Institutionalism in the image of Veblen, Commons and Mitchell, and furthermore to make the case that economics *should* be patterned upon biology in the future. While its success in the first endeavor is exemplary, I nevertheless harbor a few reservations about its triumph in the latter objectives. These qualms will require situation within a larger context of the place of biology in the panoply of *fin-de-siècle* sciences and Western culture. But first, as befits the canons of book reviewing etiquette, we lead off with fulsome praise.

Hodgson is quite simply the first economist who has bothered to acquaint himself sufficiently with the literature of biology, both current