UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DEUX PERSPECTIVES SUR LES FONDEMENTS DE LA THÉORIE DES JEUX

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN ÉCONOMIQUE

PAR

JEAN GUILLAUME FORAND

DÉCEMBRE 2003

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

TWO PERSPECTIVES ON THE FOUNDATIONS OF GAME THEORY

THESIS
SUBMITTED
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER'S IN ECONOMICS

BY
JEAN GUILLAUME FORAND

DECEMBER 2003

REMERCIEMENTS

Merci au directeur de ce mémoire, M. Robert Leonard. Non seulement pour l'aide directe qu'il m'a prêtée pendant sa rédaction, dont l'ampleur est incalculable, mais également pour ces nombreuses conversations qui n'avaient de liens avec le mémoire que la curiosité et le plaisir d'apprendre. Merci pour son amitié et celle de sa famille.

Merci à Myriam, dont la compagnie m'inspire et m'apaise, et qui m'en enseigne autant que ne le font mes livres.

Mes parents méritent plus de remerciements que je ne puis leur en offrir. Sans eux, bien plus que ce mémoire n'aurait pas fait son chemin jusqu'à maturité. Merci à mon père pour ses livres, et à ma mère pour sa persévérance. Merci à mon frère, parce qu'il est mon frère et qu'il faut toujours remercier son frère.

Finalement, merci à la totalité du département des sciences économiques de l'UQAM pour un séjour de deux ans en son sein qui fut très agréable. À un satanément bon troupeau de camarades étudiants. Aux professeurs aux portes desquels j'ai cogné pour poser des questions parfois bêtes, en particulier MM. Max Blouin et Stéphane Pallage. Et aux administratrices, auxquelles j'ai posé des questions encore plus confuses que celles que je posais aux professeurs.

ACKNOWLEDGMENTS

My thanks to my supervisor, Robert Leonard, not only for his inestimable help while writing this thesis, but for those many other conversations, the effect of which was to underline for me the importance of curiosity and the pleasures of learning. I thank him for his and his family's friendship.

I thank Myriam, whose companionship is a source of inspiration and comfort, and who has taught me more than have all my books.

My parents deserve more thanks than I can offer. Without them, it's not only this thesis that would have failed to reach maturity. I thank my father for his books; my mother for her perseverance. As for my brother - well, let me thank him for being my brother. After all, one must always thank one's brother...

Finally, my thanks to all in the department of economics at UQAM for a most pleasant two-year stay in their midst. I am grateful to an absolutely stellar group of student comrades; to the faculty members, in particular Max Blouin and Stéphane Pallage, on whose doors I knocked with my often silly questions; and to the support staff, my questions of whom were even more confused than those asked of my teachers.

À mes parents, à qui je dois presque tout
de ce qui, dans la vie, est réellement important

To my parents, to whom I owe virtually everything in life that is of true importance

TABLE OF CONTENTS

LIST OF FIGURES ix

RÉSUMÉ x

ABSTRACT xi

INTRODUCTION 1

0.1 THE COMPLEXITY OF DEBATES IN SCIENCE 1

0.2 LOUD AND RUDE: BINMORE AND AUMANN'S EXCHANGES 8

CHAPTER I
AUMANN: ANTEAS' VIEW OF STRATEGIC INTERACTION 12

1.1 FORMALIZING COMMON KNOWLEDGE 12

1.2 CORRELATED EQUILIBRIA 18

1.3 COMMON KNOWLEDGE OF (A LITTLE BIT OF) IRRATIONALITY 32

1.4 THE EPISTEMIC FOUNDATIONS OF NASH EQUILIBRIUM 39

CHAPTER II
BINMORE: THE LABOURS OF HERCULES 46

2.1 A QUEST BEGINS 46

2.2 TEARING EVERYTHING DOWN 48

2.2.1 RATIONALITY AS PROCESS 49

2.2.2 LEARNING AS SURPRISE 55

2.3 THE TURNCOAT MATHEMATICIAN 60

2.4 A LITTLE PHILOSOPHY HERE, A LITTLE PHILOSOPHY THERE, A LITTLE PHILOSOPHY EVERYWHERE! 65

2.4.1 IF YOU WANT SOMETHING DONE RIGHT, GET YOURSELF A GAME THEORIST 66

2.4.2 BOYS, DON'T LET THOSE ASSUMPTIONS OUT OF YOUR SIGHT 71

2.5 PARADOXES 77

2.5.1 THE GÖDEL CAN-OPENER FOR HERMETICALLY SEALED WORLDS 77

2.5.2 A BIG FUSS ABOUT THINGS THAT YOU CANNOT EVEN MEASURE 85

2.6 POSITIVE CONTRIBUTIONS 91

2.6.1 FROM MYTHOLOGY TO MUDDLERS... 92

2.6.2 ...AND BACK AGAIN 110

CHAPTER III THE PATHS MEET 115

CONCLUSION 130

REFERENCES 140

LIST OF FIGURES

Figure		Page
1.1	The strategic-form Centipede Game	37
3.1	The hundred-legged Centipede Game in extensive form	119

RÉSUMÉ

Ce mémoire a comme toile de fond une succession de débats publics acrimonieux et mouvementés au sujet des fondements de la théorie des jeux. Ceux-ci ont lieu en 1996 et mettent en scène deux importants théoriciens des jeux, Robert Aumann et Kenneth Binmore.

Robert Aumann est un théoricien des jeux éminent, actif dans le domaine depuis la fin des années 1950. À partir des années 1970, il révolutionne l'examen des fondements épistémiques de la théorie des jeux. Il apporte aux questions traitant de connaissance, de connaissance commune et de rationalité, des réponses inspirées de la théorie de la décision bayesienne. Ses modèles ne tracent pas de distinctions normatives entre les différents équilibres qui peuvent être atteints durant le jeu; ils décrivent le type d'équilibres dont pourrait témoigner un observateur extérieur une fois le jeu terminé, étant donné les conditions épistémiques dont sont dotés les joueurs au début du jeu.

Kenneth Binmore est un théoricien de haut calibre et également un critique persistant, auteur de plusieurs textes 'méthodologiques'. Insatisfait d'une théorie des jeux traditionnelle qu'il attaque sans répit, il cherche à convaincre ses collègues de se tourner vers la naissante théorie des jeux évolutionnistes. Il fait ceci tout en insistant auprès des chercheurs des différentes sciences sociales, surtout les philosophes de la politique, que la théorie des jeux doit constituer l'échafaudage logique qui unifiera leurs domaines respectifs. Deux filons importants parcourent les ouvrages qui semblent dispersés de Binmore. Le premier est un souci pour les procédés qui mènent aux décisions; les mécanismes de la rationalité. Le second s'attarde aux formes primitives de ces procédés; les cas où les forces de l'évolution se révèlent importantes.

En examinant de près les travaux de ces deux auteurs dans le champ du fondement de la théorie des jeux, ce mémoire tente de transcender et d'expliquer le semblant de confusion qui règne dans les échanges où ils s'affrontent. En recréant et en contrastant les trajets que les auteurs suivent jusqu'à leurs confrontations, en expliquant et en élucidant les questions que celles-ci évoquent, on dépasse rapidement le cadre de la théorie des jeux. Les préjugés philosophiques et intuitifs des auteurs prennent dès lors beaucoup d'importance. Ce mémoire nous confronte à la place primordiale qu'occupent des tendances psychologiques plus profondes dans la compréhension des contributions théoriques de ceux qui créent la science, même lorsque celles-ci sont très techniques et mathématisées. Le rôle et les méthodes de la science, l'importance et le maniement 'correct' des mathématiques dans la théorie des jeux, la forme et la nécessité du rapport à l'expérimentation et à l'empirique; les idées qu'entretiennent Binmore et Aumann sur ces questions pèsent lourdement dans un débat qui, à certains égards, peut sembler être préoccupé de points précis.

Mots clés: Théorie des jeux, Histoire de la pensée économique, Histoire des sciences, Psychologie de la création scientifique.

ABSTRACT

The background to this thesis is a series of dramatic and heated public debates on the foundations of game theory. These took place in 1996, and featured two important game theorists, Robert Aumann and Kenneth Binmore.

Robert Aumann is an eminent game theorist, active in the field since the end of the 1950's. Beginning in the 1970's, he revolutionized the field of the epistemic foundations of game theory. Using Bayesian decision theory, he answered questions relating to knowledge, common knowledge and rationality. His models cast no normative judgement on the different equilibria that can be achieved in a game; they describe the type of equilibria that might be observed by an external observer of the game once it has ended, give the epistemic capacities of the players at the outset of the game.

Kenneth Binmore is a first-rate theorist and persistent critic, and author of several papers on "methodological" issues. Unhappy with traditional game theory, which he attacks continuously, he has tried to convince his colleagues to turn towards the embryonic evolutionary game theory. This he has done by insisting to social scientists of various stripes, especially political philosophers, that game theory must constitute the logical scaffolding that will unite the various social scientific fields. Two important threads run through Binmore's variegated works. The first is his particular interest in the processes leading to decisions; the mechanisms of rationality. The second focuses on the primitive forms of these procedures; those cases where the forces of evolution appear to be important.

Carefully examining the works of these two authors in the area of the foundations of game theory, this thesis attempts to view illuminate and explain the appearance of confusion that has characterized the exchanges between these two theorists. By recreating and juxtaposing the two paths taken by these authors up to point where they confronted each other, by explaining and illuminating the questions raised by these confrontations, we are quickly led beyond the confines of game theory. Philosophical and aesthetic preferences take on considerable importance. This thesis addresses the important role, for those involved in creative scientific work, of deep psychological orientations, even when the work involved is technical and mathematical. The role and methods of science, the importance and "proper" use of mathematics in game theory, the form and necessity of any relationship between the theory and experimentation or other empirical work; the views held by Binmore and Aumann on these matters feature prominently in a debate that, at first blush, might, in certain respects, appear to be concerned with very precise questions.

Key Words: Theory of games, History of economic thought, History of science, Psychology of scientific creativity.

INTRODUCTION

0.1 THE COMPLEXITY OF DEBATES IN SCIENCE

> In one sense the philosophers are right; to make arithmetic, as to make geometry, or to make any science, something else than pure logic is necessary. To designate this something else we have no word other than *intuition*. But how many different ideas are hidden under this same word?
>
> Jules Henri Poincaré[1]

This masters' thesis will examine two radically different ways of looking at some *foundational* issues of game theory, in particular, those of game theorists Kenneth Binmore and Robert Aumann.[2] This lack of consensus means that 'foundational questions' need to be defined broadly. Indiscriminately combining the views of both authors, we could say they encompass questions of how agents in games *behave*, what they *know* and how they *get to know* what they know. For Binmore, a mathematician turned game theorist, experimental economist, moral philosopher and governmental advisor for the design of public auctions, this debate provides another vent through which he can channel his long-held aim of transforming game theory so as to enable it to represent players in a more 'realistic' manner; that is, to model people's imperfect reasoning capacities, as well as their ability to compensate for this by learning to adapt to different environments. He is sharply critical of what has been termed the 'epistemic' approach to foundations, the research program that Aumann supports, and looks for answers concerning players' knowledge and the *formation* of their beliefs and behavior in the various branches of evolutionary games. How Binmore's radical critique of the more traditional approaches is translated into his positive work is something

[1] Poincaré (1900), p.1015.

[2] The papers we feel to be relevant for illustrating Binmore and Aumann's opposing views are, respectively, Binmore (1984, 1987b, 1988, 1990a,b,c, 1991, 1992a,b, 1993, 1996, 1997, 1998, 1999b), Binmore and Brandenburger (1990), Binmore and Herrero (1988), Binmore, Kirman and Tani (1993), Binmore and Samuelson (1992, 1994a,b,c, 1996), Binmore, Samuelson and Vaughan (1995), Binmore and Shin (1992), Gale, Binmore and Samuelson (1995), and Aumann (1974, 1976, 1985, 1987a,b,c, 1992, 1995, 1996a,b,c, 1997, 1998a,b,c, 1999a,b, 2000), Aumann and Brandenburger (1995).

that will be examined closely in this thesis. The point to which the breadth of his interests is a factor in this debate and in his published work will also be an important part of the thesis. On the other hand, Aumann, a long-standing pillar of the international community of game theorists, unwittingly 'provokes' the exchanges by trying to define precisely and analyse formally concepts such as *beliefs*, *knowledge* and *common knowledge* and uncover their links to traditional concepts in game theory. As game theorists began to realize that traditional game theory relied on unspecified or misspecified assumptions concerning players' knowledge and beliefs, Aumann began to attempt to couch traditional theory in a new formal structure that makes these hypotheses *explicit*. In a metaphorical sense, he is 'taking a step back' from the vantage point of traditional game theory to 'observe' from a distance the 'meta-conditions' that *yield* traditional game theory.

Both theorists are unique figures in game theory. Although Aumann may be most famous outside game theory for his work on the relationship between the core and general competitive equilibria,[3] inside the discipline it is difficult to find a field of research to which he has *not* contributed. Cooperative games with transferable or nontransferable utility, coalition formation, decision theory, repeated noncooperative games, games played by boundedly rational automata, the philosophy of game theory and games of incomplete information: he has touched all of these.[4] Aumann is also an important 'backstage' figure in game theory; his name often appears in the 'acknowledgments' footnote of articles on various subjects. His work spans almost five decades.

Binmore, on the other hand, came to economics through a generalisation of Arrow's famous impossibility theorem for social choice.[5] He then made fundamental contributions to linking the cooperative concept of the Nash bargaining solution to the noncooperative concept of perfect equilibrium.[6] Then, in 1984 and 1987, came two important methodolog-

[3]See Aumann (1964, 1966).

[4]Respectively (what follows is a minuscule selection; Aumann's (2000) collected works, in two volumes, has more than 1500 pages): Aumann and Dreze (1975), Aumann (1967), Aumann and Myerson (1988), Aumann and Anscombe (1963), Aumann (1981), Aumann and Sorin (1989), Aumann (1985). On the last subject, Harsanyi (1968) credits the technical Aumann (1960, 1963) as very important ingredients for his treatment of games of incomplete information.

[5]Binmore (1976).

[6]These very important contributions to the theory of bargaining are of sufficient importance for some authors, for example Aumann (1985) and van Damme (1987), to refer to the Rubinstein-Binmore model of

ical papers, one on the foundations of game theory, and the other on experimental games.[7] Both of these express very original and unorthodox views, to which Binmore remains faithful during his career. In the early 1990's, Binmore turns to philosophy, combining his experience with bargaining models and indefinately repeated games to a profound interest for political thought and social evolution.[8] These are incursions into a domain in which, again, he is an unorthodox 'outsider'. Concurrently with this, Binmore becomes a prime mover in the development of evolutionary games.[9] From his researches into the foundations of game theory, he draws the conclusion that this field of study is the most promising for the development of the theory.

We must, at the outset, apologize for the fact that much of this thesis will focus on the work of Binmore and Aumann, although in many ways they were not originators in the foundational field and their contributions represent a small fraction of a bulging literature.[10] Yet their papers and their confrontations have something of an aura to them, as though they embody the clash of two ways of looking at game theory.[11] The interest in these two particular authors stems from the fact that, along with technical contributions, their papers bear underlying 'philosophical' intentions which are well articulated and quite distinct. They are also two highly influencial figures in game theory. The main task of this thesis will be to try to unravel and express the deep significance of this debate through the writings of Binmore and Aumann. Also, interestingly, Binmore's forays into the province of philosophers and political scientists, with the principled opposition to game theoretic models it drew in some of those quarters, will give us an opportunity to see that there are

sequential bargaining. These contributions can be found in Binmore and Dasgupta (1987).

[7]Binmore (1984, 1987a). Binmore's contributions to the theory and practice of experimental economics are in Binmore (1987a, 1999a), Binmore, Morgan, Shaked and Sutton (1991), Binmore, Shaked and Sutton (1985, 1988, 1989), Binmore, Proulx, Samuelson and Swierbinski (1998) and Binmore, Swierbinski and Proulx (2001). Experimental economics is an important aspect of Binmore's *oeuvre*. See p. 104.

[8]Binmore (1993, 1998) is his large two-volume work in this field.

[9]Binmore, Samuelson and Vaughan (1995).

[10]Scant details on this literature will be given within the text, and only when absolutely necessary. Work external to this debate will be referred to only to highlight the positions of the two participants. See Battigalli and Bonanno (1999) for a hefty survey of only the 'epistemic' approach to foundations.

[11]Their debates in published form are Binmore (1996), Aumann (1996c), Binmore and Samuelson (1996) and Aumann (1996a). On the surface, they were unleashed by a paper written by Aumann (1995) on the relationship between the principle of backward induction and common knowledge of rationality, although Aumann had been in Binmore's sights for a longer time.

relatively clear limits to what distinguishes him from Aumann. There are many things that they have in common, much that still binds their ways of analysing the outside world.

Hopefully, this paper will lead its reader's thoughts away from game theory and bring him to ponder the more general subject of the psychology of creation. More precisely, this masters' thesis could be seen as pointing, in an implicit way, to certain broad regularities in methodological debates in mathematized disciplines. A very strong form of what is meant here is found in Nietzsche:

> That individual philosophical concepts are not something arbitrary, something growing up autonomously, but on the contrary grow up connected and related to one another; that, however suddenly and arbitrarily they appear to emerge in the history of thought, they none the less belong just as much to a system as do the members of the fauna of a continent: that fact is in the end also shown in the fact that the most diverse philosophers unfailingly fill out again and again a certain basic scheme of *possible* philosophies. Under an invisible spell they always trace once more the identical orbit: [...] Their thinking is in fact not so much a discovering as a recognizing, a remembering, a return and home-coming to a far-off, primordial total household of the soul out of which those concepts once emerged.[12]

We believe that through Binmore and Aumann's papers one hears echoes of past debates and ideas about the 'meaning' of mathematics and of its practice. To this end, during this thesis, we will sometimes revisit some debates in the history of mathematics. As both authors are mathematicians by training, it is not surprising that they should, consciously or not, articulate important philosophical ideas from that discipline. The analogies with past thinkers to be presented in the sequel are mostly used to illustrate and exemplify positions defended by Binmore and Aumann. Yet it also seems that these analogies sometimes reveal things that are deeper than the immediate ideas being defended; 'attitudes' or 'styles', members of Nietzsche's "primordial household of the soul". Poincaré touched on a related chord when he said:

> It is impossible to study the works of the great mathematicians, or even those of the lesser, without noticing and distinguishing two opposite tendencies, or rather

[12]Nietzsche (1886), p.32. The last sentence is, for me, the main reason for the inclusion of this passage. If we interpret its first few sentences in as non-Platonist manner as possible, the passage is a powerful reference to the existence of basic psychological 'types' of thinkers. Thus my interpretation of Nietzsche here is essentially biological (for lack of a better word).

> two entirely different kinds of minds. The one sort are above all preoccupied with logic; to read their works, one is tempted to believe they have advanced only step by step, after the manner of a Vauban who pushes on his trenches against the place besieged, leaving nothing to chance. The other sort are guided by intuition and at the first stroke make quick but sometimes precarious conquests, like bold cavalrymen of the advance guard.
>
> *The method is not imposed by the matter treated.* Though one often says of the first that they are "analysts" and calls the others "geometers", that does not prevent the one sort from remaining analysts even when they work at geometry, while the others are still geometers even when they occupy themselves with pure analysis. *It is the very nature of their mind* which makes them logicians or intuitionalists, and they cannot lay it aside when they approach a new subject.
>
> Nor is it education which has developed in them one of the two tendencies and stifled the other. The mathematician is born, not made, and it seems he is born a geometer or an analyst.[13]

Again, what should be taken from this passage, in keeping with the purpose of this thesis, is not the exact distinction that is drawn by Poincaré, but rather the fact that fundamental cleavages of a deep psychological nature can divide thinkers who, on the surface, work in seemingly narrow and related fields.

That being said, this paper is about game theory; its aims, its actors, its history. While drawing on ideas or occurrences in the history of other disciplines will be helpful, this is not intended to drown out the uniqueness of this, or any other, debate in science.[14] Neither is this clash one of 'the old against the new'. Both types of theories involved are on the cutting edge of research at the moment in game theory.

A short section, in actuality the second part of this introduction, will start by looking at only a few aspects of the two (published) direct confrontations between Binmore and Aumann; in the exchange of notes between them in a 1996 issue of *Games and Economic*

[13]Poicaré (1900), p.1012, my it.

[14]Similarly, this paper will draw, for explanatory, illustrative or other reasons, on remarks from many mathematicians who were involved in the debate on the foundations of mathematics. But why should Hilbert, Brouwer or others only be called in to present their views on philosophical controversies? They were all prominent mathematicians, whose views on their work cannot and should not be forced into the single mould of the debate on the foundations of mathematics. That debate is of course of great importance, for mathematics in general and for the work of the authors involved, and many of the quotes to be heard in this thesis may well have it as a backround. However, one should not perceive in all of their interventions in the sequel that automatic links are being woven between two debates on foundations, that of mathematics and that of game theory.

Behavior and in an article by Binmore and Larry Samuelson accompanied by a reply by Aumann prepared for an International Economic Association conference on rationality held in Turin, Italy.[15] Its aim is to illustrate the alacrious tone of the debate, its *form*, while a more precise elaboration of the *content* will come later. In a way, this reproduces how a reader unversed in the details of the foundational literature is introduced to this confrontation; one is first struck by the apparent ambiance of confusion and the flashes of rhetoric. Then, more slowly, as one becomes familiar with the authors' other contributions, matters become clearer.[16] Once the paths that both authors have tread in and around the foundational litterature have been presented, we will come back to explore the exact issues evoked in the debates of 1995-96. However, the chosen sequence of presentation has some advantages: it sets the stage swiftly for relatively self-contained expositions of the two authors' respective works, this in the perspective of explaining the origins of their clashes; it presents their respective views in a format where their differences, one could say their profound incompatibility, are clearly visible;[17] and last but certainly not least, these exchanges being highly entertaining, they get our ball rolling in an ageeable manner.

Once the debates have been presented, in the first chapter we turn to a more thorough examination of Aumann's views. This starts in 1976 (section 1.1) when he was the first author to give a formal definition of the notion of *common knowledge*.[18] The tools he developped at that time are then used to attack foundational questions in a paper in 1987 dealing with the links between correlated equilibria and common knowledge of rationality (section 1.2).[19] After examining an example provided by Aumann which is adapted from earlier papers but meant to study irrationality in section 1.3, we look, in section 1.4, at what could be considered the apex of Aumann's efforts; his uncovering of the foundations of the Nash equilibrium.[20] The structure of the chapter on Aumann is rather chronological; his contributions are treated as 'signposts' on a 'path'. Convenience accounts for much

[15]Respectively, Binmore (1996), Aumann (1996c), Binmore and Samuelson (1996) and Aumann (1996a).

[16]Unfortunately, this tends to last only until one decides to read these papers again, as they are quite rich in subtle details.

[17]This is particularly important given that this paper is as interested in the contrasts between the authors' positions as in these latter themselves.

[18]Aumann (1976).

[19]Aumann (1987a). What is meant by these terms will be made clear.

[20]Respectively, Aumann (1992) and Aumann and Brandenburger (1995).

in this choice, even if the sequence of his contributions is far from irrelevant, and past papers provide stepping-stones to further developments. Discussions of individual papers provide springboards to subjects that are important to the whole of Aumann's carreer in the foundations of game theory.

The second chapter looks at Binmore. Here we face the challenge of faithfully representing the work of a man whose interests are widely distributed. After a short discussion, in section 2.1, of the *style* of Binmore's writings, we examine, in section 2.2, his important critical work in the foundations of game theory. His work in foundations proper is mostly critical; in particular, he objects to how rationality and learning are treated in traditional game theory. Section 2.3 exposes some of Binmore's views on mathematics as applied to game theory. In section 2.4, we sketch some of Binmore's tastes for philosophy. We examine two ways in which references to philosophy make it into his work. Section 2.5 looks at some types of questions, and types of mathematics, of which Binmore is particularly fond; what we call 'paradoxical' questions and mathematics. In the final section of the chapter, we look at Binmore's positive work, mostly in evolutionary games. This is done with an eye to exploring how Binmore's many methodological beliefs, presented in the preceding sections, mold and shine through the models he chooses to construct.

The third chapter comes back to the debates on backward induction. Here, Binmore and Aumann quarrel about the validity of Aumann's (1995) theorem that common knowledge of rationality implies backward induction in games of perfect information. This account of the debate draws heavily on the portraits of both theorists painted in the first two chapters.

0.2 LOUD AND RUDE: BINMORE AND AUMANN'S EXCHANGES

> Now, however, we come to the foundations of game theory in which gamesters comment on their own crazy activities and propose further apparently bizarre projects for the future. When they get into this topic, their conversations with each other have a tendency to resemble, not so much the famous interchange between Alice and Humpty Dumpty in which the later argued that words should mean whatever he found it expedient for them to mean, as the interchange that would have taken place if Alice had been similarly minded.
>
> K. Binmore, A. Kirman and P. Tani[21]

An interesting aspect of this debate on backward induction and common knowledge of rationality is that at some points both authors seem unsure of the necessity of its occurrence. Binmore, in particular, is left wondering how all this began -especially since he thought he had resolved most of these issues ten years before:

> Binmore (1987) explains why backward induction in games cannot be justified by assuming only that there is common knowledge of rationality before the game begins. In writing once more on the myth that it can, we feel like Sisyphus rolling his rock to the top of the hill again. However, it is necessary to respond to Aumann's *attempt to revitalize a fallacy which game theorists must abandon if their discipline is to continue to be taken seriously.*[22]

But his doubts run even deeper. Indeed, how could anybody even *contemplate* dreaming up, much less publishing, such views?

> Without intending any disrespect to the authors, *I believe that there is little of genuine significance to be learned from any of the literature that applies various formal methods to backward induction problems* - even when the authors find their way to conclusions that I believe to be correct. It seems to me that all the analytical issues relating to backward induction lie entirely on the surface. *Inventing fancy formalisms serves only to confuse matters.*"[23]

To reinforce similar points, even the proverbial authority of proverbs is invoked. "The proverb warns us against closing the stable door after the horse has bolted. But an even worse mistake is to bolt the stable door while the horse is still in the yard."[24]

[21] Binmore, Kirman and Tani (1993), p.14.
[22] Binmore and Samuelson (1996), p.111, my it.
[23] Binmore (1997), p.2, my it.
[24] Binmore and Samuelson (1996), p.111.

At times, it seems that Aumann is quite confused by what Binmore has written: "we found it difficult to follow Binmore's (1996) [...] critique of Aumann (1995)".[25] And on occasion, things get even worse: "Clearly, this argument is absurd."[26] But, although the battle is pitched and bloody, Aumann still manages to show generosity towards his opponent. "Though Binmore makes an unconvincing case, the malaise that he evinces is not *totally* groundless."[27] Or elsewhere: "These positions are argued by Binmore and Samuelson at great length, with fervour and flair *and even some reasoned arguments*[!!!]." Happily, however, on these rare occasions, "*I have promoted many of them myself* - some in the very paper they are attacking (Aumann (1995)) and others previously".[28]

One cannot help but notice a debate between eminent specialists, dealing with very deep aspects of an abstract mathematical theory, in which one of the debators stops to demonstrate to the other some of the A's, the B's and the C's of English grammar.

> Another error of Binmore is his failure to distinguish between the subjunctive and indicative moods. [...] Our intention could not have been made clearer. The indicative mood is heavily stressed: "are actually reached." "Are"; not "would", not "were to be". For the case that somebody might still misunderstand, the word *"actually"* is thrown in. And if, by some stretch of the imagination, *somebody* might *still* misunderstand, our very next sentence (not cited in Binmore) would surely clear thing up: "Under common knowledge of rationality (CKR), vertices off the backward induction path *cannot* be reached; and when CKR does not obtain, the results do not apply."[29]

This is followed by "Binmore is 180° off the mark."[30] And finally, the exasperated *finale*: "There is nothing more to say; I cannot defend myself against an attack that has no substance - is all smoke and no fire."[31]

Leave it to a debate on the 'foundations' of a theory to elicit exchanges in which it seems that anything and everything opposes the participants; in this case, even the form of the opponent's argument leads to quarrel. Aumann admonishes Binmore for the casual,

[25]Aumann (1996c), p.138.
[26]Aumann (1996c), p.143.
[27]Aumann (1996c), p.141, my it.
[28]Aumann (1996a), p.130, my it.
[29]Aumann (1996c), p.140.
[30]Aumann (1996c), p.140.
[31]Aumann (1996a), p.131.

non-formal style of his arguments:

> Before responding to this [Binmore's argument] substantively, we make a methodological point. It is difficult to evaluate the validity of this kind of contorted reasoning using verbal tools only. *That is a function of mathematical formalisms.* In a formal model the conclusions are derived from definitions and assumptions [, but] with informal, verbal reasoning as complex as the above, one never knows for sure whether the argument is sound. One can argue until one is blue in the face, without convincing one another, because there is no criterion for deciding the soundness of an informal argument.[32]

Unperturbed, Binmore accuses Aumann of confusing matters by "inventing fancy formalisms":[33]

> At the root, the difficulties are philosophical in that they arise from the manner in which the nature of the problems to be resolved is perceived. Usually the problems are framed in abstract mathematical terms and then attacked à la Bourbaki. Such a definition-axiom-theorem-proof format closes the mind, the aim being to exclude irrelevancies so that attention can be focused on matters of genuine importance. My contention is that the conventional approach misses this aim, not only by leaving unformalized factors which matter, but also by introducing formal requirements that cannot be defended operationally exept in terms of mathematical elegance or simplicity.[34]

The examination of these differences in methodology, where the two authors differ fundamentally in their ideas on the role of assumptions in formal models, the goals of game theory and the importance of formalization, will illuminate many of the details of this debate and will figure prominently here.

What should become clear before the end of this paper is that we are not facing a dry wrangle between specialists. The authors' differences stem just as much from the 'gut' and the imagination as from the brain and the intellect. Thus to understand this debate, one needs to look beyond the papers involved to bring to life more basic and fundamentally imprecise drives, intuitions, and 'visions'. This paper attempts to present an interpretation of these questions; *truth*, if it exists in these matters, it not something that will be attained here. In our case, let the 'story' told here be honestly intentionned, thought-provoking and a good read. This paper is the fruit of the intense interest spawned by the reading of this

[32]Aumann (1996c), p.142, my it.
[33]Binmore (1997), p.2.
[34]Binmore (1987b), p.180.

debate and the two authors' related papers. It should therefore be read not so much as an complete 'explanation' of these but rather as an exhortation to go out and read them accompanied by a broad suggested perspective with which to do so. For a paper such as this could never be a substitute to reading the original articles. As Courant and Robbins (1941) have remarked regarding mathematics:

> Some splendid books on biography and history and some provocative popular writings [about mathematics and mathematicians] have stimulated the latent general interest. But knowledge cannot be attained by indirect means alone. Understanding of mathematics cannot be transmitted by painless entertainment any more that education in music can be brought by the most brilliant journalism to those who never have listened intensively. Actual contact with the *content* of living mathematics is necessary.[35]

[35] Courant and Robbins (1941), p.v.

CHAPTER I

AUMANN: ANTEAS' VIEW OF STRATEGIC INTERACTION

1.1 FORMALIZING COMMON KNOWLEDGE

One of Aumann's most important contributions to foundational questions is that of having formalized the notion of common knowledge in a paper published in 1976 and titled *Agreeing to Disagree*. Philosopher David Lewis had, with seven years' priority,[36] given to the notion of common knowledge the verbal form with which it is still rendered today. That is, a state of affairs is common knowledge between certain people if all know it, all know that all know it, all know that all know that all know it, and so on to infinity. In *Agreeing to Disagree*, Aumann gives the concept a formal definition, and then shows that this latter definition is equivalent to the verbal one.

Nominally, the purpose of the paper was not to define common knowledge. That notion was required, however, for stating the three-page paper's main theorem. The latter being: *if two people have the same priors, and their posteriors for an event A are common knowledge, then the posteriors are equal.* "In brief, people with the same priors *cannot agree to disagree.*"[37] Roughly illustrated, the theorem states something like this. Suppose two

[36] In Lewis (1969). Aumann has said that he was unaware of Lewis's book until several years after his 1976 article. He adds that "there is of course no question of Lewis's priority; even the term "independent discovery" does not apply. Ideas percolate through the scientific world in many ways, including seminars and dinner conversations, and it is quite possible that some of Lewis's ideas filtered through to me without attribution, in ways of which I was not conciously aware. Needless to say, Lewis himself did not work in a vacuum; people were groping towards the idea of common knowledge during much of the sixties. But unquestionably, the first coherent formulation of this concept, as well as the recognition of its importance (a vital part of the scientist's task!), must be credited to Lewis." (Aumann (2000), vol. I, p.555)

[37] Aumann (1976), p.1236.

people place the same prior probability on the occurrence of some event, A.[38] Then put them in two different rooms from which they can observe the occurrence or non-occurrence of A. With the help of this new information, they update their priors by traditional Bayesian updating. Once they exit their rooms, a way to make these newly-formed posteriors common knowledge must be found, for example "if 1 and 2 tell each other their posteriors and *trust* each other".[39] In that case their posterior probabilities will be adjusted until they are equal. With a somewhat characteristic[40] lightly self-depreciating tone, Aumann says of this path-breaking paper that "we publish this observation with some diffidence, since once one has the appropriate framework, it is mathematically trivial."[41] He recognizes, however, that in the paper, the "key notion is that of common knowledge".[42]

We now introduce part of the basic notation for Aumann's contributions to foundational questions, which is in fact that of probability theory. A set Ω, usually called *the universal event*, has elements ω, called *states of the world*. These give a complete specification of all variables of importance to a player and about which he is uncertain. If a person is trying to decide whether or not to crack an egg that may be rotten,[43] there are only two *relevant* and mutually exclusive states of the world; the egg is rotten or the egg is not rotten.[44] Thus when a player is informed that a certain state of the world obtains,

[38] In an implicit way, this is common knowledge between the players. That is, it would be specified in all states of the world, since it is an assumption of the model. On this see later, pp.27-8.

[39] Aumann(1976), p.1236, my it. The word *trust* is used in a very strong sense here. The trust must be absolute, the players cannot suspect that (or suspect that the other may suspect, or that the other suspects that he suspects the other suspects, etc) in any way that the revelation of posteriors is in any way strategically motivated. In a way, then, this 'trust' is just as absolute as the knowledge it permits players to obtain. In its context, the comment is not that remarkable; after all, the paper is published in a mathematical journal by an author with considerable experience in cooperative game theory, therefore not averse to agreements of a 'binding' kind. While it may be anachronistic to do so, the contrast with Binmore is nonetheless interesting to note. The latter has devoted much of his publications (in particular the two large volumes of a book on social ethics and conventions (Binmore (1993, 1998)) to questions concerning the evolution of social behavior, such as that of 'trust', in strategic frameworks.

[40] Similar statements are to be found in Aumann (1987a) and during some of Aumann's presentations of preliminary versions of Aumann (1995) (see Binmore (1992)).

[41] Aumann (1976), p.1236.

[42] Aumann (1976), p.1236.

[43] One of the examples in Savage (1954).

[44] The states are mutually exclusive in what the individual is uncertain about. The dog on the kitchen floor watching his master crack an egg can stay where he is. That is unless the decision-maker holds some sort of super-natural belief concerning canines' influence over the freshness of eggs. In that case, the precise

he knows all that he *wants* to know; in a game this includes the other players' strategies, what they know, what they believe, etc. Within this framework, it is also assumed that *anything* that could possibly happen in the 'world' is somehow already specified and accounted for.[45] A player with a prior probability distribution on Ω knows all that *could* happen, even if he will rarely know the precise state of the world obtaining.[46] However players are not directly informed of ω, instead they are told that a certain *event* has occurred, this being a set of states of the world. The set of all events a player i can distinguish from one another forms a *partition*, $\wp^i$, of Ω, called player i's *information partition.*[47] This means that the only way a player i can know directly that the specific state of the world $\omega \in$ E has obtained is if ω is the sole member of the event E and $E \in \wp^i$.

Set $\mathbf{P}^j(\omega)$ as the element of $\wp^j$ that contains ω. When an event includes $\mathbf{P}^j(\omega)$, Aumann says that player j *knows* that event. To say that $E \supset \mathbf{P}^j(\omega)$ means that all states of the world contained in $\mathbf{P}^j(\omega)$ are also contained in E. Thus, when E occurs and $\omega \in E$ is the true state of the world, the player is informed of $\mathbf{P}^j(\omega)$ and *knows* E in the sense that he knows that the true state of the world is one of the states of the world in E that is also in $\mathbf{P}^j(\omega)$.

We can now give Aumann's 1976 definition of common knowledge. *For* n *players and given a state of the world* $\omega \in \Omega$*, an event* E *is said to be* common knowledge *at* ω *if* E *includes that member of* $\bigwedge_{i=1}^{n} \wp^i$ *that contains* ω*.*[48] $\bigwedge_{i=1}^{n} \wp^i$ represents the meet of the n partitions, their finest common coarsening. The meet of the n information partitions is, in a sense, 'the best we can do' to regroup the players' different capacities to acquire

effect of the dog on the egg would need to be specified in each state of the world.

[45]For Savage (1954), this points to "the practical necessity of confining attention to, or isolating, *relatively simple situations* in almost all applications of the theory of decision developed in this book." (pp.82-3, my it.) These "relatively simple situations" he names "small worlds". Binmore strongly objects to this 'closed' aspect of statistical decision theory, considering that it severely impairs that theory's potential for application to game theory (see Binmore (1991a), titled *De-bayesing Game Theory*, and later p.55).

[46]Even when a probability measure takes the value of zero it is still defined. Intuitively, this means that if you asked a player about an event to which he ascribed a subjective probability of zero, he would not answer you with a blank 'I have no idea what you are talking about' expression but would say 'I consider that event to be impossible.'

[47]A partition of a set B is defined as a collection of mutually disjoint nonempty subsets of B whose union equals B.

[48]Aumann (1976), p.116. There it is defined for two players.

information. More formally,[49] define a function $f : \Omega \longrightarrow 2^{\Omega}$ such that $f(\omega) = \bigcup_{i=1}^{n} \mathbf{P}^i(\omega)$. To every $\omega \in \Omega$, f associates the union of those elements of players' information partitions that contain ω. The image $\mathbf{I}$ of f is a collection of sets. Define function $g : \mathbf{I} \longrightarrow 2^{\Omega}$ such that:

$$g(I) = \begin{cases} I \bigcup (\bigcup_{k=1}^{m} J_k) : & if I \bigcap \mathbf{I}/I \neq \emptyset, where\ J_k\ is\ as\ subset\ of\ \mathbf{I}\ such\ that \\ & I \bigcap \mathbf{I}/I \neq \emptyset \\ I : & if I \bigcap \mathbf{I}/I = \emptyset \end{cases}$$

The meet of the n partitions $\wp^i$ is the image of the composite function $g \circ f$. Clearly, it covers all of Ω. And it is a partition of Ω since it is composed of unions of partitions. It is thus, generally, a 'coarser' partition.[50]

Let us see how Aumann's definition can express the infinite regress usually associated with common knowledge. Suppose that E includes the member of $\bigwedge_{i=1}^{n} \wp^i$ that contains ω. Thus for all players j we have $\omega \in E \supset \mathbf{P}^j(\omega)$, the latter being a subset of an element of $\bigcap_{i=1}^{n} \wp^i$. In our case, when any ω in E obtains, all players know E, since that event contains a set of states which forms an element of each of their information partitions. Since ω is an *exhaustive* description of the world, the fact that when it obtains all players know E must be included in the specification of ω. This must be true for all ω in the element of $\bigcap_{i=1}^{n} \wp^i$ that is contained in E, since all these states of the world are indistinguishable from one another for the players.[51] Occurrence of ω now implies that all players know E *and* that all know that all know E, this for all ω in that element of $\bigcap_{i=1}^{n} \wp^i$ that is contained in E. Again, since ω is a complete description of the world, this latter knowledge must be specified by it. Repeating the previous exercise gives us another level of knowledge about other players' knowledge. It can be pursued forever, which is what the verbal definition of

[49]The main texts in the epistemic foundations of game theory never really define what the *meet* of a partition is. I have figured out what it means, but to define it took me longer. I give here what I came up with (I have not gone to look at set theory textbook because, firstly, none of the basic textbooks deal with the subject and, secondly, this was much more fun).

[50]Here are two simple examples. Say $\Omega = \{1,2,3,4,5,6\}$, $\wp^1 = \{\{1\},\{2,3\},\{4,6\},\{5\}\}$ and $\wp^2 = \{\{1,2\},\{3\},\{4,5\},\{6\}\}$. Then $\bigwedge_{i=1,2} \wp^i = \{\{1,2,3\},\{4,5,6\}\}$. And if $\wp^1 = \{\{1,4\},\{2,3\},\{5\},\{6\}\}$ and $\wp^2 = \{\{1,2\},\{3\},\{4,5\},\{6\}\}$, then $\bigwedge_{i=1,2} \wp^i = \{\{1,2,3,4,5\},\{6\}\}$. Computing f and g in these examples gives a good idea of the finest common coarsening's 'purpose'.

[51]Actually, they are indistinguishable for *some* players, and for others they can be 'broken down' into smaller events. This does not matter, however (See the definition of the meet of a partition).

common knowledge leads to.[52]

With this definition in hand, Aumann proves his paper's main result in four lines. As he claims, the proof itself is indeed 'simple', but then again Aumann has said elsewhere that "the most lasting and important mathematical ideas are often also the simplest."[53] And again, at another time: "In my opinion the important mathematics is not necessarily the most complex, involved gymnastics that one can do. The really important pieces of mathematics are those that can be reduced to at most a few pages."[54] Or, as the mathematician Felix Klein once remarked about a proof of Hilbert's, it is "wholly *simple* and, therefore, logically compelling."[55] There is much that is rhetorical in the modesty that leads Aumann to trivialize his results. He does refer to his apparatus as intuitively "not quite obvious",[56] and proposes that it could be "of some interest in areas in which people's beliefs about each other's beliefs are of importance, such as game theory and the economics of information".[57]

In this exercise, Aumann has produced a prime example to illustrate the power of mathematical methods; he takes an unwieldy concept that seemed intrinsic to game theory, yet remaining somewhat exterior to it by dint of its being stated in a language different from that of the rest of the theory, and turns it into a simple and compact expression. The links between *this* definition of common knowledge and traditional game theory can be formally examined, and this is the door that Aumann opens. There seems to be little doubt that this would be regarded as elegant by mathematicians. The simplifying aspect of some contributions to mathematics, that of translating a concept from a language in which it was cumbersome[58] into one in which it can be elegantly and simply stated is an oft-cited factor in discussions surrounding results generally agreed to be esthetically pleasing

[52] Aumann's demonstration is entirely different. In fact, this is not a formal demonstration, since it uses the things that are 'inside' states of the world, which have no formal content *within* the model. States of the world are assumptions of the model, and knowldege refers to events, sets of states. What is inside states does not formally exist. On this see later, pp. 27-8. This argument is only meant as an illustration.

[53] Aumann (1985), p.42.

[54] Aumann (1987b), p.135.

[55] In Reid (1970), p.33, my it.

[56] Aumann (1976), p.1236.

[57] Aumann (1976), p.1236.

[58] Be it a literary one or another mathematical symbolism.

to mathematicians. This drive to reach out, to forge links between concepts, is also very important to Aumann personally. Indeed, in the preface to his collected writings, he writes that "the idea of relationship pervades my work".[59] The latter has dealt with "how ideas - strands of thought- are inextricably tied together".[60] "When you look closely at one scientific idea, you find it hitched to all others. It is these hitches that I have tried to study."[61] The form and content of his mathematical activity are fused so that, in some way, the identities of the theories he "hitches" to others loose a bit of their importance. "I love correlated equilibria and I also love subgame equilibria; I have written papers, which I hope the world will enjoy, on both subjects."[62] One feels through his papers the need to fully explore the concepts advanced, to draw the maximum number of implications, *for its own sake.* Axioms or hypotheses whose interpretation he is not fully comfortable with would not cause him to halt his deductive work; he makes sure the readers fully *understand* the axioms and what he thinks they imply, then he draws as many of the logical consequences as he can. The moral troubles he leaves to others.

Aumann has shown the two definitions of common knowledge to be equivalent, yet it seems that they strike the mind differently. In the formal definition, everything is contained and motionless. In a Venn-diagramm type of way, we picture a huge but closed set that represents *everything* and *anything* we want the 'world' to be; time and space are arbitrary attributes that we give to the elements of this set in order that it be capable of representing particular problems. We can then imagine this 'world' as completely covered with a series of layers of other sets, each layer representing a certain player's information partition on Ω. If all these 'layers' overlap in a particular way on top of an area of the 'world', then when a state of the world in that area obtains, it is common knowledge among all players that a state of the world in that area has occured. The verbal definition, on the other hand, is expressed as intertwined sequences of players' thoughts about their own thoughts and the others' thoughts about more thoughts and so on. The effect on the mind's eye is wholly different. Even though the sequence can and will never stop, one experiences

[59] Aumann (2000), vol I, p. x.

[60] Aumann (2000), vil. I, p. x. The image used in this quote comes from an analogy with knot theory, the subject of Aumann's doctoral thesis.

[61] Aumann (2000), p. x.

[62] Aumann (1998c), p.202.

a curious kind of anticipation; the fact of 'not seeing the end' leaving us some room to imagine the impossible. In this case, since the verbal definition specifies that some entity is *thinking*, it reminds us of our 'real', less than rigorous, thought processes. Thus we can be left wondering:[63] what if one of the players gets tripped up in his thoughts at the next step? Such thoughts are fleeting (or should be) and barely at the concious level, but the point of these remarks is simply that the formal definition changes not only the language used to describe common knowledge, but our way[64] of 'seeing', or more aptly, 'feeling' common knowledge.

1.2 CORRELATED EQUILIBRIA

> Above all, there was a *practical* wisdom about their doctrine, a *responsible* long-run view and a manly *tone* that contrasts favorably with modern *hysterics*.
> Joseph A. Schumpeter[65]

Aumann (1987a), a paper titled *Correlated Equilibrium as an Expression of Bayesian Rationality*, is another fundamental stepping stone in his epistemic contributions. It is in this paper that the 'partition' theory of knowledge reveals its potential for tackling epistemic questions in game theory. His approach has breadth, it brings together insights from different fields. Aumann claims that its "roots extend back to Radner (1968, 1972), to Harsanyi (1967-68), to von Neumann and Morgenstern, to formal epistemologists such as Kripke and Hintikka, and to probabilists such as Kolmogorov".[66] In this paper, Aumann clearly formulates a view of foundational questions and a strategy to resolve the problems occurring within them which will characterize much of his future work in this area. Aumann considers that this paper "puts together the fundamental notions"[67] of the 1976 paper on common knowledge and a 1974 paper intitled *Subjectivity and Correlation in Randomized Strategies*. In that paper, Aumann explored randomization of strategy choices by two approaches;

[63]Even if the verbal definition of common knowledge is still only a *definition*, in itself it implies nothing more than what it states. The fact that these 'wonderings' are 'wrong' is not the point here.
[64]At least mine. Obviously, the mental pictures I have given here are purely introspective.
[65]Schumpeter (1943), p.76, my it.
[66]Aumann (1992), p.215.
[67]Aumann (2000), vol. I, p.556.

through differences in subjective probability evaluations and through correlation, that is, through the use of a random device common to some or all of the players.

Aumann (1987a) begins by asking a fairly common question about Nash equilibria, that, although not of an epistemic nature, quickly leads to questions relating to players' knowledge and beliefs. He explains that "though at first the definition [of the Nash equilibrium] seems simple and natural enough, a little reflection leads to some puzzlement as to why and under what conditions the players in an n-player game might be expected to play such an equilibrium."[68] Given that such questions are usually answered by responding that all players believe that others are playing their equilibrium strategies and thus must choose their own best-response (equilibrium) strategies, Aumann asks: "Now why should any player assume that the other players will play their components of such an n-tuple, and indeed why should they?"[69] Aumann develops a strategy for coming to grips with these types of problems; the 'partition' theory of knowledge provides him with a tool with which to examine game theory's solution concepts in order to pin down precisely *what conditions* on agents' knowledge permit various solutions to games. It is true foundational work in that he aims to *clarify* much of traditional game theory by rendering explicit a series of assumptions implicitly made about agent's knowledge. His is not a moralizing approach, he does not aim to *justify* the basic assumptions used in game theory but to determine *which* ones are indeed *necessary*. The point of his method is therefore to look *under* game theory's main solution concepts, taking them as given, to find out what holds them up. In the process, individual players as centers of thought are somewhat overshadowed by properties of and relations between information sets. In a sense, by digging 'deeper and deeper' we observe the players from 'further and further' away.

Aumann's main theorem in this paper is that *if it is common knowledge that all players are bayesian rational, then the strategy distributions that result from play correspond to correlated equilibria*. Let us first look at correlated strategies and equilibria. For an n-person game G in strategic form with S^i being player i's $(i = 1, ..., n)$ pure strategy set, $h^i(s)$ her payoff for an n-tuple $s = (s^1, ..., s^n)$ of strategies and setting $S = S^1 ... \times S^n$ and

[68]Aumann (1987a), p.1.

[69]Aumann (1987a), p.1.

$h = (h^1, ..., h^n)$, a *correlated strategy n-tuple* in G is defined as a function $f : \Gamma \longrightarrow S$, where Γ is a finite probability space. This means f is "a random variable whose values are n-tuples of actions."[70] A correlated strategy replaces the *individual* distinct random events each player observes in a traditional mixed strategy with a random event from the probability space Γ that can be commonly observed by all or some of the players.[71] Aumann imagines the process through which players are informed of the results of the random event thus: "chance chooses an element γ of Γ, then suggests to each player i that he take action $f^i(\gamma)$. [...] If all players follow the suggestions, the correlated strategy n-tuple f results."[72] A *correlated equilibrium* obtains when no player has any incentive to deviate from his 'suggested' pure strategy, given that others will follow theirs. Formally, a correlated equilibrium is given by the condition that:

$$Eh^i[f(\gamma)] \geq Eh^i[f^{-i}(\gamma), g^i(\gamma)]$$

where g^i is a function from Γ to S^i and $[f^{-i}(\gamma), g^i(\gamma)] = [f^1, ..., f^{i-1}, g^i, f^{i+1}, ..., f^n]$.

One can define the *distribution* of the correlated strategy n-tuple f as probabilities attached to all strategy n-tuples s that form the range of f. These are the probabilities that a player i considers to compute his expected gain from respecting or deviating from a correlated strategy. That is, informed that γ has occured, player i only considers as possible those states of the world where the other players play their components in $f(\gamma)$, with probabilities given by the distribution of f. A *correlated equilibrium distribution* is the distribution of a correlated equilibrium strategy. Thus here are the definitions of the terms used in one of the two parts of the statement of Aumann's (1987a) theorem, the 'then' part. The 'if' part is that it is common knowledge that players are bayesian rational. Let us turn to that now.

Aumann defines bayesian rationality in terms of the 'information partition' model exposed in the last section. For all players i, he introduces a partition $\wp^i$ on a finite state

[70] Aumann (1987a), p.3. Aumann uses the term actions to denote strategies in the usual sense.

[71] As Aumann shows, correlated strategies are *more general* that mixed strategies. This is seen by setting $\Gamma = \Gamma^1 \times \Gamma^2 \times ... \times \Gamma^n$, where Γ^i is the random event player i observes for his mixed strategy. With correlated strategies, one can obtain any level of dependence/independence between the random events on which players condition their decisions. That is, mixed strategies are special cases of correlated strategies.

[72] Aumann (1987a), p.4.

space Ω and a probability measure p^i on Ω; players' prior probability distribution on all possible states of the world. What was not present in Aumann (1976) and is required to turn a subjective probability problem into a game-theoretic one is a function $\mathbf{s}^i(\omega) : \Omega \longrightarrow S^i$, an "exogenously specified"[73] function that gives player i's strategy at all states of the world. This function is in fact implied by the definition of a state of the world; i.e., as a complete description of all that obtains. Further set $\mathbf{s}(\omega) = (\mathbf{s}^1(\omega), ..., \mathbf{s}^n(\omega))$. A player is defined as:

> *Bayes rational at* ω if his expected payoff given his information, $E(h^i(\mathbf{s})|\wp^i)(\omega)$, is at least as great as the amount $E(h^i(\mathbf{s}^{-i}, s^i)|\wp^i)(\omega)$ that it would have been had he chosen an action s^i other than the action $\mathbf{s}^i(\omega)$ that he did in fact choose; in brief, if he chooses an action that maximizes his payoff given his information.[74]

Note that the uncertainty in this model is due only to the uncertainty of the players concerning the ω that actually obtains, not with regards to the information contained in ω. Thus if Bayes rationality is common knowledge, players cannot entertain any doubts about the rationality of others. Call the event 'all players are Bayes rational' event E. If E is common knowledge, then all ω in E, that are also in the meet of the players' information partitions,[75] contain the information 'it is common knowledge that all players are Bayes rational'. A player's uncertainty is due solely to not knowing what ω within his information partition obtains, his uncertainty concerning what the other players believe about the true ω, what they believe about his beliefs, and so on. But the states of the world he will consider in order to make his choice are the ones in which there is common knowledge of rationality.

As in Aumann (1976), the proof of Aumann's (1987a) main theorem is not long, straightforward and relatively simple. But yet again, the true difficulties, and innovations, lie in the intuition behind the formal approach. Aumann anticipated that some people would have difficulty coming to grips with a game theoretic model in which strategies for specific states of the world are *exogenously* given, that is, specified in the description of the state. Indeed, "this sounds like a restriction on the decision maker's freedom of action; at a given state ω, it is as if the model forces him to take the decision dictated by ω."[76] Such discomfort fails to appreciate the change of perspective which Aumann's treatment of

[73] Aumann (1987a), p.7.
[74] Aumann (1987a), p.7.
[75] That is, the states of the world that the players will think of as likely.
[76] Aumann (1987a), p.8.

epistemic issues imposes on the theory:

> The model describes the point of view of the outside observer. Such an observer has no a priori knowledge of what the players will choose; for him, the choices of the players are part of the description of the states of the world. This does not mean that the players cannot choose whatever they want, but only that the observer will not in general know what they want.[77]

Or, put differently, in a sentence Binmore incessantly quotes to characterize Aumann's position: "The point of view of this model is not normative; it is not meant to advise the players what to do. *The players do whatever they do*; their strategies are taken as given."[78] Thus states of the world could be interpreted as an exhaustive list of possible outcomes built by the theorist in order to map his ignorance of what happens *during* play. Assumptions such as Bayes rationality, and further, of common knowledge of Bayes rationality, serve to narrow down the list, making the analyst's job more tractable. These assumptions provide consistency requirements on the interaction of players' beliefs, and thus actions. They eliminate certain outcomes which an outside observer, *after* the dust kicked up by the play of the game has cleared, could not observe. But it says nothing about the process through which the 'consistent' outcome could come about. In a sense, the ouside observer is not only 'too far' to see anything, he is also 'too late'.

Aumann gives another, related, interpretation of why a player's strategy is included in the descriptions of states of the world. "That we include his action as part of the description of the state of the world is only a convenient way of expressing the fact that the other players do not know which action he wishes to choose."[79] In this case, players' probability distributions on states of the world are seen as conjectures by opponents regarding the possible realizations of pure strategies of a given player.

Aumann was right to foresee that the question of a 'lack of free will' in his approach would create problems for some on an interpretative level.[80] Indeed, almost ten years after the publication of Aumann (1987a), in a paper written jointly with Adam Brandenburger,[81]

[77] Aumann (1987a), p.8.
[78] Aumann (1992), p.215, my it.
[79] Aumann (1987a), pp.8-9.
[80] Binmore being one of them. He writes that "Aumann explicitely denies his players "conscious choice"." (Binmore (1991a), p.8)
[81] Aumann and Brandenburger (1995).

he is still defending his method against such criticism:

> It has been objected that since the players' actions are determined by the state, they have no freedom of choice. But this is a misunderstanding. Each player may do whatever he wants. It is simply that whatever he does do is part of the description of the state. If he wishes to do something else, he is heartily welcome to it: but then the state is different.[82]

Aumann's (1987a) results require the important assumption, also present in Aumann (1976), that players all have *common priors*. That is, $p^1 = p^2 = ... = p^n = p$. This does not mean that if player i is informed of $\mathbf{P}^i(\omega)\in\wp^i$ and player j of $\mathbf{P}^j(\omega)\in\wp^j$, that $p(\omega;\, \mathbf{P}^i(\omega))=p(\omega;\, \mathbf{P}^j(\omega))$, since the elements of their information partitions which contain the true state of the world ω will in general encompass different states other than ω. Once i is informed of $\mathbf{P}^i(\omega)$, he redistributes his beliefs about the states in $\mathbf{P}^i(\omega)$ according to p. However, the common prior assumption states that were the players to be informed of the same subset of Ω that would belong to both their information partitions, then their probabilities for all the states of the world in that subset conditionned on their partition structure would have to be equal. Aumann terms this the *Harsanyi Doctrine*, in honor of the latter's being the first to defend the use of the assumption in his series of three papers laying the groundwork for the analysis of games of incomplete information.[83] On an interpretative level, he justifies it on the basis that assuming that players have common prior probabilities allows the analysis to focus solely on differences in posteriors that are due to differences in *information* relevant to the game, not to differences in initial beliefs.[84]

---

[82] Aumann and Brandenburger (1995), p.1174.

[83] Harsanyi (1968). On a more humorous note, Binmore writes in a footnote to his game theory textbook that "although Robert Aumann hung this label on him, John Harsanyi is not in the least doctrinaire."(Binmore (1992b), p.476)

[84] In section 14 of Harsanyi (1968c), the author presents three arguments in favor of the assumption of common priors. Firstly, since the model with common priors is easier to solve, "player j can greatly simplify his analysis of the game situation by using n mutually consistent distributions [, that is] unless he feels that the information he has about the game situation is incompatible with this assumption." (p.494) This "rational laziness" justification is not relevant to Aumann (1987a). Secondly, players (and analysts) could simulate the effect of divergent priors by building a model with common priors, but with properly adjusted information endowments. This justification is also used by Aumann. Thirdly, if we imagine players as individuals randomly drawn from certain populations by some "social process", "instead of trying to estimate each player's subjective probability distribution [...] separately, player j should rather try to estimate directly the objective probability distribution [...] governing the random social process in question." (pp.494-5) In this case players are trying to estimate the "average" initial beliefs of a population;

Aumann's discussion of this assumption also furnishes a few observations that highlight his model-building practice. He first makes sure the reader knows that, when it comes to the common priors assumption, over and above the principled rhetoric against it, *everybody does it.* "Common priors are explicit or implicit in the vast majority of the differential information literature in economics and game theory. As soon as one writes "let p (rather than p^i) be the probability of...", one has assumed common priors."[85] He then cites thirteen areas of economics in which one form or another of the common priors assumption is used.[86] As to the *perhaps*[87] more 'realistic' assumption of different priors: "Occasionally the definitions do pay lip-service to the possibility of distinct priors p^i; but usually this is quickly abandoned, and in the theorems and examples, one returns to common priors."[88] Aumann seems to be criticizing a form of 'philosophical posturing', i.e., the principled "lip-service" paid to philosophically attractive ideas promptly followed by practically-minded applications of a useful analytical tool. For example, the mathematician Lebesgue, in a letter to another mathematician Borel, wrote: "I believe that we can only build solidly *by granting that it is impossible to demonstrate the existence of an object without defining it.*"[89] Yet two lines later he confesses: "I would have nothing more to say if the convention that I mentioned were universally adopted. But I must admit that one often uses, and that I myself have often used, the word *existence* in other senses."[90] Aumann will submit to the practice of the profession later on as well. Referring to both Harsanyi's (1967) and his (Aumann (1974)) attempts to model games without the common priors assumption, he writes: "Neither idea had any considerable echo. Apparently, economists feel that this kind of analysis is too inconclusive for *practical use*, and side-steps the major economic issues."[91] Thus, quite simply, the assumption is *used* and *useful* in many branches of economic theory.[92]

---

they will use only the information that is commonly available and will compute identical average values for priors.

[85]Aumann (1987a), p.12.

[86]The fourteenth being a "and what have you".

[87]And only so. See Aumann (1987a), p.13.

[88]Aumann (1987a), p.13.

[89]Lebesgue (1905), p.1081. This is a reference to the debate in mathematics at the turn of the century concerning the validity of pure existence proofs. These are proofs that show the existence of a mathematical object but fail to provide a method by which these objects can be constructed.

[90]Lebesgue (1905), p.1081.

[91]Aumann (1987a), p.15, my it.

[92]And even outside economics. Aumann ascribes to Savage a concern for *usefulness* as a guide for

The 'usefulness' of the resulting model is elsewhere called in by Aumann to support the common priors assumption. His theorem still holds with divergent priors. However, common knowledge of rationality now yields *subjective* correlated equilibria. Referring to this version of the main theorem of Aumann (1987a), he writes: "While such an approach is mathematically perfectly consistent, it yields results that are far less sharp that those obtained with common priors."[93] Aumann does not consider it worthwhile to jettison an assumption for the sake of 'realism', when doing so weakens results to an extent that he is not willing to accept. "It's not that assumptions don't count, but that they come *after* the conclutions; they are *justified by* the conclusions."[94] To Aumann, that theories be *useful* is the aim of all sciences. "In my view, scientific theories are not to be considered "true" or "false". In constructing such a theory, we are not trying to get at the truth, or even to approximate to it; rather, we are trying to organize our thoughts and observations in a useful manner."[95]

In economics, Aumann proposes the case of utility maximization as an example where 'unrealistic' assumptions end up supporting much important and useful work, unlike some of the assumption's rivals:

> For example, an objection that has been raised to the fundamental notion of utility maximization is that, for one reason or another, individuals do not *really* maximize utility. Alternatives such as satisficing have been proposed, which sometimes seem more appropriate as descriptions of true individual behavior. But the validity of utility maximization does not depend on its being an accurate description of the behavior of individuals. Rather, it derives from its being the underlying postulate that pulls together most of economic theory; it is the major component of a certain way of thinking, with many important and familiar implications, which have been part of economics for decades and even centuries. Alternatives such as satisficing have proved next to useless in this respect. While attractive as hypotheses, there is little theory built on them; they pull together almost nothing; they have few interesting consequences.[96]

modelling analogous to his own. "Savage is dead -so much the worse for us- and one can only speculate as to how he would have regarded the CPA. [However,] contrary to modern vogue, Savage was not a minimalist; he did not try to make his axioms as few and weak as possible, but as *useful* as possible." (Aumann (1987a), p.13, my it.)

[93] Aumann (1987a), p.14.

[94] Aumann (1998c), p.206.

[95] Aumann (1985), p.31.

[96] Aumann (1985), p.35. The 'satisficing' alternative to optimizing rationality was proposed by Herbert A. Simon. See p.51 in the sequel, where Binmore recognizes an important debt toward him. Note also

Indeed, Aumann remarks that such assumptions, quite hair-raising when interpreted as being 'realistic', underlie much of economic theory.

> Solid, down-to-earth conclusions are routinely drawn from theoretical models whose complexity -or simply sheer size- defies the imagination. General equilibrium theory requires that economic agents have complete preferences over commodity spaces whose *dimension* may run into the tens of thousands.[97]

Thus, concludes Aumann by citing the New Testament, when it comes to theories: "By their fruits ye shall know them."[98] Interestingly, more than seventy years earlier, in a talk setting out his plans for attacking problems in the foundations of mathematics, the mathematician David Hilbert had cited the same biblical passage in a context remarkably similar to this one, in the sense that a theory is judged by its outcome, rather than its starting point. "The acid test of a new theory is its ability to solve problems which, though known for a long time, the theory was not expressely designed to solve. The maxim "By their fruits ye shall know them" applies also to theories."[99]

Aumann, trying to define his method, suggests that

> the most appropriate term is perhaps "analytic"; it asks, what are the *implications* of rationality in interactive situations? Where does it lead? This question may be as important, or even more important than, more direct "tests" of the relevance of the rationality hypothesis.[100]

The last sentence[101] expresses yet again an important trait in Aumann with which Binmore will find himself completely at odds. That is, for Aumann, *what* happens is usually more important than *how* it happens. An example from Aumann (1987a), where he is dicussing the fact that his model's information system is exogenous and pondering the possible advantages of endogenous information systems, illustrates this point further:

> The distinction between "exogenous" and "endogenous" is often useful in economics, but it should not be pushed too far. In the natural sciences, such a

the importance Aumann places on the accumulation of accomplishments from past theorists. For some of Binmore's views in this regard, see later p.48.

[97] Aumann (1998a), p.933.

[98] Aumann (1997), p.3.

[99] Hilbert (1925), p.200. In this talk Hilbert uttered the famous: "No one shall drive us out of the paradise which Cantor has created for us." (Hilbert (1925), p.191)

[100] Aumann (1992), p.215.

[101] Partly a barb aimed at experimental economics?

> distinction is little used.[fn: We are aware of the pitfalls of blindly applying the methodology of one science to another. But neither should one go to the other extreme, of blindly rejecting all parallels.] When discussing the motion of the planets, should the motion of Uranus be considered "exogenous" and that of Neptune "endogenous", or the other way around? Perhaps it is not terribly important. What *is* important is the relationship between the motions.[102]

On another occasion, while discussing the value of relaxing assumptions on players' rationality, Aumann has said that:

> This kind of work is most interesting when it leads to outcomes that are qualitatively different -not just weaker- from those obtained with the stronger assumptions, but I don't recall many such cases. It can also be very interesting and worthwhile when one gets roughly similar results with significantly weaker assumptions.[103]

"In science" he says, "it is more important that the conclusions be right than that the assumptions sound reasonable."[104] We shall see later how Binmore's approach differs markedly from this.

Aumann's (1987a) paper is about seventeen pages long, excluding the bibliography. About five pages (two sections) deal with definitions, assumptions, the theorem and the proof. The rest is devoted to interpretation and discussion. Yet Aumann shows great care to make sure that formal and informal sections of the paper do not get carelessly entangled. To say that players *know* some event E is not the same as saying that players *know* each others' information partitions. In this last case, it:

> Is not an assumption, but a "theorem", a tautology; it is implicit in the model itself. Since the specification of each ω includes a complete description of the state of the world, it includes also a list of those other states ω' of the world that are, for Player 2, indistinguishable from ω.[105]

After discussing similar points a little further, Aumann ensures no misunderstandings are possible:

[102] Aumann (1987a), p.11.
[103] Aumann (1997), p.12.
[104] Aumann (1997), p.3.
[105] Aumann (1987a), p.9.

> The reader may ask why we adduce verbal, informal arguments for these assertions, rather than proving them as formal theorems. The answer is that the assertions have no formal content. Within the model, knowledge refers to events, i.e., sets of states. One can ask whether at a given state ω, a player i knows an event A; this is the case if and only if A includes that element of $\wp^i$ that contains ω. But neither a partition nor a prior is an event; formally speaking, the concepts of knowledge and common knowledge do not apply to them.[106]

That is, in that case, ""know" has its informal, everyday meaning."[107] Aumann further states that each "player uses only his own prior and his own partition in reaching a decision; within the model, it does not matter that he "knows" the partition and priors of the others. It is only in interpreting the results that these points become significant."[108] It is important in the sense that the 'world' that is considered is characterized by implicit 'knowldedge' by the players of its structure.

Although Aumann's insistence on the need for theories to be 'useful' and the many pages devoted to discussion in the majority of his papers attest to the importance he attaches to the *interpretation* of formalisms, this meticulous manner in which he ensures that it is absolutely clear when he is doing exactly what (mathematics or interpreting mathematics) is very characteristic. This is reflected in the internal structure of his papers. Following a short introduction come definitions, results and proofs. Then comes the discussion, methodically organized in broad categories of carefully indexed subsections. All is neat, tidy and *serious*. We can feel some of this in the reproach quoted earlier[109] that Aumann levelled at Binmore for his sloppiness in presentation and his mish-mashing of formal and informal concepts.

Aumann's style is the one he ascribes to von Neumann and Morgenstern (1944): "The method of von Neumann and Morgenstern has become the archetype of later applications of game theory. One takes an economic problem, formulates it as a game, finds the game-theoretic solution, then translates the solution back into economic terms."[110] For Aumann, interpretations must be wholly grounded in an exhaustive axiomatic framework, and must not deviate from too widely from that structure. 'Wild' interpretations must be curtailed

[106] Aumann (1987a), p.12.
[107] Aumann (1987a), p.10.
[108] Aumann (1987a), p.10.
[109] See p.10.
[110] Aumann (1987c), p.466.

by a mathematical base that is as explicit as possible. "An axiomatic development [...] is ideally suited to deal with such matters [of interpretation]: by separating sharply between assumptions and the formal process of deducing conclusions from them, it enables us to focus and clarify the interpretative discussion."[111] Hilbert expressed this view very nicely when he said: "In a sense, mathematics has become a court of arbitration, a supreme tribunal to decide fundamental questions -on a concrete basis on which everyone can agree and where every statement can be controlled."[112] Or elsewhere:

> Indeed, the requirement of rigour, which has become proverbial in mathematics, corresponds to a universal philosophical necessity of our understanding; and, on the other hand, only by satisfying this requirement do the thought content and the suggestiveness of the problem attain their full effect. A new problem, especially when it comes from the world of outer experience, is like a young twig, which thrives and bears fruit only when it is grafted carefully and in accordance with strict horticultural rules upon the old stem, the established achievements of our mathematical science.[113]

Over and above the *quality* of the handling of mathematics and their interpretations in game theory and economics, Aumann worries about their *quantity.* According to him:

> Game theorists argue too much about interpretations. Sure, your starting point is some interpretation of Nash equilibrium; but when one is doing science, you have a model and what you say is: What do we *observe*? Do we observe Nash equilibrium? Or don't we observe it? Now, *why* we observe it, that's a problem that is important for philosophers, but less so for scientists.[114]

This lack of interest in what Aumann considers to be philosophical issues, or at least the opinion that they are far from relevant to his work, is characteristic of the attitude of some

[111] Aumann (1998a), p.935.
[112] Hilbert (1925), p.200.
[113] Hilbert (1900), p.1099. Or once again, elsewhere: "On the contrary I think that wherever, from the side of the theory of knowledge or in geometry, or from the theories of natural or physical science, mathematical ideas come up, the problem arises for mathematical science to investigate the principles underlying these ideas and so to establish them upon a simple and complete system of axioms, that the exactness of the new ideas and their applicability to deduction shall be in no respect inferior to those of the old arithmetical concepts." (Hilbert (1900), p.1100) This last quote should underscore that for Hilbert (as for Aumann), the relationship of mathematics to the disciplines studying the 'real world' is crucial. Where Aumann and Binmore disagree is on the details of the relationship.
[114] Aumann (1998c), p.192.

mathematicians to their own foundational 'crisis'[115] in the first half of the twentieth century. Jean Dieudonné, a French *bourbakiste*, has said that:

> Les philosophes et les logiciens ont une tendance, parfaitement naturelle et excusable, à croire que les mathématiciens s'intéressent beaucoup à ce qu'ils font. Détrompez-les, ce n'est pas vrai: 95% des mathématiciens se moquent éperdument de ce peuvent faire tous les logiciens et tous les philosophes. Cela ne les intéresse absolument pas.[116]

One reason for this lack of interest, says Dieudonné, is that mathematicians just *use* Zermelo-Fraenkel set theory to build theories in many areas of mathematics, and the (non-)debate on 'foundations' ends at that.

> Ce système répond exactement aux besoins de tous les mathématiciens, excepté, bien sûr, les logiciens et aussi ceux que leur attitude philosophique empêche d'accepter les prémisses d'un tel systême, c'est-à-dire les mathématiciens dits intuitionnistes ou constructivistes.[117]

A whole field of human thought, like mathematics, or a smaller sub-field like game theory, does not stop and stand still when a debate on "foundations" breaks out. The clashes usually occur between a relatively small minority of practitionners. And the people who occupy themselves applying these theories sometimes hardly even hear of these rumblings in the higher spheres. As Poincaré wrote while efforts to surmount the paradoxes of set theory and logic were in progress:

> Needless to say, Cantorism and logistic are alone under consideration; real mathematics, that which is good for something, may continue to develop in accordance with its own principles without bothering about the storms which rage outside it, and go on step by step with its usual conquests, which are final and which it never has to abandon.[118]

Hans Lewy, who was a Privatdozent in Göttingen when Brouwer came there to deliver a talk at the height of his confrontation with Hilbert, said afterwards:

> It seems that there are some mathematicians who lack a sense of humor or have an over-swollen conscience. What Hilbert expressed there seems exactly right to

[115]The quotation marks would be from their point of view.
[116]Dieudonné (1982), p.16.
[117]Dieudonné (1982), p.17.
[118]Poincaré (1906), pp.1062-3.

> me. If we have to go through so much trouble as Brouwer says, then nobody will want to be a mathematician any more. After all, it is a human activity. Until Brouwer can produce a contradiction in classical mathematics, nobody is going to listen to him.[119]

For Aumann, and in contrast to Binmore, although he undoubtedly considers his work in foundations to be important and interesting, no sense of urgency, of impending doom for game theory, radiates through his papers. He believes that game theory's concepts, both in an empirical and more intuitive sense, have generally been quite successful as tools for research in the social sciences and in economics in particular, irrespective of the uncertainty surrounding their foundations.

> In the end, I think that the ordinary laws of economic activity apply to our fields as well. The world will not long support us on our say-so alone. We must be doing something right, otherwise we wouldn't find ourselves in this beautiful place today.[120]

Aumann intuitively, and genuinely, approves of most results that Binmore's own intuitions bar him from accepting.

> Up to now, most of the implications of game theory and economic theory have been not quantitative but qualitative. For such an implication to be convincing, it is important that it be supportable in common sense terms. Game Theory is most satisfying when the formal analysis suggests new insights -insights that, while not obvious, do eventually make sense on the common sense, verbal plane.[121]

[119] Cited in Reid (1970), p.184.
[120] Aumann (1985), p.37. See also Aumann (1998c).
[121] Aumann (2000) vol. II, p.61.

1.3 COMMON KNOWLEDGE OF (A LITTLE BIT OF) IRRATIONALITY

> The source of all great mathematics is the special case, the concrete example. It is frequent in mathematics that every instance of a concept of seeminly great generality is in essence the same as a small and concrete special case.
> Paul Halmos[122]

Aumann's next important contribution to the epistemic litterature is, in its structure, intimately tied to Aumann (1987a). Actually, it is more of an illustration of possible applications of epistemic models to special cases. Its originality, and what contrasts it with Aumann (1987a), is that it looks at the implications of considering as possible states of the world at which players are *irrational.* Indeed, "what is proposed here is to take this framework [that of Aumann (1987a)], remove the rationality hypothesis, and see where we are led."[123] In particular, he means to show, by using a simple example, that permitting players to believe, even with a tiny probability, that one of them will act irrationaly at a given state of the world can lead to outcomes that are substantially different from those which result if that particular state of the world is considered impossible, i.e., if there is common knowledge of rationality. This result will be useful for him in order to stress the radical difference between common knowledge of rationality and non-common knowledge of rationality, whatever form the latter may take.

In writing this simple yet imaginative paper, Aumann shows great flexibility. Having greatly contributed to the formation of a coherent and holistic approach to reasoning about rationality and knowledge of rationality, attacked by some for the rather herculian cognitive capacities with which it endows players, he turns around and explores the meaning of irrationality within the very same framework. On top of this, his results have unambiguous, easily grasped and intuitively satisfying qualitative implications. That is, he shows that he can obtain outcomes that are not the backward inductive equilibrium in a strategic form version of the Centipede Game.[124] Producing seemingly simple examples that are deep enough to illustrate the whole of a theory, and even to extend its scope to areas not immediately

[122]Halmos (1985), p.324.
[123]Aumann (1992), p.215.
[124]The Centipede Game in its extensive form will be very important in the debates with Binmore on backward induction. For a tree of the game, see p.119.

evident from its abstract formal structure, is usually considered a sign of the profundity of a mathematician's thought. It demonstrates a firm grasp of the foundations of a theory, both their formal relationships and their informal interpretation.

Essential to the creation of deep examples is the drive to do so. That is, having erected a very pretty, generally esteemed and very abstract model, Aumann was clearly not enticed to leave it at that. Neither did he tread the more traditional path and generalize a subsection of his model, or extend its range in a formal way, "taking the uncompromising stance of a mathematician [...] in his drive for ever weaker assumptions and ever stronger conclusions, and in his compulsive quest for simplicity."[125] That kind of work feeds the hungry mouths of many mathematical economists' families, and in many cases such results can yield deep insight into the structure of models. Henri Poincaré, grappling with the difficult question he had posed himself, namely; "What is mathematical creation?", began to answer by stating what to him it was *not*:

> It does not consist in making new combinations with mathematical entities already known. Any one could do that [!], but the combinations so made would be infinite in number and most of them absolutely without interest. To create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority. Invention is discernment, choice.[126]

Some ideas of interest to us emerge from this last quote. However, it was not introduced into this text as implying criticism of the more purely formal aspects of mathematics, in itself and as applied to economics. More importantly, it should illustrate the fact that the criteria retained for picking a few fascinating and illuminating members out of the 'infinite' number of additions, combinations and extensions of given mathematical ideas need not automatically select the most technical or general of them. Other factors are at work, and Poincaré (as would Aumann) cites discernment in favor of *useful* developments. Of course what advancements indeed possess such a distinction can sometimes depend on who you ask. A classic case of this is the debates between Kroenecker and Wieirstrass on the 'usefulness' of modern analysis.[127] These things seemingly will always remain nebulous and somewhat arbitrary. As Poincaré has noted elsewhere:

[125] Debreu (1992), p.110.
[126] Poincaré (1956), p.2043.
[127] See Bell (1937), chapters 22 and 25.

> Among all the constructs which can be built up of the materials furnished by logic, choice must be made; the true geometer makes this choice judiciously because he is guided by a sure instinct, or by some vague consciousness of I know not what more profound and more hidden geometry, which alone gives value to the edifice constructed.[128]

Neither are interesting examples 'easy' or 'evident' to produce once one knows the relevant axioms and mathematical objects. An important example is a mathematical creation of its own, often a very difficult and challenging one. Take the case of this example as an example. Aumann relates that he was incited to produce his paper by:

> A conversation at Stanford during the summer of 1987, [with] Jay Kadane of Carnegie-Mellon University [who] suggested that it might be worthwhile to investigate the consequences of relaxing the assumption of common knowledge of rationality that underlies Aumann (1987a). At the time, the framework for carrying out such an investigation was not clear. I suggest that the framework of the previous section is appropriate for this purpose.[129]

Thus even Aumann, the 'Formaliser of Common Knowledge', whose knowledge of the intimacies of the mathematical tools involved in examining knowledge in game theory cannot be doubted, had to take some time[130] to come up with a seemingly simple example, 'tautologically' implied by the framework of one of his models. On top of that, his first attempts contained a "serious error",[131] the finding of which he attributes to Adam Brandenburger. As another example, consider the mathematician Halmos' recounting of his reaction to a paper that "made me angry: it struck me as wordy and pretentious."[132] So he attempted to prove that paper's result in a more straightforward manner. He "dashed off" a note to the Bulletin of the American Mathematical Society with a proof that was "a lot slicker" and "a lot shorter".[133] However, he was subsequently told that one of the lemmas on which his proof rested was false. He describes the offending lemma as "a statement for which it's not only easy to find counterexamples but it's hard to find any non-trivial instances where

[128] Poincaré (1905), pp.1024-5.
[129] Aumann (1992), p.218.
[130] As an indication only, there are five years in between his conversation with Kadane and the *publication* of his example.
[131] Aumann (1992), p.227.
[132] Halmos (1985), p.156.
[133] Halmos (1985), p.156.

it is true."[134] Thus, here we have a mathematician, quite respected in the field in which he was intervening,[135] who not only fails to come up with a convincing example to illustrate a result, but also misses every one of the purportedly 'easy' counterexamples.[136]

The framework within which knowledge is addressed in this example is not the same as in Aumann (1987a). Knowledge is defined here in probabilistic terms, as belief with probability 1. Knowledge partitions of the state space Ω are not used in this information system, their function is taken by a probability distribution for each player conditionned on the occurence of certain events. Formally, an *information system* is defined as:

(i) an n-person strategic form game G;

(ii) for each player i, a set whose members, s_i, are called *information states* of i;

(iii) a function whose domain is the set of all information states of player i and whose image is the product of the pure strategy set of i in G and the set of probability distributions on *(n-1)-tuples* of information states of the other players.

That is, to each information state of player i is associated one and only one pure strategy[137] and one probability distribution on the other players' information states, which implies that he can deduce a probability distribution on the other players' pure strategies.[138] The former distribution, Aumann terms that player's *theory*, and the latter, derived from his theory, his *belief*. Player i *knows* an event E at a given information state if he assigns a probability of 1 to the occurence of E given that he was informed of the content of his information state. A *state of the world* in this model built up from information states, precisely it is an *n-tuple* of such states. Thus at a given state of the world, each player is assigned a pure strategy and a theory about the pure strategies of the other players and their theories. Thus he deduces from their theories, their theories about theories and so on,

[134] Halmos (1985), p.156.

[135] The paper was on a subject related to measure theory, at a time when this was one of Halmos' main preoccupations.

[136] Even if he wanted to come up with examples for purely personal motives, i.e., not for the published paper.

[137] This is an assumption that players know their own strategy.

[138] Since to each of these players is also associated a function which maps information states into pure strategies. These functions being part of the structure of the model, they are implicitely common (informal) knowledge.

their beliefs, their beliefs about beliefs, etc. And i is rational at a given information state if he is maximizing his payoff given his information at that state. Aumann assumes that the basic probability distribution which players update given their information is common to all players, that is, he employs the *common priors assumption*. Aumann states that "information systems are essentially equivalent to the standard partition models of information in games. [...] Formally, information systems are slightly more parsimonious that partition models, and are better suited for our current purposes."[139]

Aumann's main example corresponds to a strategic form version of the Centipede Game, reproduced in Figure 1.1. The players are Alice (row) and Bob (column), the payoff to Alice for a given cell of the matrix being stated in the top left corner, and the payoff to Bob in the lower right corner. The letters v, y, x, y and z attached to given cells mean that the players assign positive (common) prior probabilities equal to v, w, x, y and z to the occurrence of the states of the world where the strategies by which these cells are reached are played. Notice that the only Nash equilibrium of this game, which is also the only subgame perfect Nash equilibrium in its extensive form version when Alice has the first move, calls for Alice to 'take the money and run' at her first move.

[139] Aumann (1992), p.218.

Bob

		a	b	c	d
	A	10 / .50	10 / .50	10 / .50	10 / .50
	B	5 / v 100	1 000 / w 50	1 000 / 50	1 000 / 50
Alice	**C**	5 / 100	500 / x 10 000	100 000 / y 5 000	100 000 / 5 000
	D	5 / 100	500 / 10 000	50 000 / z 1 000 000	0 / 0

Figure 1.1

The Centipede Game in strategic form.

Aumann insists that although high gains could be achieved further on in the game were Alice to stay in, he does not consider her decision to withdraw at the first node to be 'paradoxical'. For:

> Although Alice may feel quite frustrated, a considered analysis will nevertheless lead her in the end to pick up the money at the first opportunity. After all, there is a difference between frustration and paradoxality: the players of a one-shot prisoner's dilemma will certainly feal frustrated, but the logic of that situation *does* inexorably point to playing "greedy", and the time has long passed since this was considered paradoxical.[140]

However, aware that a growing number of game theorists were calling into question the relevance of backward induction for games in extensive form like that of the Centipede, he adds that "others will feel that if this is rationality they want none of it -or, more to the point, that it represents an approach that is of little practical interest, at least in this example."[141]

Coming back to the example, what Aumann shows is that there can exist a very

[140] Aumann (1992), p.220.
[141] Aumann (1992), p.220.

small probability z such that Alice is irrational at the state of the world which occurs with probability z but is rational at all other states of the world which occur with positive probability, while Bob is rational at all such states of the world. Alice is not rational at z since she knows at z that Bob will play c and her best response to c is to play C. However, she is rational at y where she plays C if

$$(100000)y + (500)x \geq (1000)x + (1000)y;$$

since at y she believes with probability x that Bob will play b, to which her best response is to play B. Thus she is rational at y if $\frac{x}{y} \leq 198$. Proceeding this way for Alice's other non-null states of the world, and doing the same for Bob, grants the desired result if none of the ratios $\frac{v}{w}$, $\frac{w}{x}$ and $\frac{y}{z}$ are larger that 198. Since the value of z is tiny in this case,[142] and yet solutions that are not backward inductive equilibria are not only possible but rational, Aumann claims that "this attractive resolution for these paradoxes gives a rigorous justification to the elusive ideal that, whereas one should certainly play rationally at the end, it seems somehow foolish to act from the very beginning in the most pathologically pessimistic, "play it safe" way."[143] Not only are the solutions not subgame perfect equilibria, they are not even Nash equilibria.

It is interesting to use this example to illustrate what happens when one assumes common knowledge of rationality. Then the state of the world that previously occurred with probability z *cannot* occur, thus $z = 0$. But the fact that z was different than zero, although very small, was what made playing c rational for Bob even when he believed that the probability that Alice would play C was high, to which his best response would usually be to play b. Therefore Bob is now irrational at the state of the world occuring with probability y, so the latter vanishes. But now alice is irrational at the state of the world that occurs with probability x, and so on. Common knowledge of rationality makes the particular prior probability distribution assumed in this example impossible, since the only state of the world where this knowledge is true is when Alice plays A and Bob plays a, is assigned probability zero in it. In the previous example both players considered it *a priori* impossible that Alice terminate the game on her first move.

[142] If all the above ratios are set at 198, the differences in probabilities are huge (I.e., $z = \frac{v}{(198)^4}$).

[143] Aumann (1992), p.223.

Aumann, in conclusion, gives an informal interpretation of his formal interpretation of the model of Aumann (1987a).

> Most of us have experienced situations where some harmful fact is perfectly well known but is studiously overlooked by everybody. In this case, the harmful fact is the players' rationality (!). More precisely, the fact itself need not be harmful, but common knowledge of it would be. The above approach enables us to understand this phenomenon within the context of the theory.[144]

1.4 THE EPISTEMIC FOUNDATIONS OF NASH EQUILIBRIUM

The next of Aumann's papers to be reviewed is co-authored with Adam Brandenburger, and it is titled *Epistemic Conditions for Nash Equillibrium.* Deeply linked to Aumann (1987a) and the rest of the epistemic literature, it stands in some sense in a class of its own, as a form of *aboutissement.* For although correlated equilibria are very interesting objects in themselves, the bread and butter of game theoretic analysis, especially that which is applied to economics and elsewhere, is based on the Nash equilibrium or one of its refinements. The paper's fundamental question is: "just what epistemic conditions *lead* to Nash equilibrium"?[145] The strategy is that of Aumann (1987a): "we seek sufficient epistemic conditions for Nash equilibrium that are in a sense as "spare" as possible."[146] They attempt to *justify* Nash equilibrium in a truly simple and (in principle) straightforward way. With their approach, "you get equilibrium when certain informational assumptions are satisfied",[147] and that is the end of the matter. *Discovering* the afore mentioned assumptions is the job of mathematicians, *verifying* and *discussing* their validity is left to empirical economists and to philosophers. "So, the story of *why* we have Nash equilibrium, it's an important philosophical problem, but as science it is not that important."[148] To establish the foundations of game theory, Aumann sees no need to look outwards to philosophy. Rather, he looks at game theory *from the outside*, with the help of a new enveloping layer of mathematics that permits the clarification of the more familiar concepts. In this there

[144] Aumann (1992), p.226.
[145] Aumann and Brandenburger (1995), p.1161, my it.
[146] Aumann and Brandenburger (1995), p.1161.
[147] Aumann (1998c), p.204.
[148] Aumann (1998c), p.192.

is much that resembles the idea behind David Hilbert's 'metamathematical' approach to the debate on the foundations of mathematics. Over and above all, there is the sincere conviction that what is needed is not less mathematics and more philosophy but more mathematics, period. *Within* the realm of mathematics shall be found the answers to riddles that appear during its development. As Hilbert put it: "Let us remember that *we are mathematicians* and that as mathematicians we have often been in precarious situations from which we have been rescued by the ingenious method of ideal elements."[149] John von Neumann remarked that under Hilbert, the question of the foundations of mathematics, "in and of itself philosophico-epistemological, is turning into a logico-mathematical one."[150] Or, as has been stated by a historian studying Hilbert:

> En remplaçant la question du rapport à l'expérience par celle de la non-contradiction, Hilbert transforme le fondement, d'un problème épistémologique en un problème logique. Le problème des fondements peut devenir un exercice mathématique susceptible d'une solution par démonstration.[151]
>
> Pour Hilbert, la force du programme formaliste tient à ce que celui-ci rétablit les lois logiques et donne aux mathématiques un fondement possédant une rigueur mathématique et dépourvu de présupposés extérieurs aux mathématiques.[152]

The framework of this article is much like that of Aumann (1992), i.e. that of an *information system*, termed *belief system* in this paper. As there, each of the n players' *theory* is a probability distribution on the *(n-1)-tuple* of other players' *information states* (in this paper called *types*). From his *theory*, a player can deduce a *belief* (in this paper called a *conjecture*) on *(n-1)-tuples* of pure strategies (in this paper called *actions*) of the other *n-1* players.[153] Note that a player's conjecture considers all of his opponents' actions together, at once. He does not form conjectures about the others' actions individually. This will be important when there are more than two players. In this paper, as in Aumann (1992), a player is said to *know* an event if he assigns probability 1 to that event.

The paper contains two main theorems, one for games of two players and one for

[149] Hilbert (1925), p.195.
[150] von Neumann (1983), p.61.
[151] Cassou-Nogués (2001), p.67.
[152] Cassou-Nogués (2001), p.106.
[153] Remember that some function maps a player's type space into his action space, so that to each type is associated one (and only one) action. These functions being a part of agents' 'world', they are implicitely taken into account when the latter make decisions, i.e., they are informally 'common knowledge'.

games of more than two players. The first, Theorem A, states: "*Suppose that the game being played (i.e., both payoff functions), the rationality of the players, and their conjectures are all mutually known. Then the conjectures constitute a Nash equilibrium.*"[154] The authors state what must have struck many people about this result:

> In Theorem A [...] common knowledge plays no role. This is worth noting, in view of suggestions that have been made that there is a close relation between Nash equilibrium and common knowledge -of the game, that players' rationality, their beliefs, and/or their choices."[155]

The second theorem of this paper, Theorem B, states: "*In an n-player game, suppose that the players have a common prior, that their payoff functions and their rationality are mutually known, and that their conjectures are commonly known. Then for each player j, all the other players i agree on the same conjecture* σ_j *about j; and the resulting profile* $(\sigma_1, ..., \sigma_n)$ *of mixed actions is a Nash equilibrium.*"[156] Again, common knowledge does not necessarity play the role that was expected of it. "So common knowledge enters the picture after all, but in an unexpected way, and only when there are at least three players. Even then, what is needed is common knowledge of the players' conjectures, not of the game or of the players' rationality."[157] Notice also that the assumption of common priors is required for Theorem B. This stems from the fact that even though each player plays only one action in equilibrium,[158] the equilibrium is represented by players' *conjectures*, that is, distributions on the other players' actions, induced 'mixed actions'. The actions of the players, which individually are perfectly determinate, are perceived by the other players as particular realisations of random actions. This probably explains part of the reason why the term "action" is substituted for "strategy" in many of Aumann's papers, since the role played by strategies in the traditional Nash equilibrium is taken up here by players' beliefs. The assumption of common priors is required "since j's component of the putative equilibrium is meant to represent the conjectures of other players i about j, and these may be different for different

[154] Aumann and Brandenburger (1995), p.1162. An event is mutual knowledge if all players *know* it, but not more.

[155] Aumann and Brandenburger (1995), p.1162.

[156] Aumann and Brandenburger (1995), p.1163.

[157] Aumann and Brandenburger (1995), p.1163.

[158] I.e. a pure strategy.

i, it is not clear how j's component should be defined."[159]

In a footnote to the article are cited many examples of well known papers and texts about and within game theory, including some of the authors' own, in which common knowledge of just about everything in the game was assumed to be a given. Indeed, these types of assumptions were taken for granted by just about everyone in the field, from the top theorists to graduate students teaching microeconomics classes. Does it not seem that these results of Aumann and Brandenburger toss a great big monkey wrench into the well-oiled machine that was the 'standard' introductory approach to game theory? One could find it interesting to ponder on how long these results will take to filter down into the classrooms, and in what shape they will be in when they arrive. It has become more difficult to state that beliefs and the game itself are common knowledge among the players by saying something of the form: "The structure of the game is common knowledge".[160] Instead of this large blanket of assumptions, professors will now have to qualify: "Well, since in this case there are x players..."; "No, no, only the beliefs, not the payoffs"; "It is just mutual knowledge, "I know that ..."". And if students wish to know where these rather precise conditions come from, professors will need to get into a theory that is rather complex and subtle, and probably entirely new to students. In fact, in some of the newer game theory manuals, these issues have been getting more attention.[161]

We feel we should reproduce the proof of Theorem A here to give the reader an idea of the type of reasoning used in this literature. It is much simpler than the proof of Theorem B, it is clear and intuitive, yet it illustrates the types of issues and pitfalls one encounters in these types of models. The first step is to show that *player* i *attributes probability* π *to an event* E *if and only if he indeed attributes probability* π *to* E (Lemma 1).

Proof: *If*: Suppose player i attributes probability π to E. Take the event $F = \{t : P(E : t_i) = \pi\}$.[162] At any state of the world t where our hypothesis is true

[159] Aumann and Brandenburger (1995), p.1163.

[160] Although, as a professor has told me, common knowledge of the game and the players' beliefs is *sufficient* to ensure a Nash equilibrium. However, it becomes false to state that Nash equilibrium *requires* such common knowledge.

[161] Two examples are Binmore (1992b) and Osborne and Rubinstein (1995).

[162] The distribution $P(F : s_i)$ is extended from player i's theory. That is, for a given event E, it evaluates the probability of the event $\{s^{-i} \in S^{-i} : (s_i, s^{-i}) \in E\}$. Recall that i's theory is a distribution on all

we have $P(E : t_i) = \pi$, therefore $t \in F$ and we have $P(F : t_i) = 1$. So *i knows F* at t.

Only if: Suppose player *i knows* he attributes probability π to *E*, but that 'in reality' he attributes probability ρ to *E*. By the proof of the 'if' part of this lemma he must *know* that he attributes probability ρ to *E*. Thus for some state *t* where he knows that he attributes both probabilities to *E*, $P(E : t_i) = \pi = \rho$.

This lemma easily extends to the following one (Lemma 2). *Let ϕ be an n-tuple of conjectures. Suppose that at some state s, it is mutually known that the conjectures are ϕ, then $\phi(s) = \phi$, i.e., they are indeed ϕ.* Lemma 2 translates the more general Lemma 1 into the the form it needs in a multi-player game. Each player i's component of the *n-tuple* ϕ is a probability distribution on the product of the strategy spaces of the *n-1* other players. Thus when a player *knows* his component ϕ^i of ϕ, by Lemma 1 his conjecture is truly ϕ^i. Since this is the case for all players, the *n-tuple* of conjectures is indeed ϕ.

The next step in the proof of Theorem 1 is another lemma (Lemma 3): *Let g be a game, ϕ an n-tuple of conjectures. Suppose that at some state s, it is mutually known that* $\boldsymbol{g}=g$,[163] *that the players are rational, and that* $\boldsymbol{\phi} = \phi$. *Let a_j be an action of a player j to which the conjecture ϕ^i of some other player i assigns positive probability. Then a_j maximizes g_i against ϕ^j.*

Proof: Since *i knows* ϕ^i, he indeed attributes positive probability to $[a_j]$(by Lemma 2). He also attributes probability 1 to the events $[\phi^j]$, [*j is rational*] and $[g]$. Therefore there exists at least one state of the world where these four events occur simultaneously,[164] that is, where the game is g, j is rational, his conjecture is ϕ^j and he chooses a_j. This means that a_j is optimal against ϕ^j.

The proof of Theorem A will come, in a manner similar to the passage from Lemma 1 to Lemma 2, from applying the results of Lemma 3 to both players.

Proof of Theorem A: By Lemma 2, $\phi = (\phi^i, \phi^j)$ is the true conjecture pair. By Lemma

(*n-1*)-tuples s_{-i}.

163 This notation uses the fact that since g depends on an uncertain state space, it can be viewed as a random variable. Thus **g**=g is the event that the random variable **g** takes takes the value g. This event will also be written $[g]$.

164 That is, since all these probabilities are evaluated at s_i, i's information state, there must be an (*n-1*)-tuple s^{-i} of other players' information states where the four events occur.

3, every action a_j with positive probability in ϕ^i is optimal against ϕ^j in g, and every action a_i with positive probability in ϕ^j is optimal against ϕ^i in g. Thus (ϕ^i, ϕ^j) is a Nash equilibrium of g.

Why do only conjectures need to be common knowledge when $n \geq 3$? Mutual knowledge of conjectures alone, even common knowledge, is not sufficient because of the problems the authors noted earlier: conjectures can be common knowledge even if all players attribute different probabilities to the use by a given player of a given action. But by adding a common prior to common knowledge of conjectures, what do we obtain? We get a version of Aumann (1976)![165] The latter told us that if players share a common prior and if their posteriors for a given event (say, the event that player i takes action a_i with probability π) are common knowledge, then these must be equal. Thus the common prior will allow players to coordinate onto a common conjecture.

Thus with conjectures the problem we face is that players could potentially be examining entirely different states of the world, unless we impose some form of coordinating mechanism. What about knowledge of rationality and of the game? Why does that remain 'mutual' when $n \geq 3$? Take three players, i, j and k. Suppose that rationality is mutual knowledge. Then i *knows* the event *[j is rational]*, and k *knows* the event *[j is rational]*. Evidently, i and k now have access to information about the same states of the world, the ones in which j is indeed acting rationally. So the kind of confusion we met with conjectures cannot occur. In a sense, rationality and payoffs[166] relate to the 'general environment' surrounding the players, not to actual occurrences. Rationality says something about the *types* of decisions a player may make (what he will *not* do), not, in general, about precisely *what* decisions he indeed takes or intends to take. By requiring knowledge of that form, Nash equilibrium is more stringent.[167] So the thought process behind the justification of Nash equilibrium that Aumann (1987a) brought up, which says that 'if all players believe *for some reason* that all other players will play their components of the equilibrium and they still want to play theirs, then those strategies constitute a Nash equilibrium', is not at all evinced by Aumann and Brandenburger (1995). They just change the way we *say* it.

[165]Even though Aumann (1976) is not mentioned in this paper.
[166]I.e., the game.
[167]I.e., a 'shorter' list of states of the world corresponds to its occurence.

Common knowledge of conjectures and a common prior replace the: ‘for some reason this player believes that this other player will play the strategy ...’. It is probably in this case that the true aim of Aumann’s method in the epistemic literature is at its clearest.[168] His plan’s aim is very deep, in its structure, its applicability, its abstractness, and yet its scope is, in the sense pointed to above, remarkably narrow.

[168]Maybe because we know Nash Equilibria more than correlated equilibria or other less common solution concepts.

CHAPTER 2

BINMORE: THE LABOURS OF HERCULES

2.1 A QUEST BEGINS

In this part of the thesis, we turn to Binmore. Here, quotes from Binmore's texts will be plentiful and rather long. The author's many more philosophical papers, as well as his style of writing, seem to make this necessary. Quoting Binmore at length, at the same time permitting the reader to 'take in' more of his personality, seems preferable to rephrasing his ideas in order to avoid excessive quotation.

Binmore, like with Aumann, was an early contributor to the debate on the foundations of game theory. Or rather, he could be likened to an early and loud alarm bell, since his role in that field is essentially that of the thought-provoking critic. He voices his concerns on foundations in some of his first contributions to game theory and economics. And, in some ways, it could be said that 'Binmore the Critic' arrives on stage with his script fully mastered. In 1984 he wrote a short paper, *Equilibria in Extensive Games*,[169] which already contained much of the criticism he would later level at the principle of backward induction, at the Aumann-inspired theory of knowledge and at the traditional model of a player; in brief, his distinctive stance in the debate on foundations. From that point on, he became 'a man on a mission', unfailingly driven towards his goal of refounding game theory. Individual arguments are developed and refined, but at the core very little really changes.

Quite a few of his papers contain whole sections that hammer away at identical points. Sometimes even the same sentences or expressions reappear. He is quite conscious of this: "The reader will find the same issues addressed over and over again, sometimes in almost

[169] Binmore (1984).

the same words."[170] Or: "To pluck yet again the single string on this minstrel's harp,"[171] he kids before stating yet again his position. He explains his 'strategy' in his books on philosophy: "I have made a point of repeating the really important ideas again and again, because I am weary of being misunderstood, and see no way to insist that I literally mean what I say other than by straightforward repetition."[172]

He seems to fancy his war on what he perceives to be the dominant methodology of game theorists as one of David against Goliath. Note the epic tone of the following passage where Binmore describes the toilsome task which awaits him. It should also give a taste of Binmore's cultured yet funny literary flurries which lend his papers a distinctive 'flavour':

> Writing on the foundations of game theory is a Herculean task. I don't know which of the labours of Hercules provides the most apt metaphor. It is tempting to cite the Augean stables, which housed three thousand oxen but had not been cleaned for thirty years, but this would imply too harsh a judgement. The nine-headed Hydra, which grew two heads for each that was struck off, would do very well for a piece on refinements of Nash equilibrium [...]. I shall [...] settle for the wrestling match with the giant Anteas, although I fear his feet are so firmly entrenched that it would truly take a Hercules to move them.[173]

It seems that Binmore-Hercules has settled on the strategy of repeatedly stomping on one of the giant's toes.

In a similar spirit and using an identical analogy, Poincaré had attacked the schools of logicists and formalists that were forming in his time to address the problems in the foundations of mathematics:

> It is time to administer justice on these exaggerations. I do not hope to convince them; for they have lived too long in this atmosphere. Besides, when one of their demonstrations has been refuted, we are sure to see it resurrected with significant alterations, and some of them have already risen several times from their ashes. Such long ago was the Lernaen hydra with its famous heads which always grew again. Hercules got through, since this hydra had only nine heads, or eleven; but here there are too many, some in England, some in Germany, in Italy, in France, and he would have to give up the struggle. So I appeal only to unprejudiced men of good judgement.[174]

[170] Binmore (1990c), p.viii.
[171] Binmore (1992b), p.18.
[172] Binmore (1998), p.xix.
[173] Binmore (1992b), p.1.
[174] Poincaré (1905), p.1023.

Binmore's lack of orthodoxy extends to the tone of his writings; his less technical papers are liberally sprinkled with humour and light-hearted banter, often in slightly satirical form. Although this seems only to communicate a certain *joie d'écrire*, he suggests that scholars who take offence at this 'happy-go-lucky' style are the ones most likely to be unable to appreciate novel work.

> I hope that my more sophisticated readers will appreciate that the purpler passages with which I try to brighten up the arguments are sometimes actually meant to raise a wry smile. I know the risks that accompany such departures from the deadpan style of respectable intellectual inquiry. Why else do Mill and Kant have a reputation of intellectual rigour denied to Bentham and Hume? But my experience is that minds which take their scholarship very seriously are unreceptive to challenges to traditional wisdom in whatever form they may be expressed.[175]

Binmore does not bow to established wisdom in any domain. He has written that:

> My childhood experiences had already taught me to distrust authority in all its shapes and forms. Looking back, I see now that this is the most valuable lesson I ever learned, although I have to admit that this particular piece of learning is best neglected if one is seeking a smooth passage to the top ot the tree.[176]

For example Binmore does not feel that deference is automatically owed to the important texts and authors of a discipline. Of von Neumann and Morgenstern (1944), he writes: "I have read the great classic from cover to cover, but I do not recommend the experience to others! Its current interest is largely historical."[177] "Nor" in philosophy, he admits, "am I very respectful of the great thinkers of the past. [...] Nor does a great thinker have to be right about everything in order to be great."[178]

2.2 TEARING EVERYTHING DOWN

In this section, we look at those parts of the foundations of game theory that Binmore considers to be wrong. Binmore has a clear perception of what game theory *should* be,

[175] Binmore (1998), p.x.
[176] Binmore (1999b), p.119.
[177] Binmore (1992b), p.xxix.
[178] Binmore (1993), p.x.

and he devotes most of the space in his papers on foundations to showing where the traditional theory fails to live up to that ideal. Here, we group these issues around two themes; rationality and learning, the subjects, respectively, of the two subsections that follow.

2.2.1 RATIONALITY AS PROCESS

The purpose of Binmore's first salvo in 1984 was "to comment on the fundamental difficulties which exist in the foundations of the theory of extensive form games."[179] His concern is thus with games that have a sequential structure, where actions taken during play have the potential of revealing information.[180] This dynamic twist to his thought is important for understanding his work, as we shall have the opportunity to show at many points in the sequel.

If Binmore's early foray into the foundational field is largely uncited by him and others in the subsequent literature, it is mostly because the questions adressed there are taken up again, in much more detail, in a very important paper published in 1987 entitled *Modelling Rational Players*.[181] In 1984, his paper had been published in the Economic Journal, a leading journal in economics. In 1987, Binmore chooses a journal dealing with the philosophy of economics to publish the long,[182] literary, pugnaciously critical and rhetorical paper questioning "some of the shibboleths that underlie current game-theoretic research and that seem to me to obstruct further progress."[183]

In 1984, Binmore's prime worry regarding the foundations of game theory is fully formed. Namely, "the problems in justifying the use of the various equilibrium ideas can be traced to a considerable degree to the fact that the game structure has been over-abstracted. In particular, *the mechanism by means of which the players find their way to an equilibrium*

[179] Binmore (1984), p.51.

[180] Even if the solution concepts applied to games in extensive form generally specify equilibrium stategies decided upon before the start of play. In the interpretation of such games, questions regarding 'learning' during play can become important. In any case, they are important to Binmore.

[181] Binmore (1987b, 1988).

[182] About 80 pages, in two parts which differ in aim; the first is more critical, while the second attempts to find some positive solutions to problems outlined in the first.

[183] Binmore (1987b), p.180.

is typically abstracted away altogether."[184] The same issue is addressed again in 1987:

> Equilibrium is discussed without explicit reference to any dynamic process by means of which the equilibrium is achieved. Similarly, prior beliefs are taken as given, without reference to how a rational agent acquires these beliefs. Indeed, all questions of the procedure by means of which rational behavior is achieved are swept aside by a methodology that treats this procedure as completed and reifies the supposed limiting entities by categorizing them axiomatically.[185]

Binmore attributes most of what he considers to be game theory's woes to this essential defect. *Players*, their ability to reason, the sequences of thoughts (or lack thereof) which underpin their decisions and the sequences of actions (and thus potential 'mistakes') which forms their beliefs: all crucial elements which should be at the heart of game theory. And these are ill-served by imposing the impossibility of 'incorrect' (i.e. irrational) choices or beliefs, as Aumann does when he assumes a common prior, and then excludes from consideration states of the world in which players are not Bayesian rational. "This [method] abstracts away the question of *how* rational players reach their conclusions and seeks instead to characterize axiomatically *what* the conclusions must be directly."[186] To paraphrase what was said earlier about Aumann; to Binmore, *what* action is indeed rational is usually less important than *how it got* to be rational.

If traditional game theory's model of players who, *ex post*, are never wrong, is insufficient, then what should be done about it? Binmore's wish that the reasoning processes of players in games be modelled explicitly leads him to insist that "such an approach forces rational behavior to be thought of as essentially *algorithmic*. This makes it natural to seek to model a rational player as a suitably programmed computing machine"[187] If thought is algorithmic, then a decision is the result of a procedure, a finite computation that uses certain inputs and produces certain outputs. Players and their strategies, seen respectively as hardware and software, offer the possibility of retracing the steps that lead to their decisions; it then becomes possible to examine precisely *how* an equilibrium is *contructed.* Later, in his books on the social contract, these views lead Binmore to make broad claims about

[184] Binmore (1984), p.51, my it.
[185] Binmore (1987b), pp.180-1.
[186] Binmore (1992a), p.3.
[187] Binmore (1987b), p.181.

the basis of human consciousness. Yet, both arguments are founded on the idea of looking at players (in his game theory papers) or brains (in his philosophy) as types of computing machines.

Binmore, thinking generally about *types* of equilibrating mechanisms, devises a taxonomy of game-theoretic solution concepts to represent these. The latter is divided into what he calls "eductive" and "evolutive" types. That is, he distinguishes between equilibria that are achieved through chains of thoughts and counter-thoughts by rational players, and those brought about by evolutionary pressure, either by an exogenous 'fitness' selection mechanism or by simple types of learning by imitation. "The word *eductive* will be used to describe a dynamic process by means of which equilibrium is achieved through careful reasoning on the part of the players."[188] Binmore goes to the trouble of reassuring his readers that the word 'eductive' really does exist; it is "a well-documented word to be found in the concise Oxford dictionary. It is chosen to suggest the circle of ideas built into the words "eductive", "deductive" and "inductive"."[189] On the other hand:

> The word *evolutive* will be used to describe a dynamic process by means of which equilibrium is achieved through evolutionary mechanisms. It is intended to include not only the very long-run processes studied by evolutionary biologists [...], but also medium-run prosesses in which the population dynamics are not necessarily based on genetic considerations [...], as well as the very short-run processes by means of which markets achieve clearing prices [...]. The linking consideration is that adjustment takes place as a result of *iterated* play by *myopic* players.[190]

Binmore credits H. A. Simon for inspiring his distinction between eductive and evolutive processes. The latter spoke in terms of *substantive* and *procedural* rationality.

> The former is concerned only with finding what action maximizes utility in the given situation, hence it is concerned with analysing the situation but not the decision maker. It is a theory of decison environments (and utility functions), but not of decision makers. Procedural rationality is concerned with *how* the decision maker generates alternatives of action and compares them. It necessarily rests on a theory of human cognition.[191]

[188] Binmore (1987b), p.184.
[189] Binmore (1984), p.53.
[190] Binmore (1987b), p.184.
[191] Simon (1997), p.18. Earlier (see p.25), Aumann criticized Simon's 'satisficing' idea as being philosoph-

Simon, considering the repercussions of studying the *process* of rationality, suggests that future research goes in a direction that Binmore would support: "Does that lead to a theory of the mind? Yes, indeed it does. Unless we have a theory of how the human mind operates, we have few grounds on which to build an economic theory that will talk about the kind of uncertain world we live in".[192] However, Simon's theory of the mind is based on his research in Artificial Intelligence, while for Binmore, such a theory should be based on the mathematics of computation.

As noted earlier, during his career, Binmore is concerned almost solely with games with dynamic aspects. The word 'dynamic' even gets into his definition of 'eductive'. In this way, Binmore gives to games normally seen as static a form of temporal structure that consists of sequences of *thoughts* that occur before the game, instead of actions during play. In contrast, Aumann's attempts to provide foundations for game theory, apart for his paper on backwards induction,[193] are concerned with static solution concepts.

> Such [eductive] considerations are *internal* to a player. The dynamics of an eductive libration are therefore invisible to an observer. But the point of view we are pushing here insists that this is no good reason for proceeding as though the dynamics were absent altogether. [fn: It is not denied that common knowledge conditions on beliefs allow a satisfying static analysis of certain static games. But an evasion of the study of the process by means of which beliefs are formed leads to difficulties for dynamic games.][194]

Thus if solution concepts are 'only' suitable for static games, that is not enough. To him, the study of dynamics is the only road that leads to insight about learning.

The two types of equilibriating mechanisms are, in fact, related.

> Of course, the distinction between an eductive and an evolutive process is quantitative rather than qualitative. In the former, players are envisaged as potentially

ically attractive yet poor in terms of the results it yields. Here Binmore praises a two-component theory classification scheme due to Simon which would, from Aumann's point of view, leave virtually *all* of the models of economics in the first basket, and almost none in the second. However, by speaking of 'evolutive' processes Binmore opens the door to models in which agents process no or very little information. Simon probably had something richer in mind with his 'procedural' rationality.

[192] Simon (1997), p.26.

[193] Which, with this in mind, not surprisingly is the one that provoques the most public clash between the two authors.

[194] Binmore (1988), p.11.

> very complex machines (with very low operating costs) whereas, in the latter, their internal complexity is low. It is not denied that the middle ground between these extremes is more interesting than either extreme. However, only isolated forays have so far been made into this area.[195]

Although both parts of Binmore's *Modelling Rational Players* are concerned with the thought processes of rational players, and thus with eductive theory, he makes no secret of his own sympathies. Already in 1984 he states that it is the field of evolutive theory:

> Which is of greatest significance in so far as applications in the foreseeable future are concerned. [fn: And possibly in so far as any considerations are concerned if one takes the view that a "correct" reasoning process is one that cannot be bettered by any alternative reasoning process.][196]

Note how relativistic his idea of rationality is. One does not *postulate* rational behavior, one *observes* entirely context-dependent instances of it in action. And again, in 1987:

> But a stricly eductive environment is seldom encountered in the real world. What intuitions we may glean about "reasonable" properties of equilibria from empirical observation are therefore contaminated by the intrusion of evolutive factors. My view is that much of the confusion concerning equilibrium ideas in game theory can be traced to this fact. Intuitions which belong in an evolutive theory have been defended from an eductive position. In offering unorthodox views on eductive issues, I am therefore not necessarily denying the validity of such intuitions: I am merely suggesting that they be properly recognized as evolutive.[197]

Binmore never strays from evolutionary thinking for very long. As in an eductive context he has in mind not one ideal of a rational player but many distinct types of 'rationalities' trying to 'outthink' each other, the question of *selecting* the 'best' players from this pool crops up naturally. And Binmore, as he did at the end of the last quote, does not hesitate to use the opportunity to bring back eductive processes to evolutive ones:

> At the risk of confusing matters hopelessly and completely, I [...] propose to argue that, even in an *eductive* context, evolution has a central role to play. One might even take the view that, without some evolutionary story in the

[195] Binmore (1987b), p.185. We shall see that the constraints of evolutionary games makes for very stupid players. Exploring the 'middle ground' proves very difficult.

[196] Binmore (1984), p.56.

[197] Binmore (1987b), p.185.

> backround, rationality, in the sense required for an eductive analysis, makes little sense. Given several rival models of rationality, how is one to make a choice?[198]

Binmore leaves very little room in game theory for a purely eductive theory of equilibrating processes. What he demands from such a theory may be unreachable.[199] This could both help to explain and be explained by his obvious and declared preference for evolutionary models of equilibriating processes. While his work in evolutionary games is positive and creative, much of his writings on eductive games[200] has a strongly polemical flavour. In fact, it can seem as though an important aim of his writings on eductive theory is to use its capacity to focus attention on the *process* of rationality to highlight the shortcomings of traditional game theory. For example, he writes:

> In what follows, it is heretically argued that *none* of these [that is "the principles of "backwards induction", "the successive elimination of dominated strategies", and "the equivalence of games which can be mapped into one another by "strategically inessential transformations"""] is acceptable in a stricly eductive context.[201]

The field of evolutionary games was relatively new and undeveloped in the beginning of Binmore's career in game theory. This meant that much of what he contributed there was in some way 'unconstrained' by past theories and accomplishments. Most of game theory and economics, on the other hand, falls under the broad heading of 'eductive'. Here, Binmore, who obviously prefers to blaze new trails rather than follow established roads, will slash away mercilessly at the status quo with which he is profoundly dissatisfied, with little regard for past accomplishments.

---

198 Binmore (1988), p.16.

199 This will be apparent in his debates with Aumann on the principle of backward induction. See Chapter III.

200 Exceptions are provided in Binmore (1988) and Binmore and Herrero (1988).

201 Binmore (1987b), p.185. All the principles he mentions here are fundamental in traditional theory.

2.2.2. LEARNING AS SURPRISE

> Self-evidence [...] can become dispensible in logic, only because language itself prevents every logical mistake. -What makes logic a priori is the *impossibility* of illogical thought.
>
> Ludwig Wittgenstein[202]

Binmore mounts another offensive against the general frame of game theorists' practice, attacking that which he terms 'Bayesianism'. The term is coined by Binmore (1991) to characterize the types of models that Aumann analyses in the foundational literature. These models, which take as starting points 'universes' composed of states of the world which together enumerate every possible and imaginable occurrence, give players, through probability distributions and information partitions defined on the totality of this universe, potential, or 'informal'[203] knowledge about *all* aspects of their 'world'.

> It [...] needs to be explained that [his criticism] is not an attack on Bayesian decision theory as commonly used in analysing particular games. I am a Bayesian myself in such a context. The paper is an attack on *Bayesianism*, which I take to be the philosophical principle that Bayesian methods are always appropriate in all decision problems.[204]

For "just as a mathematician can use Pythagoras' theorem without subscribing to the Pythagorean metaphysics that apparently made pure number the essential stuff of the universe, so one can be a Bayesian without being a Bayesianismist."[205] Bayesianism, as he sees it, is "a metaphysical doctrine that hinders advances in the foundations of game theory."[206] "How Bayesianismists manage to convince themselves that this trivial algebraic manipulation somehow unlocks the secret of rational learning is beyond my comprehension."[207] He admits, however, that one will not find a 'Bayesianismist' manifesto among game theorists, in which this "creed" is explicitly laid out and defended.

[202] Wittgenstein (1961), p.57.
[203] See pp.27-8, where Aumann discusses this point.
[204] Binmore (1991a), p.1.
[205] Binmore (1993), p.306.
[206] Binmore (1991a), p.20.
[207] Binmore (1993), p.306.

> But I do not think I am merely attacking a straw man. What matters for this purpose is not so much what people say about their philosophical attitudes, but what models they choose to construct. As Robert Aumann likes to say of game theoretic concepts in general: By their fruits shall ye know them.[208]

Binmore's principle line of attack, while not yet fully pursued at that time, was already stated in 1984. "Bayesian decision theory is a "closed universe" theory. It takes for granted that we can describe our fundamental domain of uncertainty completely, precisely and with no prospect of revision in the light of new data."[209] As we shall see, what engages Binmore is how the word "new" in the last quotation should be interpreted. The bone he picks with 'Bayesianismists' stems from what he perceives to be their interpretation of it.

Recall, from the earlier sections on Aumann, that Savage (1954) also dealt with a *universe* consisting of *states of the world* that encompassed *all* that is subject to uncertainty for a decision maker. The difference in Aumann's framework is that states of the world are *all-inclusive*, that is, they also include the actions of players. This is necessary because of the multi-agent interaction that takes place in games; as the actions of opponents are sources of uncertainty for a given player, these must be exhaustively catalogued in the states of the world he considers in his decisions. But in order that players, roughly, 'live in the same universe', that is, be capable of assessing the probabilities of similar events (i.e., those belonging to the same game), a player will examine states which contain his own actions as well, even though they are there in some sense 'for the sake of the others'. Savage, considering only single-person decision problems, could work with *acts* defined as functions mapping states of the world into consequences. States were completely exogenous to agents, and consequences were the result of combining 'nature' with a 'decision' from the agent.

On the issue of 'small worlds', Binmore refers "to the scriptures"[210] of Bayesianism, that is, to Savage (1954) himself. He remarks that Savage was fully aware of the difficulties involved in granting decision makers the capacity to foresee, and thus discount the impact of, any possible future events.

> As has just been suggested, what in the ordinary way of thinking might be

[208] Binmore (1991a), p.1.
[209] Binmore (1984), p.53.
[210] Binmore (1991a), p.2.

> regarded as a chain of decisions, one leading to the other in time, is in the formal description proposed here regarded as a single decision. [...] Carried to its logical extreme, the "Look before you leap" principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not -as some might think- because there might later be cause for regret, if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled by human possibility.[211]

Savage did believe that considering 'small worlds' was valid for "relatively simple problems of decision",[212] although he did not attempt to give a clear-cut process for deciding which problems were "relatively simple". He thought that this was maybe best seen as "a matter of judgement and experience".[213] For Binmore, problems regarding knowledge, belief and rationality in games cannot be treated in 'small worlds', or as he calls them, 'completable universes'.[214] Surprises, mistakes and ignorance are central to his views.

In fact, for Binmore, Bayesian updating in 'small worlds' should not even be referred to as a learning process. Since, in such a world, a decision maker can set up lists of all possible strings of events, he can, while deliberating *before* taking his decision, ask himself: "For every conceivable possible course of future events, what would my beliefs be *after* experiencing them?" However, "such an approach automatically discounts the impact that new knowledge will have on the basic model used to determine beliefs: i.e., it eliminates the possibility of being *surprised* by an event whose implications have not previously been considered."[215] Suppose that, upon examining this large mishmash of interwoven beliefs, the decision maker notices some inconsistencies, that is, some beliefs that violate one or some of Savage's axioms. In that case, should he wish to be consistent, he should adjust his beliefs until no further violations of consistency persist. Exactly *what* changes are made is a 'personal' matter, they depend on an act of introspection by the decision maker with regards to his preferences and judgements of likelihood. "With Savage's definition of consistency, this is equivalent to asserting that the adjusted system of contingent beliefs can be deduced,

[211]Savage (1954), pp.15-16.
[212]Savage (1954), p.16.
[213]Savage (1954), p.16.
[214]The reason for such a change in terminology will be made clear in Section 2.5.
[215]Binmore (1987b), p.210.

using Bayes' rule, from a single prior."[216] That is, Savage shows that if his axioms are accepted, then a personal probability distribution can be defined on the events of a space of states of the world. In that case, decision makers can be said to 'possess' a prior probability distribution on the possible events, which, when the decision maker is faced with realized events, can be updated via Bayes' rule.

Binmore doubts the interpretative value in terms of learning of this type of adjustment process. To him, it all sounds rather fabricated, orchestrated. "At the end of the story, the situation is as envisaged by Bayesianismists: the final "massaged" posteriors can indeed be formally deduced from a final "massaged" prior using Bayes' rule."[217] However, the natural interpretation of 'priors' and 'posteriors' are jostled about and turned on their heads in the works of "a complex adjustment process that operates until consistency is achieved."[218] The process is not 'open-ended' enough for him. *In advance* of the decision, events are looked at from 'two sides', before and after, and then beliefs are adjusted so that both perspectives concord. But that is one facet too many for him.

> What is certainly false in this story, is the Bayesianismist view that one is *learning* when the massaged prior is updated to yield a massaged posterior. On the contrary, Bayesian updating only takes place *after* all learning is over. The actual learning takes place while the decision-maker is discounting the effect that possible future surprises may have on the basic model that he uses to construct his beliefs, and continues as he refines his beliefs during the massaging process. Bayesianismists therefore have the cart before the horse. Insofar as learning consists of deducing one set of beliefs from another, it is the massaged *prior* that is deduced from the unmassaged *posteriors*.[219]

This world in which all thinking takes place *before* the action is entirely unappealing to him. It evacuates all dynamics from the learning process, since the aspect of this 'story' which he considers to be important, namely the interpretation of the acts and judgements through which consistency in beliefs is achieved, is considered to be *completed before* the start of the game. What interests Binmore is *where* the subjective prior distribution comes from, how it was *formed*. And as he believes that the steps preceding a decision, an action, should

[216]Binmore (1991a), p.5.
[217]Binmore (1991a), p.5.
[218]Binmore (1991a), p.5.
[219]Binmore (1991a), p.5.

be modelled explicitly, he also holds that the origins of beliefs should also be modelled. Without which:

> Naive Bayesian rationality apparently endows its fortunate adherents with the capacity to pluck their beliefs from the air. But this will not do for game theory. One might almost say that what game theory is *about* is the massaging process (via "if I think that you think..." arguments) by means of which beliefs are constructed.[220]

To Binmore, when one ventures into the radically new uncertainty of an 'open' world, consistency loses some of its normative appeal. For

> after all, scientists are not consistent, on the grounds that is is not clever to be consistently wrong. When surprised by data that shows current theories to be in error, they seek new theories that are inconsistent with the old theories. *Consistency, from this point of view, is only a virtue if the possibility of being surprised can somehow be eliminated.*[221]

Here again we find in Simon's work remarks closely consonant with Binmore's. After giving examples of the progression of models in economics designed to explain the same phenomena, Simon argues that just as economists change their minds as to the relevant factors that should be included in a model, "in the course of history economic actors may change their ways of looking at choice situations, [...] and their decisions may thereby change."[222] Thus if acquiring knowledge through mistake and basic reassessment is good enough for scientists and scholars, why should it not be good enough for agents in a model?

It should now be plain that for Binmore, "learning" is understood "in the sense of "adding to one's understanding" rather than simply "observing what happens"".[223] That is, learning is intimately tied to error, but even more so, to error concerning 'truths', or 'true' events, that cannot even be imagined before their occurrence makes their existence undeniable. One is led into an unending chain of uncertainty and essentially guaranteed ignorance, as all 'corrected' theories are just as fragile and open-ended as their predecessors.

One may ask: 'How could such ideas possibly be *formalized*?' One can understand the irritation of someone like Aumann when faced with such criticism, however appealing

[220]Binmore (1987b), pp.211-2.
[221]Binmore (1991a), p.4, my it.
[222]Simon (1997), p.22.
[223]Binmore (1991a), p.5.

it may seem from a philosophical point of view. The gap is truly wide between Binmore's *wishes* and what, as we will hear him say later, is "a practical possibility at this stage of the art." We shall review his positive work later, much of which is in evolutionary games. Agents in such a framework 'learn' quite well in Binmore's open-ended sense of the word, but only because they cannot *think* (much less entertain counter-thoughts of an infinite extent). Every step in the game comes as a complete surprise to them. They are entirely incapable of behaving *strategically*.

2.3 THE TURNCOAT MATHEMATICIAN

Binmore often takes issue with the crisp-toned and formalized papers that appear in the leading journals of economics and game theory, at least with regards to the foundational questions of game theory. He opposes the position that they implicitly defend concerning the place, the importance and the use of mathematics in game theory. He takes aim at Aumann who he considers to be a representative example:

> The character of Aumann's mathematics is beside the point. We focus on the wrong question if we look for an error in how Aumann deduces his conclusion from his hypotheses. An axiom-definition-theorem-proof format is designed to close the mind to irrelevancies so that attention can be concentrated on the issues that really matter. But if an inappropriate formalism is chosen, one necessarily closes one's mind to issues that it is perilous to neglect.[224]

Thus, "the result is the construction of magnificient mathematical edifices of which a medieval scholastic might justly be proud, but little in the way of genuine progress."[225] Within game theory, Binmore fears the spread of a "serious disease that manifests itself as a worship of mathematical formalism".[226] In his day, Poincaré had also mocked the recourse to formalism of some of his contemporaries:

> They have accumulated formulae and they have thought to free themselves from what was not pure logic by writing memoirs where the formulae no longer alternate with explanatory discourse as in the books of ordinary mathematics, but where this discourse has completely disappeared.[227]

[224] Binmore and Samuelson (1996), p.111.
[225] Binmore (1988), p.10.
[226] Binmore (1991a), p.8.
[227] Poincaré (1905), p.1023.

Binmore confesses to not always having been so rebellious when dealing with mathematics: "As a mathematician, I used to write in a very formal style, and it may be that, like many converts, I have allowed the pendulum to swing too far in the other direction."[228]

Binmore takes a very critical stance towards standards of mathematical rigour and aesthetics when these are employed to judge the success and 'worth' of game-theoretic models. For example, considering Kohlberg and Mertens' (1986) concept of *Strategic Stability*, he says that "although [their] framework [...] is certainly *mathematically coherent*, I would argue that it is far from being *conceptually coherent* in that the *plausibility* of their mathematical criteria is hard to evaluate."[229] In another case, discussing the differences between the Nash and Kalai-Smorodinsky bargaining solutions, Binmore asserts:

> Kalai and Smorodinsky (1975) were [...] tentative in the claims they made for their solution. Their aim was simply to point out that axiom systems other than Nash's (1950) can lead to bargaining solutions that are no less elegant from the mathematical point of view than the Nash bargaining solution. Once this has been pointed out, even mathematicians have no choice but to ask interpretive questions about the *meaning* of rival axiom systems.[230]

Or again, referring to games where credible 'threats' are allowed: "This is such an elegant piece of work that it is an enormous pity to have to say that it has little or no relevance to practical questions."[231]

In judging game theory, mathematical coherence, elegance and axiomatics should therefore not be placed before 'relevance'. Other mathematical benchmarks should also be jettisoned:

> Consider, for example, *existence*. This is regarded as a *sine qua non* for an equilibrium notion by those brought up in the Bourbaki tradition. But evolutionary stable equilibria do not always exist. Is the idea therefore to be abandoned? Clearly not.[232]

[228]Binmore (1993), p.viii. On another occasion he writes: "Mathematicians will recognize this result as saying that Abel summability implies Césaro summability. It seems a long time ago that I wrote a thesis on such arcane matters!" (Binmore (1998), p.110)

[229]Binmore (1988), p.39, my it.

[230]Binmore (1998), pp.82-3.

[231]Binmore (1998), p.108.

[232]Binmore (1987b), p.180.

As we shall see Aumann do in the case of the core later on, Binmore goes on to give a loose interpretation of the significance of the formal existence of evolutionary stable equilibria. It will not be explored here.[233] What is of interest here is that Aumann could call into question the relevance of the mathematical requirement of *existence* in game theory while interpreting the core, which in general does not always exist. That he will not, and that Binmore does in a remarkably similar context, is noteworthy.

Not only should mathematical standards not always be accorded priority relative to other criteria, but sometimes 'too much' mathematics can hinder comprehension. Thus Binmore hopes that an essay on foundations written with Adam Brandenburger[234] "will serve to demystify a subject that has already become overburdened with unnecessary formalism and misleading jargon."[235] At other times, Binmore implies that game theorists are taking 'the easy way out' by reverting to sophisticated formal theories.

> In brief, I think that the backward induction problem -like much else in the foundations of game theory- poses only a very small challenge to our powers of formal analysis. The real challenge is [...] to our ability to find tractable models that successfully incorporate everything that matters.[236]

He rejects applying Bourbaki's (1948) defense of the 'ease' provided by common mathematical structures to the foundations of game theory. There, Bourbaki wrote:

> The "structures" are tools for the mathematician; as soon as he has recognized, among the elements which he is studying, relations which satisfy the axioms of a known type, he has at his disposal immediately the entire arsenal of general theorems which belong to the structures of that type. Previously, on the other hand, he was obliged to forge for himself the means of attack on his problems; their power depended on his personal talents, and they were often loaded down with retrictive hypotheses, resulting from the peculiarities of the problem that was being studied. One could say that the axiomatic method is nothing but the "Taylor system" for mathematics.[237]

Thus, formalisation should proceed no further in game theory, as treating questions

[233] Although evolutionary stable equilibria will be presented in Section 2.6.1.
[234] Binmore and Brandenburger (1990).
[235] Binmore (1990c), p.viii.
[236] Binmore (1997), p.2.
[237] Bourbaki (1948), p.1272. Note also Bourbaki's disapproval of the 'patchwork' era, where new concepts were very personalized and their relationship with anterior work was not clear. But in some sense this is what Binmore, through his distrust of Aumann's 'unifying' approach, advocates for game theory.

regarding knowledge and rationality "axiomatically a la Bourbaki has got about as far as it can go."[238] And if that means that game theorists need to start working with patchy, 'improvised' models, then so be it. "I would prefer to work with a game theory that has no foundations at all, than to operate using foundational principles based on a flawed methodology."[239] "Informal models are not to be despised."[240]

This idea of a need to 'stop' the impressive progress of game theory, in the sense of pausing to consolidate and mull over the paths taken, with the prospect of abandoning some of them, comes up in many different forms in Binmore's writings.

> Recent advances in game theory have ensured its recognition as a key subject across a wide range of disciplines in the social sciences. But, as is often the case when advances come quickly, scant attention has been paid to the foundations on which the theory is being erected.[241]

Some types of game theory, according to Binmore, are a waste of time. Game theorists should therefore be exhorted *not* to work in these 'bourbakist' fields. The spirit of this is in some ways similar to an attitude towards the foundations of mathematics that irritated Hilbert considerably; that of putting barriers to mathematical creation based on *a priori* philosophical and personal beliefs. Hilbert accused opponents of Cantorian set theory of falling prey to "the expedient of prohibitions, of dictatorship."[242] Thus, of Kronecker: "Of the intuitive methods of Riemann, which at the time had the most splendid success, or of the newly-arisen Cantorian set theory, he wishes to know nothing; he remains closed to their accomplishments. Here he pursues an ostrich-politics."[243] And as for Poincaré, "on these matters he conducted himself in a manner that was merely carping, negative, wholly unproductive. He produced no new ideas, and the new, fruitful scientific approach of Cantor he branded as "Cantorism". Like Kronecker, he dictated prohibitions."[244] For Hilbert, his struggle against 'prohibitionism' aims to defend intellectual freedom in mathematics:

> We rather say that the prohibitions must be formulated in such a way that

[238]Binmore (1992b), p.2.
[239]Binmore (1991), p.3.
[240]Binmore (1984), pp.51-52.
[241]Binmore (1990c), p.vii.
[242]Hilbert (1920), p.943.
[243]Hilbert (1920), p.944.
[244]Hilbert (1920), p.945.

> the contradicitions are eliminated but everything valuable remains- and not only must all the valuable results remain standing, but the freedom of concept-formation and of the methods of inference ought not to be limited beyond what is necessary.[245]

Aumann's views on the progress of science is akin to this. As one never knows which concepts, after many applications, will have proved their 'usefulness', the *a priori* restraints on research should be kept to a minimum.

Binmore feels that the analytical tools and axiomatic structure of game theory have in some way 'outpaced' and 'overtaken' the discipline's basis of ideas, concepts and intuitions. In fact, a 'healthy' state of affairs in a foundational enterprise should have things the other way around, as was the case, Binmore contends, in laying the foundations of mathematical analysis. In that case:

> Mathematicians like Euler discovered many wonderful results without agonizing over precisely what real numbers are, or what it means for a series to converge. [...] Later mathematicians, like Dedekind or Weierstrass, put analysis on a proper basis by showing how real numbers can be constructed from the rationals, and how sentences containing the word "infinity" can often be translated into carefully worded statements in which only finite magnitudes occur. Their approach was therefore *reductionist* or *constructive*. Only later, after their innovations had been thoroughly absorbed by the mathematical community, did coherent *axiomatic* treatments of the concepts become possible.[246]

For Binmore, game theory should be entering into the second of these three stages. As he puts it, strongly, "too many of our *wonderful* results are *wrong* for foundational issues to remain neglected."[247] Note that at the same time he questions (again) the relevance of aesthetics in evaluating game theory's achievements. Thus, "perhaps the time has come to embrace the reductionist methodology of Dedekind or Weierstrass wholeheartedly and to stop seeking intellectual short-cuts that are unlikely to be found."[248] To Binmore, a former mathematician, the abundance of formal methods in game theoretic papers can in some ways

[245]Hilbert (1920), p.945.

[246]Binmore (1992b), p.2.

[247]Binmore(1992b), p.2, my it. He means 'wrong' from an intuitive point of view. This could be misunderstood, as is sometimes done by Aumann in his exchanges with Binmore, as saying that the *mathematics* are wrong.

[248]Binmore (1992b), p.2.

be seen, not as illustrating the theory's success and sophistication, but rather the poverty of its ideas and the unwillingness of its practitioners to stray from the safety of a well-beaten path. Comparing the hold mathematical formalisms have on game theory to that which *newspeak* had on the characters of Orwell's 1984, he states that "the trap is to proceed as though anything that is not expressed in the formalism to which he is accustomed does not exist at all."[249]

2.4 A LITTLE PHILOSOPHY HERE, A LITTLE PHILOSOPHY THERE, A LITTLE PHILOSOPHY EVERYWHERE!

References to 'philosophy', in many forms, abound in Binmore's work. 'Philosophy' is to be understood widely here, as both a concern for the problems dealt with by professional philosophers and as a taste for general conceptual discussions. Binmore (1987b) states clearly that he wants to write "a philosophical piece about the foundations of game theory".[250] Nor does a former professional mathematician, now game theorist and economist, feel the need to be apologetic about devoting much time to philosophical matters. For "at root, the difficulties are philosophical in that they arise from the manner in which the nature of the problems to be resolved is perceived."[251] In an "apology" at the start of the first volume of *Game Theory and the Social Contract*, Binmore begs the forgiveness of his more formally-minded collegues for often preferring long and tortuous sequences of words to clean strings of symbols:

> In so far as apologies are necessary on the mathematical content of the book, they need to be tendered to mathematicians, and in particular to mathematical economists. [...] It is certainly correct that the truly formal parts of the discussion are infrequent and that these are patched together with appeals to intuition that some mathematical economists will find irritatingly imprecise. If these appeals to intuition were capable of being defended by obvious formal arguments, such irritation would be unjustified. But this will seldom be the case. It will not even always be "obvious" what the formal setting for a rigorous discussion would be. I agree that the need to reach a wider audience is not an adequate excuse. But I don't think this book would have got written if I had tried to do a better job on the formal side. Nor do I have much sympathy with the view

[249] Binmore (1991a), p.8.
[250] Binmore (1987b), p.179.
[251] Binmore (1987b), p.180.

> that one should either say something formal or else say nothing at all.[252]

These do not seem like the words of a man much burdened with a guilty conscience. In fact, after similarly apologizing to political philosophers for not being 'one of their own', and to many others for various reasons, he admits that "in spite of all these apologies,the truth is that I do not really feel very apologetic."[253] Some years later he confesses that "for what it is worth, *Just Playing* [the second volume of his work on philosophy] is my *magnum opus*".[254]

In the two subsections to follow, we shall distinguish two uses in game theory for which Binmore has philosophy in mind. The first is to be seen in the context of a broadening of the scope of applications of game theory advocated by Binmore. The second is in his insistence that models and their assumptions be critically and extensively discussed.

2.4.1 IF YOU WANT SOMETHING DONE RIGHT, GET YOURSELF A GAME THEORIST

In one of his papers, before even starting the actual analysis of a model, Binmore downplays the shortcomings of his innovative approach by abandoning these at the doorstep of philosophy: "Since sharp answers [to the questions he wishes game theory to address] woud provide an incidental resolution of the problem of scientific induction, [their not being solved in his paper] is perhaps not surprising."[255] Nevertheless, Binmore sees in game theory a tool which *could* successfully be brought to bear on these longstanding philosophical problems, and he suggests that game theorist *should* probe these questions, for the sake of the theory's progress.

> A properly founded game theory would have aswers to questions like: *What is the self? What do we have in common with others? What does it mean to know something? How do we learn? How should we learn?* Right now we do not even know to what extent such questions are meaningful. But genuine progress is unlikely if we continue to regard the problem of scientific induction or the

[252]Binmore (1993), p.viii.
[253]Binmore (1993), p.xi.
[254]Binmore (1999b), p.136.
[255]Binmore (1988), p.13.

problem of personal identity as difficulties best left to philosophers.[256]

Very broad and very ambitious questions are put to game theory elsewhere:

> What follows in this section clearly has wider implications than the game-theoretic applications for which it is proposed. In particular, it has some significance for the software-hardware approach to the mind-body problem [...], and some relevance to the explanations advanced by some biologists for the evolution of self-consciousness in humans. However, I have nothing particularly original to offer on these general issues. The aim is to codify what seems to be an emerging consensus on these questions, in a manner that, although far from formal, is sufficiently precise to allow the possible applications in game theory to be sensibly evaluated.[257]

Experimental and theoretical psychology are also ripe to provide game theory with applications. In his opinion, there do not "seem [to be] grounds for anything but pessimism about the prospects for fundamental advances in psychological theory in the foreseeable future."[258] He explains:

> "A prime difficulty is that we know very little about how people think and learn. Introspection curiously provides little guidance in such matters. Psychologists offer various theories, together with greater or lesser quantities of evidence in their support, but those theories that are formalizable strike most game theorists as being too mechanical to reflect the way people really think in game-like situations. Any theory that offers no role for looking ahead and anticipating the actions of others is clearly inadequate for game-theoretic purposes. In constructing models, one therefore works largely in the dark. My feeling is that the time is past when experimental work in this area can be left to psychologists. They wear different blinders to economists that focus their attention on different matters and lead them to construct different research agenda. If learning models are to be properly informed by empirical data, I suspect economists will need to free themselves of their traditional prejudices by running experiments of their own.[259]

And it is precisely when game theory tries to tackle these new problems that 'Bayesianism', 'completable universes' and all the evils of 'Aumann's method' become a problem:

> That is to say, Aumann's system works well in providing foundations for that part of game theory *as-it-is* which holds together fairly coherently. Where it

[256] Binmore (1992a), p.2, my it.
[257] Binmore (1988), p.20.
[258] Binmore (1990b), p.14.
[259] Binmore (1992a), pp.24-5.

> does not work so well is in dealing with those problems that have reduced game theory as-it-is to a state of confusion and disarray: namely, the problems of equilibrium selection and equilibrium refinement.[260]

It is on these latter problems that "progress is necessary if game theory is to break out of the beachhead it has established in the social sciences."[261] Poincaré had similarly questioned the relevance of Cantorian set theory and formal logic by questioning their relevance to 'real-world' applications:

> So he asks me if I have found the [...] error among the orthodox. No, I have not seen it in the pages I have read; I know not whether I should find it in the three hundred pages they have written which I have no desire to read.
>
> Only, they must commit it the day they wish to make any application of mathematics. This science has not as sole object the eternal contemplation of its own navel; it has to do with nature and some day it will touch it. Then it will be necessary to shake off purely verbal definitions and to stop paying oneself with words.[262]

For Binmore, the progress of game theory is intimately tied to the drastic broadening of the scope of its applications. He is, in a sense, game theory's zealous reformer, crusader and missionary. Although he may be very critical of game theory's 'orthodoxy', he constantly tries to reach out to and convert novices and outsiders. When addressing them he makes it clear that his ends are to be achieved from *within* the theory's fortress; even if an unsettlingly noisy one, he is a hen in the henhouse and not a wolf stalking the roost.[263] Indeed, even though his comments about the theory can be quite damning, "this is not to say that game theory has nothing to contribute to the major debates of our time. On the contrary, it is inconceivable that these debates will ever be conducted in a remotely scientific

[260] Binmore (1992a), p.14. "Aumann's system" in this case refers to the whole 'epistemic' approach to foundations.

[261] Binmore and Samuelson (1992), p.282. Binmore speaks here of the social sciences, although he would also extend this to the fields reviewed above.

[262] Poincaré (1906), p.1058.

[263] "The things that excite game theorists are diverse and complicated -so much so that outsiders often notice only the issues that divide them, because these are the issues about which game theorists want to talk. And when they talk, gamesters talk a great deal, since they are a quarrelsome and unruly breed. However, there are good reasons why they do not split into rival clans. Underlying their disputes and differences of approach there is a deeply felt sense of common purpose that holds game theorists together." (Binmore, Kirman and Tani (1993), p.2)

manner without the intervention of game-theoretic ideas."[264] In other words, "I believe that there is "only one game in town", and that we had better learn how to play it or shut up altogether."[265]

Interestingly, when 'colonizing' other fields of social study, Binmore does stress the fundamental role played by mathematics in game theory. While he admonishes professional game theorists for overemphasizing mathematical criteria for judging the *results* of game theory, he makes it clear to beginners that the logic and clarity of a mathematical *approach* to social science is indispensible.

> Thinking seriously about game theory requires some experience, knowledge, and sympathy with the manner in which mathematicians reason. This does not mean sophisticated knowledge of specific mathematical techniques is required. [...] What is necessary is an understanding of the *nature* of a mathematical argument: in particular, the necessity in a mathematical argument of sometimes taking, as hypotheses, conditional statements which one might very well suspect of being false.[266]
>
> Unfortunately few social scientists seem to properly understand or appreciate either the aims or the techniques of applied mathematical reasoning. On the one hand, there is a tendency to apply simple models without much consideration to complex situations for which they are ill suited. On the other hand, there is a tendency to denigrate mathematical thinking altogether.[267]

In these discussions with social scientists and philosophers, we find Binmore defending the use of certain methods on grounds of mathematical simplicity. Referring to the modern economists' approach to utility via preference relations, he says that "such a sophisticated view makes it hard to measure payoffs in real-life games, but its advantage in keeping the logic straight is overwhelming."[268] Axiomatic thought, as well, is stressed by Binmore:

> Indeed, the major motivation for expressing intuitions concretely within a formal model is to discipline our wilder flights of fancy, by exposing the incomplete

[264] Binmore (1990b), p.5.

[265] Binmore (1990c), p.viii.

[266] Binmore (1990b), p.4.

[267] Binmore (1990b), pp.4-5.

[268] Binmore (1993), p.98. Simplicity, as Hilbert might say, is just good mathematics. "An old French mathematician said: "A mathematical theory is not to be considered complete until you have made it so clear that you can explain it to the first man whom you meet on the street." The clearness and ease of comprehension, here insisted on for a mathematical theory, I should still more demand for a mathematical problem if it is to be perfect; for what is clear and easily comprehended attracts, the complicated repels us." (Hilbert (1900), p.1097)

> and inconsistent nature of the castles in the air that our imaginations so freely construct. [...] In the moral sphere, where we are much more likely to be led astray by our prejudices, it seems to me that there is even less room for laxity. [...]
>
> What constitutes rigor in moral geometry? In such a context, practical reasoning matters as much as pure reasoning. Axioms for rational decision-making must therefore take their place alongside the axioms for logic and mathematics.[269]
>
> [...] What makes us [...] geometers is not our *conclusions*, but our common understanding of the geometrical *ethos*.[270]

Binmore can be quite brusque in his treatment of scholars in social sciences or philosophy who turn a deaf ear to his gospel. He invariably faces stiff resistance during his incursions into these fields. In these moments, Binmore lectures these heathens with terms and a spirit which are evocative of Aumann's reprimands towards him in the more restricted field of game theory. He makes light of some of his non-mathematician opponents, with a slight tone of exasperation reminiscent of Aumann's reactions to 'informal' arguments, or to Binmore himself.

> The charge of logical incoherence has actually been put to me in public by more than one political scientist. These occasions can be embarrassing, since debating logic with an untrained opponent is like dueling with an enemy determined to impale himself on one's sword. [...] I must therefore sometimes defend propositions that everybody, myself included, would prefer to be false. But the fact that one does not like a conclusion does not make the argument that leads to it a *reductio ad absurdum*.[271]

As Binmore's opponents seem mostly unversed in mathematics, his counterattacks tend to turn more quickly to ridicule. Under a engraving by William Blake[272] which "shows some of Plato's students exercising the mathematical skills without which they were denied entry to his Academy", he suggests that "modern writers of footnotes to Plato might usefully go and do likewise."[273] Similarly, he recounts episodes where he is questioned about using game theoretic models based in discrete time:

[269] Binmore (1993), pp.319-20.

[270] Binmore (1993), pp.9-10. In section 2.5, we shall see Binmore, again, praise the role of mathematics in game theory, but in a sense different than here.

[271] Binmore (1998), pp.vii-ix.

[272] These are littered accross the two volumes of his philosophical work.

[273] Binmore (1998), p.532.

> Time is continuous, so the story goes, and it is therefore a mistake to treat the case when time moves in discrete jumps as fundamental.
>
> It is not easy to respond in a manner that such critics can understand. In particular, drawing attention to the way physicists do things cuts no ice with critics who have little or no scientific training. Still less does it help to suggest that the metaphysical nature of time is open to debate. Usually it is thought that a joke is being made. However, I have sometimes found it effective to point out that a derivative is *defined* as a limit. In writing down a differential equation, one is therefore making a statement about the manner in which certain limits are related. If one wants to get these relationships right, one therefore has little choice but to look at the limiting procedures that led to them.[274]

These passages where Binmore is defending his approach to philosophy against attacks by social scientist that have no mathematical training displays some points of contact with Aumann's defense against Binmore's charges. In both cases, the accused is countering indictments which attempt to undermine the *relevance* of the entire *underpinnings* of his creations. The opponents are in essence telling them that they don't really understand why they even bothered to expend so much thought on so hopeless an enterprise. If one spends *years*[275] working on what one considers to be novel and interesting approaches, and if opponents' charges are seen as chimerical at best, their frustration is easily understood.

2.4.2 BOYS, DON'T LET THOSE ASSUMPTIONS OUT OF YOUR SIGHT

> But I cannot stand forward, and give praise or blame to anything which relates to human actions, and human concerns, on a simple view of the object, as it stands stripped of every relation, in all the nakedness and solitude of metaphysical abstraction.
>
> Edmund Burke[276]

With Binmore, everything is brought back to questions. *How* are decisions in a game made? Using *what* information, obtained *where*? *Why* would a rational player act in such and such a way? These are the questions that underly Binmore's contribution to the foundational debate. For example, discussing some claims he identifies with the "traditional

[274] Binmore (1998), p.115.
[275] As is the case with Aumann and Binmore.
[276] Burke (1909), p. 156.

response" to a critique of "perfect rationality", he charges them of sweeping "under the carpet the question which really matters here: namely, *how did the perfect machine get to be perfect?*"[277] And elsewhere: "Given that a guessing-algorithm-cum-stopping-rule is required, *where* does it come from? In particular, *what is the origin* of the basic guessing rule?"[278] No assumption or principle is left unexamined; they must all be justified. "From my earliest years", he says, "I have always ached to know the *why* of everything. It is like an itch that must be scratched."[279] This illustrates another way in which 'philosophy' makes its way into his game-theoretic papers; one that deals not with potential applications of the theory but with the critical examination of its underpinnings.

As Binmore's main concern in the literature on foundations will be to try and understand rationality, it is no surprise that his proposed use of the word 'rational' illustrates his philosophical inclinations. Firstly, always the iconoclast, he declares:

> A glance at any dictionary will confirm that economists, firmly entrenched in the static viewpoint [...], have hijacked this word and used it to mean something for which the word *consistent* would be more appropriate. Such an inversion is certainly useful as a rhetorical device. Who can argue with someone's advocating the rational course of action? But, insofar as scientific enquiry is concerned, such ploys can only be a source of unnecessary confusion.[280]

Modestly, he then proposes that "a *rational decision process* will be understood in this paper to refer to the *entire* resoning activity that intervenes between the receipt of a decision stimulus and the ultimate decision, *including* the manner in which the decision-maker forms the beliefs on which the decision is based."[281] For Binmore, in diametrical opposition to Aumann, "it is *not* enough that a model yield *results* that are intuitively, empirically, or mathematically satisfying. Intellectual honesty requires a critical appraisal of the *structure* of the model".[282]

Binmore illustrates these positions: "For example, the core implements Walrasian equilibria and there have been those who have attached much significance to this fact. But the

[277]Binmore (1988), p.25, my it.
[278]Binmore (1988), p.25, my it.
[279]Binmore (1999b), p.69.
[280]Binmore (1987b), p.181.
[281]Binmore (1987b), p.181. I did not need to put italics in the passage, they were already in the original!
[282]Binmore (1984), p.52.

mechanism by which the core is achieved bears little resemblance to empirical studies of the manner in which markets actually work."[283]

> Just as an argument is not refuted by showing that it leads to an unwelcome conclusion, so it is not confirmed by showing that it leads somewhere attractive. I do not doubt that economists are right to devote much of their time to studying Walrasian equilibria, but I think they deceive themselves when they offer a core analysis of a market game as a reason for doing so. The market-clearing price in large markets is demonstrably *not* achieved as a consequence of rival coalitions registering objections to this or that proposed allocation. Nor would it be in a world inhabited by farsighted superbeings.[284]

On this very subject, Aumann has written that "conceptually, the core expresses the idea of unbridled competition; non-emptiness of the core expresses the idea that such competition can lead to stability, that there is an outcome consistent with it."[285] But, constant in his methodological conceptions, he further writes that "this image of the core as expressing competition emerges from the applications; it is not in any sense obvious from the definition."[286] Binmore makes an *a priori* interpretation of the concept of the core based on the elements by which it is defined. Since this interpretation does not square with 'real markets', the core as applied to general equilibrium models is not convincing. For Binmore, "axiom systems always seem to make good sense when contemplated in the abstract. To uncover their hidden secrets, one has to worry them like a fretful terrier."[287] Aumann, suspicious of commenting on the 'meaning' of the core from its definition, waits for a sufficient amount of applications of the concept to emerge, and then gleans from what he considers to be the substance of these an *a posteriori* interpretation.

On another occasion, after discussing his views on players who 'make mistakes', Binmore adds that "the ideal of "perfect rationality" excludes such possibilities, *but the fundamental tenet of this paper is that this ideal is unattainable.*"[288] We have seen earlier that Aumann as well considers that 'perfect rationality' is an unrealistic ideal, yet that does not keep him from using it in his theories. His reasons for doing so, based on 'usefulness', were

[283] Binmore (1984), p.52.
[284] Binmore (1998), p.40.
[285] Aumann (1985), p.53.
[286] Aumann (1985), p.53.
[287] Binmore (1998), p.173.
[288] Binmore (1988), pp.21-22, my it.

also examined. Binmore's reaction is the opposite; if 'perfect rationality' is unacceptable as a postulate guiding agents' choices, then he does not wish to carry on the analysis until he has found an alternative method which is more satisfactory.

On these points, Binmore agrees closely with H. A. Simon. Referring to modern modeling methods among economists, the latter says:

> The model itself is manipulated with great mathematical formality, and if it is tested with quantitative data, high standards of sophistication are imposed on the statistical methods employed. *What is omitted is any serious testing of the validity of the assumptions of the model itself* [...]. This might be all right if the quantitative, econometric tests were generally sharp and decisive. Almost always, they are not.[289]

A similar point is taken up by Binmore, but in a slightly different context. In this case he is taking on both game theorists' reluctance to discuss the 'meaning' of basic assumptions and all of cooperative game theory at once.

> In the literature on cooperative games, such questions are seldom asked. Axioms are stated in formal terms with only a sentence or two of motivation, since the serious business is thought to be the proving of theorems that follows. The attention of trained mathematicians is then diverted to the details of the proofs, while lay folk read only popularized versions of the work because they are intimidated by the algebra. But perhaps the discussion of the Kalai-Smorodinsky bargaining solution will suffice to show that one can sometimes get by with almost no algebra at all, but that it never makes sense to leave the axioms unquestioned.[290]

Whereas some game theorists, Aumann in particular, refuse to diminish the importance of cooperative game theory, Binmore has branded much of it as next to useless. He is also one of the main proponents of the so-called 'Nash program', the effort to look for noncooperative foundations to cooperative games. For example, of books on game theory which ignore cooperative games (such as Binmore (1992b), Aumann writes: "The people who write these books are missing some very important sides of game theory; they are representing only their own knowledge and their own interests. These books do not give a balanced summary of what game theory has accomplished."[291] Whereas for Binmore: "If

[289] Simon (1997), p.21, my it.
[290] Binmore (1998), p.92.
[291] Aumann (1998c), p.197.

noncooperative game theory had advanced to a stage which allowed us to write down the "solution" of any contest, then we would also have solved all the problems of cooperative game theory. The appropriate cooperative solution concept for any game would simply be the "solution" of the contest obtained by prefixing the original game with an appropriately formalized negociation period."[292]

Binmore is aware of the complexity involved in rendering formally what he feels to be the correct method for modelling players and games. Yet, as we could have guessed after considering his remarks in the preceding discussion, that does not drive him to shy away from the task. He would rather have informal 'improvised' models he believes are 'on the right track' than neat and tidy formal models that he feels are basically unsound. "Ideally one would wish to offer a formal model of the equilibrating mechanism, but this will seldom be a practical possibility at this stage of the art."[293]

He believes game theory should be striving to develop some form of 'menu' of different equilibrating processes. For his conclusions "would certainly not be very promising if one were hoping for a method of analyzing games which was independent of the environment in which the game is played. But this aim seems to be both hopeless and misguided."[294] What is needed is:

> To take into account the possibility that the type of equilibrium to be used will be a function of the equilibrating mechanism operating in the environment under study. This is to be contrasted with attemps to provide blanket equilibrium definitions intended to be applicable in all abstract games independently of the equilibrating mechanism.[295]

The traditional 'unified' method:

> To my mind, [...] is like trying to discuss animal anatomy without a Linnaean classification scheme. Or, to take a more homely example, like trying to decide which of the roots of a quadratic equation is the "right" solution without reference to the context in which the quadratic equation has arisen.[296]

[292]Binmore (1998), p.43.
[293]Binmore (1984), pp.51-52.
[294]Binmore (1988), p.38.
[295]Binmore (1984), p.52.
[296]Binmore (1987b), p.183.

"Nothing would seem to guarantee that a good theory will be simple, and those who aspire to a unique theory of "perfect rationality", devoid of arbitrary or culturally determined events, are simply baying at the moon."[297]

Thus, in principle, one would choose this or that game 'framework', depending on the structure of the 'real life' situation one wishes to analyse. Although this sounds like other similar methodological propositions by other eminent game theorists,[298] what to Binmore constitutes the principal distinguishing mark of a situation is the *process* through which equilibrium is acheived. Essentially:

> A classification of game-playing environments [...] needs to take account of numerous issues, of which perhaps the most difficult concern the manner in which players process data (i.e., the modeling of their reasoning processes) and the manner in which data that serve as input to this process are generated and disseminated.[299]

On the other hand, for some game theorists, what was more important was categorizing what phenomena could and/or should be modelled as cooperative or non-cooperative games. Thus, they tended to look at the 'real world' situation's coalition structures, possibilities for communication or cooperation, the possibility to enforce commitments, etc. For example, Shubik (1982) argues that the "most important division in the classification of solutions is between cooperative and noncooperative solutions."[300]

> The extensive form stresses the fine structure of the game, the details of the moves and information. The strategic form suppresses much of the detail and highlights the strategic choices of the players, the details of the payoffs, and the possibilities for threats. The coalitional form suppresses strategic detail and highlights the joint gains that can be made by the formation of coalitions.[301]

Each of these "three major representations of a game stresses different features of the phenomenon to be studied."[302]

---

[297]Binmore (1992a), p.16.
[298]For example Aumann (1985), or Shubik (1982).
[299]Binmore (1987b), p.183.
[300]Shubik (1982), p.364.
[301]Shubik (1982), p.359.
[302]Shubik (1982), p.362.

2.5 PARADOXES

> [Philosophy] must set limits to what can be thought; and, in doing so, to what cannot be thought.
> It must set limits to what cannot be thought by working outwards through what can be thought.
> Ludwig Wittgenstein[303]

> Comment que'je fais pour mett' ma poubelle aux vidanges?
> Pierre Légaré, stand-up comic

This chapter, the preceding section in particular, has tried to show that Binmore enjoys questions. But certain types of questions seem to elicit in him a special fascination. These are interrogations that have a certain paradoxical flavor. He enjoys paradoxes of self-reference, of complexity defying the bounds set upon it. To express these, Binmore summons to his aid very specific types of highly abstract mathematics. These are the many types of mathematics dealing with impossibility and incompleteness. That is, he refers to forms of mathematics which, in the broader mathematical community's or his interpretation, support his ideas on the universes in which players and people play games.

2.5.1 THE GÖDEL CAN-OPENER FOR HERMETICALLY SEALED WORLDS

For Binmore, the 'world' within which games are played must, of necessity, be 'incompletable'. The idea that every possible state of the world is, in principle, 'knowable' by players, is very distasteful to him. In his papers on foundations, Binmore repeatedly defends this contention with the same argument, based on reasoning inspired by Gödel's famous 'incompleteness' results for formal systems representing arithmetic. Clearly, the idea behind these results is fundamental to him. First we look at a derivative of Gödel's result that Binmore is particularly fond of.

As Binmore thinks players should be seen as algorithmic, any decision taken by a player should be the result of a finite chain of calculations. More precisely, for some purposes, he wishes to model players as Turing machines.

[303] Wittgenstein (1961), p.30.

> Essentially, this is a computing device with no predetermined upper bound on the amount of storage it may use in a calculation. Turing envisaged a machine with a *finite* number of internal states equipped with a device for reading or writing letters from a fixed alphabet on a paper tape of indefinite length. Before the calculation, the tape is left blank except for the input data. After the calculation, it is to be hoped that the tape will contain an answer to the problem the machine is designed to resolve. During the calculation, the tape may be used to store interim results. What happens at any step in the calculation depends on the internal state of the machine and the symbol on the tape at the beginning of the step. These factors determine what the machine does with the tape *and* the next internal state to which the machine moves. The machine may overwrite symbols on the tape and/or scroll the tape one space to the left or right.[304]

"No apology would seem to be necessary for modelling a player as a Turing machine" for "it is orthodox for mathematicians to subscribe to the Church-Turing thesis, which asserts that any formal calculation possible for a human mathematician can be aped by a Turing machine."[305] Here Binmore defends a link between computing machines and the human mind; he is no longer speaking only of players in a game.

What Binmore offers by way of an argument to show that players' 'worlds' are 'incompletable' is what he calls: "a rather crude adaptation of the standard "halting problem" for Turing machines";[306] or "nothing more than the Liar's paradox and Newcomb's paradox dressed up in mathematical language."[307] "Indeed, many mathematicians would regard an argument as unnecessary, seeing the conclusions as an essential consequence of Gödel's observation that consistency is incompatible with completeness."[308] Here Binmore refers to Gödel's famous theorem on the incompleteness of formal systems. His judgement, that most mathematicians would consider his argument to be unnecessary, is interesting. The authors whose work he attacks with it, Aumann included,[309] are mostly mathematicians by training. None of these theorists would deny the *logic* of the argument, which is, as he says, derived from longstanding mathematical results. Yet what he is attempting to impose on

[304]Binmore (1987b), p.204.
[305]Binmore (1987b), p.205.
[306]Binmore (1984), p.54.
[307]Binmore (1987b), p.204.
[308]Binmore (1987b), p.209. This example shows up in Binmore (1984), Binmore (1987b), Binmore (1991a), Binmore (1993), Binmore and Brandenburger (1990) and Binmore and Shin (1992), in formulations that differ slightly (Binmore (1987b) being the most unique).
[309]Aumann (1999) even writes a paper on formal logic.

other game theorists is not the *comprehension* of these standard results, but an *acceptance* of his interpretation of them, and his application of this interpretation to game theory.

His argument runs as follows. *Suppose*, Binmore starts, that we model a 'perfectly rational' player with the help of a suitable Turing machine. Then this machine, by its very nature, should *always* provide answers, say by responding with a *YES* or a *NO*, to appropriately posed questions asked in a decision-theoretic framework. If that were not the case, it could not be 'perfectly rational'. Consider any question relevant to Turing machine *N*, and let the computer-coded form of this question be $\lfloor N \rfloor$. Just as legitimately, we can formulate a question about that question, such as: *Is it possible that Turing machine M will answer* NO *to* $\lfloor M \rfloor$*?* Call the computer-coded version of that question $\lceil M \rceil$. Contruct a machine T that outputs $\lceil z \rceil$ when it receives an input in the form $\lfloor z \rfloor$. This machine will only play the role of a 'transmission belt' of sorts. Now create a machine $R = ST$, that is, a machine R that is equivalent to first running machine T which feeds its output directly to machine S which uses it as input. Now we feed question $\lfloor R \rfloor$, a question about itself, to R. Machine T receives this question, and feeds question $\lceil R \rceil$ to machine *S*. The answer that S gives to $\lceil R \rceil$ will constitute the answer that R gives to $\lfloor R \rfloor$. Suppose R responds to $\lfloor R \rfloor$ with *NO*. Then S has stated that it is impossible that R responds *NO* to $\lfloor R \rfloor$. Suppose now that R respond to $\lfloor R \rfloor$ with *YES*. Then machine S has stated that it is possible that R responds to $\lfloor R \rfloor$ with *NO*, and has thus been proven wrong again. Thus the opening hypothesis, that Turing machines which give definite answers to all *relevant* questions of decision theory exist, is untenable. "In summary, the claim is that, if attention is restricted to players who *always* give an answer to problems that make proper sense, then sometimes such players will get the answer *wrong.*"[310]

> This echoing of Gödel is no accident. The halting problem for Turing machines, from which the preceding example is adapted, is closely related to part of Gödel's reasoning. Note, in particular, the self-reference involved in asking a machine how it will respond to a question about how it responds to questions.[311]

Indeed, the results of Gödel are an obvious source of inspiration for Binmore, and he rarely tires of citing them. Given their nature, and considering Binmore's aims and ideas,

[310]Binmore (1987b), p.209.
[311]Binmore (1991a), p.13.

this is not so incredible. The:

> Results of Gödel are obtained by a kind of metamathematical reasoning which goes more deeply into the structure of the formal system as a system of objects. [...] The objects of the formal system which we study are various formal symbols, formal expressions (i.e. finite sequences of formal symbols), and finite sequences of formal expressions. There are an enumerable infinity of formal symbols given at the outset. Hence, [...] the formal objects form an enumerable class. By specifying a particular enumeration of them, and letting our metamathematical statements refer to the indices in the enumeration instead of to the objects enumerated, metamathematics becomes a branch of number theory. Therewith, the possibility appears that the formal system should contain formulas which, when considered in the light of the enumeration, express propositions of its own metamathematics.[312]

Thus Gödel's results are based on fairly complex interactions between different 'levels' of theories; the formal logical system and its metamathematics, ordinary number theory. The crux of the result lies in considering statements in number theory that have, in a sense, two 'meanings' in, or two links to, the formal system. A first link between natural numbers and the formal system comes through the concepts of *recursive relations*, that is, relations between natural numbers that can be verified mechanically, through an algorithm. An important lemma for Gödel's result establishes a relationship between a true recursive relation between natural numbers and a provable assertion in the formal system. A second link is provided by the fact that the structure of the formal system is countable, thus expressible with natural numbers, once suitably coded.[313] This 'double link' is exploited by constructing a recursive relation between natural numbers that is only true when these numbers, when 'translated' back into formal propositions through the coding system, say something meaningful within that system. By cleverly constructing that relation and its corresponding formulas in the formal system, and by using the relation that always holds between a recursive relation and a provable statement in the formal system, Gödel obtains his result that either the formal model can prove and disprove one of its formulas (i.e., it is inconsistent) or that it is incapable of proving or disproving one of its formulas, that is known to be true in some of its models (e.g., in number theory). Thus it is incomplete.

[312] Kleene (1950), p.205.

[313] Through 'Gödel numbering'.

We can discern the attractiveness for Binmore of such reasoning. To him it accurately represents the principle that there will always be bounds to knowledge; and that every bound which, through experience, imitation, deduction or any other form of learning, is punctured and surmounted is unquestionably superceded by further bounds the whereabouts of which is not necessarily known. "Gödel provided a principle that allows extra axioms to be appended to any sufficiently complex formal deductive system in such a way that true statements that were unprovable in the original system become provable in the expanded system."[314] Gödel's results are also of a deep and philosophical nature in that they give a clear answer to a question of the form: 'What is is possible for us to *know*', albeit in a precise mathematical sense. And since Binmore sees players in a game[315] as computing machines, that is, as pieces of hardware that processe information through formal systems, it gives him a precise answer to the question: 'What is it possible for *players* to know?'

The last question is posed and answered in Binmore and Shin (1992). The remarkable aspect of this paper is that the authors append a game-theoretic interpretation directly to the structure of Gödel's results. "Informally, an algorithmic player will be understood to *know* that the relation R holds among $x_1, x_2, ..., x_k$ if and only if [the characteristic function of the relation R defined on k-tuples of natural numbers] C_R is a recursive function and $(x_1, x_2, ..., x_k) \in R$."[316] As in Gödel's proof, a lemma ensures that "recursiveness has a formal counterpart in the notion of proof"[317] in the formal system. Thus, only if the formal complement of a certain relation between natural numbers is *provable* will a player be said to *know* that relation. To embed game theory in that structure, all relevant information in a game is to be coded into natural numbers. "It is important that not only objects like games or strategies can be coded. Algorithms can themselves be coded as natural numbers, since they can be viewed simply as listings of computer instructions. This allows *players* to be admitted as part of the domain of enquiry."[318] The formal structure that they consider is "a first order logical system known as elementary Peano Arithmetic [...]. Apart from the usual symbols and axioms of first-order logic, the language of PA contains arithmetical

[314] Binmore (1993), pp.235-6.
[315] And humans and their brains, as we shall see later.
[316] Binmore and Shin (1992), p.147.
[317] Binmore and Shin (1992), p.149.
[318] Binmore and Shin (1992), p.147.

symbols"[319] and axioms describing the permitted manipulations of these symbols. "Most true formulas of *PA* are provable. [...] However, there are formulas which are true but which cannot be proved. The celebrated "Gödel sentence" is an example."[320]

> Since we are interpreting the knowledge of a player in the game as the provability of the appropriate formula in *PA*, to say that some player knows that another player knows some fact about the game is to say that one player can prove in *PA* that some other player can prove some formula in *PA*. In other words, we must have some way of expressing statements about proofs in *PA* *within* the system *PA*.[321]

This is done through the Gödel-numbering of statements about knowledge. "In effect, the Gödel number serves as the unambiguous "name" for each of these objects. In turn, each Gödel number has a corresponding numeral that can be expressed in the language of *PA*. In this way, certain expressions in *PA* can be interpreted as making claims about other expressions in *PA*."[322]

With such a setup, Binmore can *guarantee* that players will be incapable of achieving any form of *complete* knowledge, since by Gödel's result there will always be unprovable formulas (and thus 'unknowable' in Binmore and Shin's interpretation) in Peano arithmetic. And exactly *what* types of formulas are put on the chopping block is clear to Binmore.

> However, this common knowledge is restricted to features of the game that can be coded in the language of arithmetic. Thus, vague assertions such as: "It is common knowledge that all players are perfectly rational" are meaningless in the context of this chapter, unless the sense in which individuals are perfectly rational can be written (in coded form) as a formula in arithmetic. If the sense in which players are rational boils down to the prosaic claim that players do not play dominated strategies, then we can take this in our stride. However, if the claim is more exotic then we may not be able to accomodate it.[323]

As Binmore had done with Turing machines, not satisfied with marshalling to his service interpretations of Gödel's theorems, he enrolls Gödel *direcly* into his ranks, adapting not only the results of the theorems, but its building blocks. The same results follow, but they

[319] Binmore and Shin (1992), p.148.
[320] Binmore and Shin (1992), p.149.
[321] Binmore and Shin (1992), p.150.
[322] Binmore and Shin (1992), p.150.
[323] Binmore and Shin (1992), p.152.

have been reinterpreted from the very beginning. "No results that are not well known in mathematical logic are required. Any originality lies in the context in which they are applied."[324] This transfer between disciplines of the 'meaning' of results through interpretation is much more interesting than one that simply evoques the 'spirit' of the imported results while maintaining the importing discipline's theoretical structure unchanged.

But Gödel's service in Binmore's New Models Army is not yet completed. A dangerous and most likely fantastically controversial task still awaits him. Indeed, with some help from Gödel's theorem, Binmore sets out to deny "that the "I" presupposed by our language really exists."[325] In his opinion, most philosophical enquiries into the 'will' and the 'free will' are fundamentally misguided in that they attribute much more unity, coherence and capacity to understand its goals, wants and motives (in sum, itself) to the human individual as a center of thought. Thus, "far from being able to predict where our thoughts are leading, we are often very surprised at what we find inside our heads". To the problem of the 'individual':

> My answer is not in the least original. As Spinoza (1985) put it:
>
> Experience teaches us no less clearly than reason, that men believe themselves to be free, simply because they are conscious of their actions, and unconscious of the causes whereby these actions are determined.
>
> Accepting such a resolution of the problem of free will is tantamount to denying that there is such a problem.[326]

However:

> In denying a real existence to the "I", nobody [...] would wish to dispute the *usefulness* of this little word. In programming a computer, for example, it is highly convenient to be able to invent names for subprograms that need to be called upon repeatedly. If the program is to interact with other programs in computers located elsewhere, these too can be named. If the program needs to predict the result of its interaction with these other programs, the programmer may find it necessary to simulate their operation. His program will then incorporate subprograms that model the operation of the external programs. If these external programs also simulate the operation of the external programs with which *they* interact, then the programmer will therby be forced into simulating *their* simulations. In doing so, he will need to invent a name that his own

[324] Binomre and Shin (1992), p.143.
[325] Binmore (1993), p.227.
[326] Binmore (1993), p.227.

> program can use in referring *to itself* when it appears as an element in another program's simulation. This name may well be "I".[327]

With this last illustration, we see that for Binmore:

> The chief function of the "I" is to act as a mirror of others in our own minds and to reflect the manner in which we are similarly mirrored in the minds of others. Its nature is therefore inextricably bound up with the nature of the "I"s with which it interacts. *This is why game theory is so important if we are ever to understand what lies at the root of being human.*[328]

Binmore envisions the future of game theory as modelled on his view of human beings in society; as games played between computing machines. Binmore almost disowns this controversial position in a paragraph-long disclaimer.

> The path away from psychological correctness is therefore hard to follow, but I propose to follow it anyway. I know that I thereby risk discrediting the ideas on morality that I am writing this book to promote. Nor will it help to protest that one does not need to share my views on human psychology in order to agree with what I have to say about morality. My only hope is that the discussion which has led me to this extremity will have been sufficiently dry that my more critical readers will already have skipped to the next chapter.[329]

However, the quote before this one should make it clear that Binmore was far from considering these parts of the book as "dry".[330] Had he found these opinions of slight importance, he might simply have struck out the section, and save himself the garanteed ire his ideas would provoke among some philosophers. But as we have seen, in this section his life's work as a game theorist and his fundamental epistemological beliefs meet.

Binmore needs Gödel to placate opponents of a 'man as machine' perspective of the brain. These thinkers criticize the idea of the brain as a well-programmed computer and put forward arguments relating to the 'mystery' of the subconscious. As an example, one could point to the difficulty with which we sometimes retrace the steps of our reasoning.

---

[327] Binmore (1993), p.228.
[328] Binmore (1993), p.241, my it.
[329] Binmore (1993), p.228.
[330] This reader actually found that section of the book to be among some of its most exciting. Recall also, in this regard, his comment earlier about his books on philosophy being his 'magnum opus'. See p.66.

> As great mathematicians commonly report, [...] ideas just seem to pop up from nowhere. When I try to trace the origins of my own humble attempts to be mathematically creative, I come up against the same problem. The ideas come from somewhere beneath the level of consciousness. It is the same when one seeks to pinpoint the moment at which a decision is made. Hume (1978), for example, tells us that he often repeatedly decides to rise from his bed but fails to do so, only later to find himself out of bed and dressing with no clear idea of how this came about.
>
> However, if we truly are machines, are these not the sort of phenomena that Gödel-type arguments would lead us to expect? If we are machines, then it is logically impossible for us always to know everything that is going on in our heads.[331]

Binmore does not think his arguments regarding 'complete rationalities' "conclusive". In fact, he promptly applies a type of 'paradox of incompleteness' argument to question the possibility of a "conclusive argument" in this field. "It is tempting to suggest that conclusive arguments may not be available in this area. But no argument in support of this suggestion could then be conclusive."[332]

2.5.2 A BIG FUSS ABOUT THINGS THAT YOU CANNOT EVEN MEASURE

In 1984, before giving his argument regarding 'completable universes' using Turing machines that was reproduced above, Binmore had written in a footnote:

> Actually I prefer to attack the claims made by naive Bayesians to exclusivity for their approach via the axiom of choice and the well-known paradoxes of nonmeasurability. [...] But I am aware that such considerations invite dismissal because they are deemed absurdly esoteric.[333]

Examples of non-Lebesgue-measurable sets of real numbers date back to 1905, the first produced by the mathematician Vitali. A standard version is given in all textbooks on measure theory. An informal exposition of the argument is given here. To construct such a set, one first partitions an interval into an uncountable collection of disjoint sets by means of an equivalence relation defined in a way that makes it useful later in the construction.

[331] Binmore (1993), p.236.
[332] Binmore (1987b), p.182.
[333] Binmore (1984), p.54.

Secondly, one calls on the axiom of choice in order to select an element from each of these new sets. This creates an uncountable collection of elements, say $\mathcal{G}$, which, when interpreted in the light of the equivalence classes from which the axiom of choice permitted them to be extracted, are 'non-compatible' in a particular way. The third step is to 'attach' this uncountable collection to a countable one, creating a countable number of sets by adding a rational to each of the numbers in the collection, and repeating this for all rationals in a given[334] interval.[335] The equivalence relation defined in the first step also made use of rationals, and that relation now implies that the new countable collection of sets must be pairwise disjoint as well. One of the main reasons one would wish to work with measurable sets is that a measure defined on these sets is countably additive, a crucial property in many applications.[336] Thus, in this case, if we suppose the set $\mathcal{G}$ to be measurable, and using the fact that measures are translation invariant,[337] we obtain a contradiction by summing the measures of the sets in the countable collection created in the third step. In a way, the uncountably infinite 'incompatible' elements which were cleverly 'squeezed' into a countable number of sets, also 'incompatible', jump back out and produce a paradox.

There is no automatic interpretation of this result. Especially when, as is the case with Binmore, one takes a very singular result of pure mathematics and glimpses within it major ramifications for the epistemology of interactive decisions. Formally, the example speaks for itself, yet in the interpretations the author must speak for the result. Unless the author says something about formal matters which can be proven to be false, an interpretation cannot be verified,[338] it simply expresses how an author captures the 'meaning', in an extra-mathematical sense, of a result. Interpretations can be 'farfetched', unwieldy or odd, however, and thus be poorly judged or not taken seriously. Interpretations of a single

[334]Finite.

[335]I.e., $\mathcal{G} + q_i$, q_i a rational number, $i = 1, 2, ...$

[336]That is, for any collection $\{E_i\}_i$, $i = 1, 2, ...$, where $E_i \cap E_j = \emptyset$, for all i, j, $\mu(\cup_i E_i) = \sum_i \mu(E_i)$, where $\mu(\cdot)$ is a measure. Thus, in probability theory, where disjoint sets of elements are interpreted as independent events, this property is important if one wishes to allow the consideration of countably infinite sequences of such events.

[337]Thus $\mu(\mathcal{G} + q_i) = \mu(\mathcal{G})$.

[338]Except if they make some claims that could be verified empirically (although this can mean many things). However, in this case, we will see that Binmore's interpretation, like when he applied Gödel's theorem to the study of the brain, does not easily lend itself to such verification (at least not 'at this stage of the art' in psychology and the cognitive sciences).

theorem can sometimes vary greatly depending on the applications one considers for the mathematics, or even between people who apply them similarly, without it being in any way clear which one is 'correct'. An 'accepted interpretation' can be the result of custom, popularity or authority.

Binmore, not surprisingly, interprets non-measurability in a wide, philosophical sense. After some of the quotes from Binmore used earlier to illustrate his attitude toward the abundance of formal methods in game theory, it could seem surprising that in some cases he is quite enthusiastic about the application of very abstract mathematics to the understanding of rationality. But then again, they are very particular types of mathematics, and he interprets them in a very particular way.

> In fact, there is a whole body of mathematics that cries for attention in this context. The Axiom of Choice, in Zermelo-Fraenkel set theory, *may be loosely interpreted as asserting the existence of abstract computing devices (functions) whose internal structure is beyond the capacity of the analysing mathematician to duplicate*. He or she is therefore faced with an "open universe" problem. It is clearly no accident that, with the Axiom of Choice, sets of real numbers, which are not Lebesgue measurable, exist but, without it (but with some ancillary technical assumptions), all sets of real numbers can be taken to be measurable. [...] *The infinite, in this context, serves as an idealization for nonconstructible.* This problem will not go away and nothing is served by adopting a formalism within which it cannot be expressed.[339]

Again, Binmore's interpretation of a piece of mathematics that some mathematicians choose to ignore is centered on computability; his fundamental belief is that all knowledge must be capable of being constructed, ensuring that it will always be fundamentally bounded.

It is interesting that Aumann, in 1964, had quickly referred to nonmeasurable sets. This was in his famous paper *Markets with a Continuum of Traders*, which uses measures defined on agents in a trading economy modelled as points on the real line. He *assumes* that all sets in which he is interested are measurable, something which for him "is of technical significance only and constitutes no real economic restriction. Nonmeasurable sets are extremely "pathological"; it is unlikely that they would occur in the context of an economic model."[340] This last quote is not included in order to suggest that it would represent

[339]Binmore (1987b), p.211, my it.

[340]Aumann (1964), p.164. Similarly, referring to von Neumann and Morgenstern *stable sets*, Aumann has

Aumann's response to Binmore's arguments based on nonmeasurable sets in the epistemic debate. Although the contexts and times are very different, a few issues are worth noting. For example, could a proto-Binmore have taken exeption to the measurability assumption, saying that nonmeasurable sets of agents could represent 'stange' or 'unexpected' groups of agents whose exact influence on the economy could not be ascertained? Thus by excluding these we ignore an important element symbolizing the fundamental uncertainty of an economic system. The wording of Aumann's dismissal of nonmeasurable sets is also suggestive; he considers them nifty examples, limiting cases near the boundaries of the theory. This seems to be the view of a majority of mathematicians. The word 'pathological' is frequently used to dismiss the relevance of some examples. On the contrary, Binmore sees them as being of great importance for *any* theory, stumbling blocks one *must* take into account. From an intuitive point of view, their existence confirms his fundamental beliefs about human beings and their world. This is especially the case in game theory, where he believes that his interpretations of these matters deserve to be countered by theorists whose work it aims, and that "those who deny it need to be prepared to explain why they should not be classified along with those who count angels on the end of pins."[341]

> That is to say, in the world of theorem-proving, the "open universe" is a *necessary* fact of life with which one has learned to live. One is therefore perhaps entitled to be suspicious of theories of knowledge in which this fact of life is somehow evaded.[342]

These last points are reflected in Binmore's comment on Bruno De Finnetti's (1974) book on probability theory. Many books dealing with probability will *assume* the measurability of the sets they deal with, thus automatically discarding the problem of nonmeasurable sets. "De Finnetti (1974) must be given credit for being more scrupulous than some in that he acknowledges the related Banach-Tarski paradox. But his response, that only *finite* sets be admitted into the universe of discourse, misses the point."[343] Binmore appreciates that De Finnetti, like himself, recognizes in nonmeasurable sets an issue worth confronting head-

said: "The counter-examples [proving the nonexistence of stable sets in certain cases] are ingenious, difficult and deep; but there is no question that they are contrived. They do not appear to correspond to any economic, political or social reality". (Aumann (1985), p.58)

[341]Binmore (1987b), p.209.

[342]Binmore and Brandenburger (1990), p.120.

[343]Binmore (1987b), p.211.

on. But Binmore disapproves of the road the former will take once the problem has been noticed. De Finnetti's answer is based on *his own* intuition of the meaning of probability.

One is struck by the similarities of some passages in De Finnetti's book with some of Binmore's criticisms of 'mainstream' decision theory. De Finnetti, a prominent 'subjectivist' probabilist, is also somewhat of an outsider, an unorthodox, in probability theory. For example, he discusses the position "most commonly accepted at present":

> Its success owes much to the mathematical convenience of making the calculus of probability merely a translation of modern measure theory. [...] No-one has given a real justification of countable additivity (other than just taking it as a "natural extension" of finite additivity); indeed, many authors do also take into account cases in which it does not hold, but they consider them separately, not as absurd, but nonetheless "pathological", outside the "normal" theory.[344]

Again, in terms reminiscent of Binmore's, he asks himself:

> Why is it that, at times, some people prefer [...] to adopt a fixed frame of reference, within which one assumes complete knowledge of everything, all the details, no matter how complicated, no matter how delicate, and irrespective of whether they are relevant or not? This, despite the fact that the system is only used to draw particular conclusions, which could have been much more easily obtained by a direct evaluation. All this would appear to be a purely academic exercise; far removed from realism or common-sense.
>
> In seeking the reason for this, one should probably go back to the time when fear was the order of the day, and all manner of paradoxes and doubts resulted. The only hope of salvation was to take refuge within paradox-proof structures -and this was no doubt right, at the time.
>
> We must consider, however, whether it is reasonable, or sensible, to force those who are now strolling across a quiet park to take the same precautions as the pioneers who originally explored the area when it was wild and overgrown, and were ever fearful of poisonous snakes in the grass?[345]

He rejects simply *assuming* the measurability of all sets of interest to the analyst, since to him it sounds arbitrary.

> Only differences of a logical nature could possibly justify special treatment in a probabilistic context. In general, there is no reason to discriminate between

[344]De Finnetti (1974), p.119. Recall that countable additivity is crucial in the proof of the existence of nonmeasurable sets.

[345]De Finnetti (1974), p.236.

> sets, and, in particular, this applies to sets which have, with respect to the outcomes of a random quantity X, the form of intervals, or anything else, however "pathological". There is no justification for thinking that some events merit the attributing of probability to them, and others do not; or that over some particular partitions into events countable additivity holds, but not over others; and so on.[346]

Thus, although he has other reservations on this subject:

> The above could be taken in itself as a sufficient reason for *rejecting countable additivity as a methodologically absurd condition* (as a general, axiomatic kind of property) since it sets itself agains the abscence of any logical distinctions, which alone could justify discrimination between events.[347]

His distaste for the assumption of countable additivity is found *before* the analysis starts, in his belief that there is no *a priori* reason for eliminating nonmeasurable sets. Similarly to Binmore, irrespective of the results the theory obtains, if the starting point displeases him, De Finnetti goes no further. But at this point De Finnetti's and Binmore's ways part. The former shuns 'infinity' and sets up his theory on 'finitistic' lines, based on limits and approximations.

> Indeed, in practice, it will probably turn out to be advisable to limit oneself to *even simpler* ideas, sticking to the more elementary ambit (Jordan-Peano measure, Riemann integral) where the conclusions are unexceptionable, rather that passing to the more "modern" set-up (Borel or Lebesgue measure, Lebesgue integral), given that the usual extension is based on a convention which is inadmissible as a general axiom, and difficult to justify in a realistic way as a particular hypothesis for individual practical cases. It seems to me that it is difficult to justify not only its validity, but even that possible interpretations and applications to actual and practical problems are not illusory.[348]

Thus, disliking countable additivity as an axiom, and seeing no possible justification of its use in even the most restricted of ways, De Finnetti abandons the modern theories of measure and integration that need it. He does this at the cost of encountering problems with evaluating probabilities where there are discontinuous points in the distribution, as can happen when he reverts back to the Riemann integral. But, fundamentally, he *intuitively*

[346] De Finnetti (1974), p.237.
[347] De Finnetti (1974), pp.237-8.
[348] De Finnetti (1974), pp.125-6.

feels that considering probability distributions as resulting from limiting processes, with the fundamental but 'decided' imprecision it entails,[349] is a more *realistic* depiction of the way people need and calculate probabilities. After arguing that distribution functions should be considered as indeterminate at discontinuous point, he argues:

> On the other hand, this mathematical argument is closely bound up with the point that I consider to be most persuasive both from the point of view of fundamental issues and of applications: the need for some degree of realism when we assume the impossibility of measuring X with absolute certainty. [... To] consider $F(x)$ as completely determined, apart from discontinuity points, is equivalent to thinking that X can be measured with as small an error as is desired, but cannot be measured exactly with error $= 0$.[350]

Thus we have an example of the same formal results, in this case concerning nonmeasurability, leading to very different interpretations, these being closely linked with authors' intentions and convictions. Even though De Finnetti's proposal maintains a role for ever-present uncertainty, Binmore shuns it. One reason may be that De Finnetti's vision of ignorance has it ever-decreasing in scope, while for Binmore ignorance is an *irreducible property of a system that can know.* That system cannot *control* its ignorance. One can vanquish it by going outside and above that system, but only at the price of acquiring that new system's own boundary of the 'knowable'.

2.6 POSITIVE CONTRIBUTIONS

Much of what we have seen so far from Binmore has been negative in tone and intent. The deep-seated disapproval spawned by his evaluation of game theory's contemporary development leads him to devote much of his published work to examining this movement critically. But even though there is an important positive facet to his work in the foundational field, his negative aims always remain close to the surface. This section is concerned with his positive work, keeping this last caveat in mind. Also, the gap between the captivating literary criticism of traditional theory and what is actually put into a model cannot go by unnoticed.[351] Binmore translates his concerns with game theory's foundations almost

[349] For any $\epsilon > 0$, we can find an N ...

[350] De Finnetti (1974), p.242.

[351] Although there is no doubt about Binmore's skill as a theorist. The point is rather that in his critical

exclusively in terms of evolutionary games. The first subsection, with one exception, will be concerned with some of his writings in that field. In the second subsection we show how deeply anchored Binmore's penchant for evolutionary explanations of social phenomena is, by looking at some of the main ideas behind his work in political philosophy. The point is to show how, once again, Binmore's positions in game theory can be seen as expressions of more basic worldviews on human beings and society.

2.6.1 FROM MYTHOLOGY TO MUDDLERS...

> Jupin [Jupiter] les renvoya s'étant censurés tous,
> Du reste contents d'eux; mais parmis les plus fous,
> Notre espece excella; car tout ce que nous sommes,
> Lynx envers nos pareils, et taupes envers nous,
> Nous nous pardonnons tout, et rien aux autres hommes:
> On se voit d'un autre oeil, qu'on ne voit son prochain.
>
> La Fontaine[352]

Binmore's first attempt at incorporating thought processes in a game-theoretic model is cited in Binmore (1984) as an example of an eductive model with more 'realistic' foundations, although he developed it, with M. J. Herrero, in the context of bargaining models.[353] They present a new equilibrium concept, called the *security equilibrium*, weaker than perfect equilibrium, but "in which the constructive element is explicit."[354] Thus a security equilibrium, which they show to be unique in some cases, is a more general concept than perfect equilibrium with a definition exhibiting a precise process that describes how players arrive at their decisions. That is, "when security equilibria are unique, their definition embodies a scheme for *constructing* the equilibrium beliefs of the agents based on an iteration of Von Neumann's satisfying maximin criterion."[355] Thus, even if the idea is related to Kreps and Wilson's (1982) *sequential equilibrium*, as it takes "as fundamental the notion of an *assessment profile* rather than that of a strategy profile",[356] it dispenses with the "usual"

---

work the task he sets for game theory is beyond anybody's power to accomplish.

[352] La Fontaine (1984).

[353] Binmore and Herrero (1988b).

[354] Binmore (1984), p.55.

[355] Binmore and Herrero (1988b), p.33.

[356] Binmore (1984), p.55. An *assessment profile* for a player is a pairing of a strategy with a *belief*, that is

proceedure which assumes "that "common knowledge of rationality" endows agents with the ability to pluck sharp *a priori* predictions about the behavior of other agents "from the air"."[357]

"The intuition is that the players believe, or act as though they believe, that strategy profiles in a certain set I are "impossible". A security equilibrium is then simply a strategy profile which is not impossible."[358] I is defined inductively as an infinite union of sets $I = \cup_{n=0}^{\infty} I_n$. $I_0 = \emptyset$ and I_{n+1} is defined as the strategy profiles which are *rendered* 'impossible' by the fact that the profiles in I_n *are* 'impossible'. That is, some strategies not in I_n, now that it is 'common knowledge' that strategies in I_n will not be used, may yield to a player an expected payoff that is less than his *security level.* This is the payoff that the player can garantee for himself, the one which is obtained by supposing that his opponent can guess his strategy and will act to minimize his payoff. Thus a strategy profile is deemed 'impossible' if its play entails for one player an expected payoff that is less than a rational player's 'worst-case scenario'. After a round of elimination of strategy profiles, security levels are recalculated given the new 'possible' strategy sets, and the strategies which yield payoffs beneath this new level are likewise eliminated, and so on. Thus although this approach calls for common knowledge of a certain form,[359] the justification for its use and the limitations of its extent are quite clear and natural.

Another of Binmore's early positive contributions is in a paper co-authored with Larry Samuelson called *Evolutionary Stability in Repeated Games Played by Finite Automata.*[360] In this paper, the authors follow the work with finite automata of Rubinstein (1986) and

a set that contains, for each of his information sets, a probability ditribution on the nodes inside these.

[357]Binmore and Herrero (1988b), p.33.

[358]Binmore and Herrero (1988b), p.34.

[359]For as is the case with the iterated elimination of dominated strategies, to sustain one round of deletions, players have to know that their opponents are also 'deleters'. To sustain a second round, players have to know that their opponents know that they themselves are 'deleters' (that is, they have to know that their opponents act in accordance with the game that results from the deletions of strategies in the first round), and so on. For the case of security equilibria, the members of the set I are strategy *profiles*, containing one strategy for each player. The first round of 'contributions' to I are permitted by each player's knowledge of what he himself contributes to the profiles in I, and first-level knowledge of the other's 'security rationality' permits him to know his opponent's contributions. For the second round, a player needs to know that his opponents are playing the game from which the above-mentioned profiles in I have been withdrawn. He thus needs to know that the other knows that he is 'security rational', and so on.

[360]Binmore and Samuelson (1992).

Abreu and Rubinstein (1988).

> In a two player repeated game, the players are assumed to use machines [...] to implement their strategies. A (Moore) machine consists of a finite set of internal states (one of which is specified to be the initial state), an output function and a transition function. Given that the machine of a player is at a certain state at the tth round of the game, the output function determines the tth one-shot game action of the player as a function of the state. The transition function determines the next state as a function of the current state and of the other player's move at period t. A pair of machines induces a sequence of state pairs and a sequence of stage game action pairs, starting with the initial states of the two machines and their associated outputs.[361]

Thus the true 'players' are the entities who choose the machines before the game; a pair of machines determines the entire sequence of play. The preferences of these metaplayers are lexicographic in payoffs and 'costs', here interpreted as the number of states of a machine. That is, a player will first look for a machine that maximizes the stream of payoffs received in the game, independently of its complexity. Only in case of a 'tie' in payoffs between machines will the player concern himself with their cost, and choose the machine with the least amount of states.[362] Abreu and Rubinstein point to some difficulties with the interpretation of their model:

> Notice that we ignore the complexity issues connected with *computing* optimal strategies and concentrate instead on the costs of *implementing* them. While a unified approach to both these questions would be very attractive, there are contexts in which the present formulation appears plausible. One such application is the organization of bureaucracies: sophisticated managers seek to devise simple rules of thumb which can be implemented mechanically by lower level employees operating in a strategic environment with peers in parallel hierarchies. The machine is viewed here as a set of managerial instructions in accordance with which subordinates operate. On a more individualistic and cerebral level one could think of the states in a machine as primitive representations of "states of mind" -vengefulness, conciliation, aggression, etc. Players seek to devise behavioral patterns which do not need to be constantly reassessed, and which economize on the number of states needed to operate effectively in a given strategic environment.[363]

[361] Abreu and Rubinstein (1988), pp.1259-60.

[362] Abreu and Rubinstein (1988) use the more general *weakly monotonic* preferences. In that case if two machines yield the same payoff, the player will prefer the machine whose complexity is the lowest. But this definition of preferences does not exclude, for cases other than equality of payoffs, various trade-offs between payoffs and costs. Binmore and Samuelson (1992) will restrict their attention to lexicographic preferences.

[363] Abreu and Rubinstein (1988), p.1262.

Thus to confront the apparent strangeness of rational and strategically-minded players choosing unadaptive machines, Abreu and Rubinstein interpret it as a voluntary and calculated streamlining of a 'normal' player's commitment to a strategic situation. The metaplayers of this framework are very close to players in a traditional setting. And the solution concepts that are applied are the same as those traditionally applied.

> It is worth emphasizing that the model is embedded firmly within the standard noncooperative game-theoretic paradigm. The machine game is simply a normal form game in which a player's strategy is the choice of a machine. That is, players choose simultaneously, and once and for all, machines to play the repeated game. The solution concept is just *Nash equilibrium.* The new element is in incorporating considerations of complexity explicitely.[364]

While the technical details of Binmore and Samuelson's paper follow those of Abreu and Rubinstein, they nonetheless make a very interesting departure from traditional game theory by putting evolution in the role of the metaplayer.

> The interpretation to be studied in this paper is that the metaplayers are a metaphor for an evolutionary process. That is to say, the automata represent rules-of-thumb that have evolved during past plays of the (infinitely repeated) game. If metaplayers are to be seen as a metaphor for an evolutionary process, then it is natural to replace the notion of a Nash equilibrium by an appropriate version of the idea of an evolutionary stable strategy.[365]

The paper's main theorem is that once evolutionary stability has been suitably reinterpreted to accomodate finite automata as players, only 'utilitarian' outcomes are to be expected as evolutionary equilibria. These are outcomes in which the surviving strategies achieve the maximal sum of payoffs in the stage game. But the innovative interpretation seems to be the main import of the paper. Drawing links with Binmore's methodological concerns, the authors propose to model "the thinking processes of players explicitly."[366] "How are the thinking processes of a player to be modeled? The avenue of investigation that we regard as most promising abandons the theory that people think deeply about their behavior when interacting with others in game-like situations."[367] Instead:

[364] Abreu and Rubinstein (1988), p.1265.
[365] Binmore and Samuelson (1992), p.279. Evolutionary stable strategies will be presented shortly.
[366] Binmore and Samuelson (1992), p.284.
[367] Binmore and Samuelson (1992), p.284.

> Players are identified with the finite automaton that represents the strategy they are using. However, a more generally applicable paradigm sees the finite automaton as something like a virus that controls the strategy of a player who gets infected by it. The virus then spreads at a rate determined by how well the strategy performs relative to other strategies currently in use.[368]
>
> In evolutionary game theory, a game G is seen as being played repeatedly. Each time it is played, Nature chooses its players from a population whose composition changes over time. The players do not think about how to play G. They are endowed with strategies by a process of mutation and selection that tends to eliminate strategies that are relatively less successful.[369]

The authors place their paper in the foundational field by linking it to the problem of equilibrium selection.

> When more than one equilibrium exists, the problem of selecting one from among them is not easy. But progress on this front is necessary if game theory is to break out of the beachhead it has established in the social sciences.
>
> In seeking insight into the equilibrium selection problem, it is sensible to first look at simple examples. In such simple examples, the "right" equilibrium often seems glaringly obvious, and it may be easy to give lists of plausible *ad hoc* reasons why the right equilibrium should be selected. However, such principles are notoriously unreliable when applied in general. The purpose of the enterprise is not to pluck a selection criterion from the air that happens to be intuitively satisfying in particular cases. It is to find selection criteria that are defensible *from first principles*.[370]

It is not exactly clear who they are targeting with these remarks. It could be that they have in mind certain types of 'resolutions' of the types of game-theoretic 'paradoxes' that are raised by counter-intuitive conclusions.[371] Binmore (1993) ardently decries some of these, in particular attempts by some philosophers to secure cooperation in a one-shot prisoners' dilemma. But such ideas usually come from outside game theory, and get very little sympathy from economists. It could also be the case that they have the refinements literature

[368] Binmore and Samuelson (1992), pp.284-5.
[369] Binmore and Samuelson (1992), p.286.
[370] Binmore and Samuelson (1992), p.283.
[371] This interpretation is lent credence by a footnote to this passage which reads: "Our aims are the same as the even more easily misunderstood work of Aumann and Sorin (1989) or Anderlini (1990) who are concerned with equilibrium selection in games of pure coordination in which one equilibrium Pareto-dominates the others. Nothing could be easier or less relevant than to solve the problem they set for themselves by inventing "collective rationality" principles like: reject any equilibrium Pareto-dominated by another." (Binmore and Samuelson (1992), p.283)

in mind. Yet in what sense are the different refinements to Nash equilibrium 'indefensible from first principles' other than the fact that to Binmore and Samuelson they 'happen to be intuitively unsatisfactory'? Different refinements obtain with different postulates regarding the rationality of the players, and evolutionary stability results from stripping players of *all* rationality and transfering it to the evolutionary process. All these approaches start from 'first principles' of some kind, but are some of them more *ad hoc* than others?

Note that by the definition of an evolutionary process, the 'best' machines will always survive, and thus will obtain the highest payoffs. They will therefore be 'rational' machines. But this type of rationality is entirely dependent on the game being played. The 'thinking process' of a particular machine cannot be transferred into a new game, it is only a particular set of instructions on how to play *that* game in which it is currently involved. For example, if you transfer a machine from one prisoner's dilemma into another, where in the latter only the labels of the strategies differ, and if that machine has no set of instructions to permit it to recognize and adapt to this trivial change in the game, then it will not be able to operate in the new game. What is transferable from one game to another is the optimizing rationality embedded in the selection process. This, just like the types of rationality postulates used in non-evolutionary game theory, is defined in a manner general enough to be immediately applicable in any game. The players in this case do not even have preferences, these belong to the evolutionary process. 'Nature' prefers machines with higher payoffs, and in case of a tie for payoffs, the less complex machines.

Binmore and Samuelson admit that they "follow the standard practice of evading a study of the dynamics of the evolutionary process by appealing to the idea of an evolutionary stable strategy."[372] That is, although an evolutionary stable strategy is not defined in a properly dynamic way, it is easily interpreted as representing a strategy capable of withstanding a sequence of independent 'invasions' by mutant strategies. "Mutations are *rare*, so that after each mutant invasion, the system has time to attain a new equilibrium before the next mutant invasion."[373] Formally, a strategy a is an *evolutionary stable strategy* if:
(I) $U(a,a) > U(b,a)$, or

[372] Binmore and Samuelson (1992), p.287.
[373] Binmore and Samuelson (1992), p.286.

(II) $U(a,a) = U(b,a)$ and $U(a,b) > U(b,b)$

for all possible mutants b, where $U(x,y)$ is the payoff to the carrier of strategy x when he meets a carrier of strategy y. Thus a strategy a is evolutionary stable if it is always a best reply to itself (condition (I)), or if, when mutant strategies which are also best replies to a exist, this latter is a better reply to the mutants than these are to themselves (condition (II)). Non-existence of evolutionary stable strategies can be intuitively troubling when for strategies a and b, the situation is given by:

$U(a,a) = U(b,a)$ and $U(a,b) = U(b,b)$

> Existence problems of this type seem to us to be an artificial contruct arising from a definition of an ESS that is not entirely appropriate to the situation. The standard definition of an ESS demands that any sufficiently small invading group of mutants be eventually eradicated. One certainly would wish to use a definition that precludes mutant invasions in which the original small mutant bridgehead expands at the expense of the original normal population. But what of mutant invasions after which the original bridgehead neither expands nor contracts? The original normal population and the mutant invaders will then survive together in a state of peaceful co-existence.[374]

Binmore and Samuelson give a version of an evolutionary stable strategy that allows for these eventualities, as well as being suitably modified to account for costs attached to complexity. This implies that after many successive mutations, there will in general subsist more than one strategy that coexist in more or less stable "symbiotic relationships", that is, the population will be polymorphous. The authors extend the definition of an evolutionary stable strategy to cover polymorphous populations and find that their paper's main theorem, that the machines left in equilibrium will be utilitarian, still holds. Again, however, the stability of a polymorphous population is evaluated by testing it against only one mutant strategy at a time. This point bothers the authors. Already at the beginning of the paper the constraints imposed by the concept of an evolutionary stable strategy are found to be chafing. "Is such a model appropriate in our context? We think that socioeconomic evolution would be better modeled by supposing that mutations are sufficiently frequent that the systems do *not* have time to adjust to the last mutation before the next appears."[375] The problem is that broadening the definition of evolutionary stable strategies to allow the

[374] Binmore and Samuelson (1992), pp.289-90.
[375] Binmore and Samuelson (1992), p.287.

coexistence of different strategies brings polymorphous populations, and that the dynamic evolution of the proportional importance of strategies within these can be sensitive to the order of succession and importance of mutant invasions. Referring to an example with only two strategies, the authors write:

> The proportions in which [the component strategies in a polymorphous modified evolutionary stable strategy] are present will drift depending on the shocks the system receives as different mutants appear. Although the symbiotic relationship may persist for long periods of time, it will be stressed to the point of collapse if a sufficiently adverse sequence of mutations is encountered.
>
> Perhaps more satisfactory situations exist at higher complexity levels. At the two state level, little more can be said without information about how mutations arise.[376]

At this point, says Binmore, "we have [...] put aside our program of investigating bounded rationality by studying the evolution of automata until more is known about evolutionary dynamics."[377]

A first effort in studying the dynamics of many simultaneous mutations is found in Gale, Binmore and Samuelson (1995), in a paper entitled *Learning to be Imperfect: The Ultimatum Game.* In this paper, the evolutionary selection process is interpreted as *learning* and mutations as possible *mistakes* in that process. The point of the paper is to demonstrate:

> That interactive learning processes readily lead to outcomes in the Ultimatum Game that are Nash equilibria but not subgame perfect. We argue that game theorists were therefore wrong to put all their eggs in the subgame-perfect basket when predicting laboratory behavior in the Ultimatum game.[378]

The Ultimatum game is a famous sequential game where a 'pie' is to be split between two players, with a proposer making an offer, and a responder, faced with that offer, capable only of accepting or refusing it. Although any division of the pie is a Nash equilibrium, the subgame perfect equilibrium is for the proposer to offer nothing[379] and for the responder to accept. For the responder, to refuse is a weakly dominated strategy; since to say no to any

[376] Binmore and Samuelson (1992), p.295.
[377] Binmore (1999b), p.133.
[378] Gale, Binmore and Samuelson (1995), p.58.
[379] Or next to nothing.

positive amount of money is strongly dominated and he is indifferent between refusing or accepting an offer of nothing.

> Player I [proposer] need only believe that player II [responder] will not play a weakly dominated strategy to arrive at the subgame-perfect offer. But the deletion of weakly dominated strategies is an *eductive* principle, whereas we believe that the principles to which one must appeal when predicting actual behavior, in the laboratory or elsewhere, are almost always *evolutive* in character. That is to say, the outcomes we observe are not the product of careful reasoning but of trial-and-error learning.[380]

The authors study a simplified version of the Ultimatum Game, since they will use a computer to calculate numerically the results of the dynamic process. The 'pie' is of size 40, with proposers restricted to offering positive integers from 1 to 40. The responders' strategy is to choose a threshold below which all offers are rejected, and above which they are accepted. Thus for the responders as well, strategies are numbers that range over the positive integers between 1 and 40. The unique subgame-perfect Nash equilibrium is for proposers to offer 1 and for responders to accept any offer of 1 or more. In an evolutionary setting, we picture two very large populations of responders and proposers, each composed of 40 different types of individuals. Denote the fraction of proposers (responders) who make an offer of (plan to accept any offer above) i (j) at time t by $x_i(t)$ ($y_j(t)$). The fitness of a proposer making an offer of i at time t, denoted by $\pi_i(t)$, is given by the expected payoff to making that proposal given the current distribution of responders.[381] The fitness of a responder accepting any offer above j at time t, $\pi_j(t)$, is likewise defined. The average fitness of the population of responders at time t is the weighed sum of the individual types' fitnesses, $\bar{\pi_p}(t) = x_1(t)\pi_1(t) + ... + x_{40}(t)\pi_{40}(t)$. $\bar{\pi_r}$, the average fitness of responders, is defined similarly. The dynamics of the population fitnesses is studied through the *replicator equation*. This says that the fitness of a particular type within a population increases (decreases) in time when its individual fitness is above (below) that of the population. For the fitness of a proposer of type i, the replicator equation would be $\dot{x_i} = x_i(\pi_i - \bar{\pi_p})$, and similarly for responders. Thus we have 80 first degree difference equations; all that is left

[380] Gale, Binmore and Samuelson (1995), p.58.

[381] The populations of players must be thought of as 'large' so that the proportions of different types in the populations can be regarded as the probability of drawing an agent of that type at random from the populations.

is to specify initial conditions, let the computer calculate the trajectories in time of the different proportions of types in the population, and wait for these to reach their limits. With every type in both populations being initially equiprobable, the authors find that the system converges to an equilibrium in which responders receive a little more than 20% of the pie.

A problem with this setup, that is both technical and interpretative, is that there is no mechanism to 'throw back' into the population strategies that could have a long term effect, such as proposer strategies of very low offers, but which, in the short term, are rendered practically 'extinct' by the selection process.

> However, it may appear that Nash equilibria which fail to be subgame-perfect equilibria are attractors only because the long-run operation of the replicator dynamics allows some strategies to approach extinction, and hence artificially excludes the evolutionary pressure against weakly dominated strategies that would otherwise eliminate them. We therefore turn our attention to models in which small fractions of all possible strategies are continually injected into the population -including those that test the "rationality" of responders who refuse positive offers. Only if the survival of Nash equilibria that are not subgame-perfect is robust in the presence of such noise can we realistically argue against the subgame-perfect prediction.[382]

They rectify this problem by adding 'noise' to the dynamic process. That is, every period, a fraction δ_p of proposers of every type fails to conform to the replicator equation given above, and choose with probability θ_i a strategy i, where these range over all types of proposers that are not his own. The new replicator equation is given by:
$\dot{x_i} = (1 - \delta_p)x_i(\pi_i - \bar{\pi}_p) + \delta_p(\theta_i - x_i)$. Thus a failing strategy, whose representation in the population is smaller than the proportion of the other types who are adopting it, will always be reinforced by an inflow of agents from other types (this is the last term of the equation). We can define a similar equation for responders.

This is the framework of the model, and it is readily seen that it is very simple. In contrast, the authors' interpretation of their model is rather complex. While in Aumann's texts, the main theorems are 'played up' and the interpretations 'toned down', with Binmore the opposite occurs. His interpretations bring life to questions that are not easily seen in

[382]Gale, Binmore and Samuelson (1995), p.61.

the model. They create a certain richness (or even a certain beauty) around a model that, on its own, is bare. This is how the authors present what they see in their equations:

> Noise in an interactive learning system may arise in many ways and cause perturbations of various types. We therefore think it important to be clear on the source of noise to be studied.[fn: We depart from that part of the refinement literature which follows Kohlberg and Mertens (1986) in demanding robustness in the face of all conceivable perturbations. There is no reason to suppose that a system will necessarily be adapted to types of noise that it has experienced only rarely if at all.] This in turn requires that we take a little more care than is usual in modeling the agents.
>
> *We envisage an agent as as stimulus-response mechanism* with two modes of operation: a playing mode and a learning mode. Its playing mode operates when it is called upon to choose a strategy in one of a large number of games that it repeatedly plays against different opponents. Its behavior in each game is triggered by a stimulus that is determined by the manner in which the game is framed. (By a "game frame" we mean more than the game itself. We include also *the context in which the game is encountered and the manner in which its rules are described.*) When it receives such a stimulus s it responds by playing a strategy $D(s)$. If the learning mode were absent, an agent could therefore be identified with a fixed decision rule D that maps a set of stimuli into a set of strategies. However, sometimes an agent will enter its learning mode between games to adjust its current decision rule. When learning, it takes a stimulus s and some information f about the relative success of strategies in the game labeled by s to modify the value of $D(s)$. The learning rule that it uses for this purpose is assumed to be fixed.[383]

Thus Binmore renews his emphasis on computing machines by talking as though the agents in this model were explicitly analysed. The (unmodelled) very large number of agents that compose the increases or decreases in the relative importance of certain strategies in the populations, which in biology can be referred to as births or deaths, are in this case seen as collections of machines, who sometimes ‘learn’, or otherwise ‘make mistakes’. The authors assume "that the only source of error lies in the possibility that an agent may mistakenly learn to play a strategy that is adapted to the wrong game. We do not explicitely model the situations that may be confused with the Ultimatum Game."[384] That is, θ_i reflects the proportion of the population that play the Ultimatum Game while mistakenly believing that they are really playing another game for which strategy i has evolved. Futher on in the

[383]Gale, Binmore and Samuelson (1995), pp.61-1, my it.
[384]Gale, Binmore and Samuelson (1995), p.62.

paper, the authors endogenise the δ_p's, the probability of 'making mistakes', by making them dependent on the 'costs' of making a mistake. Since near the subgame-perfect equilibrium mistakes are very costly for proposers and cost very little to responders, less of the formers' low-offer types, and more of the latters' high-demand types will be thrown back into the populations, and thus the subgame-perfect equilibrium will be harder to reach. Again, an interpretation which seems to outshine its underlying model is suggested:

> Such an assumption adds more complexity to the stimulus-response mechanism used to model an agent. The mechanism must now *incorporate a device that responds to changes in its environment by diverting computational capacity between monitoring and other tasks according to the estimated rewards from the different activities.* Like the learning rule, this device is assumed to be fixed.[385]

To the careful and tight-lipped interpretations of Aumann, these grand constructions Binmore heaps onto very simple foundations stand in stark contrast. The authors are aware of the 'leaps of interpretative faith' which they ask of their readers. For example, they caution in a footnote:

> Although the English language forces us into speaking of players' misreading the game or learning to play better, it should be emphasized that our agents do not monitor what is going on except insofar as this is modeled by the learning rule with which they are endowed.[386]

But their ultimate goal is not simply the eventual adoption by their collegues of their model, but rather the acceptance of their way of thinking about problems in game theory.

> Such a viewpoint admittedly takes a lot for granted about the long-run equilibria of evolutionary processes. But perhaps the possibility that those who think like us might conceivably be right is sufficiently attractive that other theorists may also be tempted to divert their attention away from their traditional preoccupations to the evolutionary alternatives.[387]

Once again, we see that for Binmore the *models* he constructs can themselves bear polemical intentions.

[385] Gale, Binmore and Samuelson (1995), p.65, my it.
[386] Gale, Binmore and Samuelson (1995), p.62.
[387] Binmore and Samuelson (1994b), pp.61-2.

> We certainly do not want to claim that the theoretical framework presented in this paper is adequate to "explain" the observed experimental results in the Ultimatum Game. Our aim has been rather to show that these results do not serve to refute the optimizing paradigm, as it is often claimed by proponents of *homo sociologicus*. At the same time, we hope to convince at least some game theorists that they will never make any progress with the equilibrium selection problem by inventing more and more elaborate versions of *homo ludens*.[388]

Homo ludens is for Binmore and Samuelson the result of game theorists' 'one-upping' the economists' *homo economicus*; the latter "taking for granted that economic agents are well-informed mathematical prodigies capable of costlessly performing calculations of enormous complexity at the drop of a hat."[389] *Homo ludens*, for his part, "not only takes it to be common knowledge that his fellows are prodigies like himself, but continues to hold beliefs whatever evidence to the contrary he may observe."[390] However, the sociologists' *homo sociologicus* is not much better, for "no clear consensus exists on [its] precise nature. [...] Economists therefore have little use for *homo sociologicus*, since something needs to be clearly defined before it can be slotted into a model."[391] Here, once again, Binmore wants to defend game theory from attacks emanating from without, while at the same time attempting to 'save game theorists from themselves' by preaching a wide-ranging reforms in their approach.

Additionally, the study of learning dynamics has the additional virtue of linking experimental game theory to its theoretical counterpart:

> By transferring the debate into one concerning individual learning behavior, a topic amenable to both theoretical and empirical (or experimental) analysis, this investigation holds out the promise of making important new progress on the equilibrium selection problem. This progress may well require fundamental changes in our views of game theory. In particular, it appears as if backward induction will lose its currently exalted position.[392]

The paper, while presenting a theoretical model, is clearly aimed at the debate around the *experimental* performance of the theoretical predictions of the Ultimatum Game. "In

388 Binmore and Samuelson (1994b), p.61.

389 Binmore and Samuelson (1994b), p.45.

390 Binmore and Samuelson (1994b), pp.45-6. This point is related to the criticism that Binmore will level at backward induction. On this, see Chapter III.

391 Binmore and Samuelson (1994b), p.46.

392 Binmore and Samuelson (1994a), p.866.

the case of the Ultimatum Game, the relevant experiments have been replicated too often for doubts about the data to persist. A theory predicting that real people will use the subgame-perfect equilibrium in the Ultimatum Game is therefore open to question."[393] For Binmore, "experimental results on the Ultimatum Game are disturbing because so much in economics depends on the rational expectations hypothesis built into the idea of subgame-perfect equibrium."[394]

In the debates within experimental economics concerning the performance of economic theory and game theory in particular, Binmore plays a leading yet ambivalent role. He remains very critical of traditional game theory in his experimental work, especially concerning the computational abilities and farsighted nature with which it endows players. Yet he refuses the conclusion that this rejects the relevance of the 'maximization paradigm'. He aggressively attacks the authors who would rather replace self-interest with altruism as the important force driving agents' actions. One of the important recurring ideas in these papers is that people, when placed in unfamiliar (and bizarre to the uninitiated) game situations within a laboratory environment, must be given some time to *adapt* to this new context. People are seen as doing just about anything at first, importing some random 'strategy' (mode of behavior) from 'everyday life' to the game.[395] They then gradually learn how to play more 'rationally' *within* the context of the game. The last paper cited above, Binmore, Swierzbinski and Proulx (2001), where Binmore's interests in evolutionary games and boundedly rational players make their way into the experimental design, is particularly original and thought-provoking.

Experimental results are important, and can interact directly with the theory. "When properly conceived and executed experiments fail to confirm a theory, eccentrics like me believe that there is a good case for suspecting that the theory is *wrong*."[396] And "theorists who continue to take the validity of subgame-perfection for granted get away with it only because those who read their papers don't care about data."[397] Aumann's stance towards experimental game theory is much more skeptical. Even if it means that empirical verifica-

[393] Gale, Binmore and Samuelson (1995), p.58.
[394] Binmore (1999b), p.127.
[395] Note the resemblance to the details of the model just presented.
[396] Binmore (1999b), p.128.
[397] Binmore (1999b), p.129.

tions of the theory would be much rarer, more difficult and 'messier' to work with, Aumann thinks that these should be found in the 'real world'. One of his favorite examples:

> Comes from cooperative game theory; it is the work of Roth and associates on expert labor markets. This is something that started theoretically with the Gale-Shapley algorithm, and then Roth found that this algorithm had actually been implemented by the American medical profession in assigning interns to hospitals. That happened *before* Gale and Shapley published their paper. It had happened as a result of 50 years of development. So there was an empirical development, something that happened out there in the real world, and it took 50 years for these things to converge; but in the end, it did converge to the game theoretic solution, in this case the core. Now this is amazing and beautiful, it is empirical game theory at its best. [...] That is something very different from taking a few students and putting them in a room and saying, OK, you have a few minutes to think about what to decide.[398]

For Aumann there are few *organized* and *coordinated* ways to search for empirical examples of game theory at work. In some sense, they 'fall into the lap' of the perceptive researcher. Aumann also frequently uses the word 'beautiful' when referring to these types of empirical validations of game theory, in a sense very similar to when he refers to the mathematical elegance of the theory. There is no doubt that he sees very little beauty in the idea of students 'playing around' in an experimental laboratory.[399]

The desire to model sequences of overlaping mutations culminates in the so-called *musical chairs* model, presented in a paper of 1995,[400] and whose socio-economic interpretation is spelled out in a different paper.[401] The complexity of the model's technical aspects will make the presentation of anything exept its basic properties of little use for this thesis. Once again, the basic aim of that paper is to clarify the problem of equilibrium selection. "Which equilibrium should be selected in a game with multiple equilibria? This paper pursues an evolutionary approach to equilibrium selection in which the equilibrating process or "libration" is explicitly modeled."[402] So here Binmore comes back, but in an *evolutive* context, to his goals stated in 1987 with regards to *eductive* games; that the process through which equilibrium is reached should be modeled.

[398] Aumann (1998c), pp.183-4.
[399] Note that he refers to large laboratories filled with very sophisticated equipment just as "rooms".
[400] Binmore, Samuelson and Vaughan (1995).
[401] Binmore and Samuelson (1994c).
[402] Binmore and Samuelson (1994c), p.1.

> Dynamic models are unlikely to yield much insight into which aspects of a game's environment are significant in equilibrium selection if the mathematical properties of the learning rules studied are simply plucked from the air. For this reason, we consider the question of *microfoundations* to be crucial: it is important to derive the model's equations of motion from explicitly stated assumptions about the manner in which individual learning is postulated to proceed.[403]

The major conceptual change from the noisy replicator dynamics of the last paper is that there are now two channels through which agents can change their strategies, through *imitation* and through *mutation*, and that these processes are formally linked to individual agents. "Our agents are *muddlers* rather than maximizers. For muddlers, the learning process is itself noisy, in that agents do not always choose best responses."[404] Recall that in the unperturbed replicator dynamic agents could be said to be 'learning' only because, after each period, there were proportionally more members of the population using relatively more successful strategies than in the previous one. The precise agent who was learning remained inaccessible to the modeler. When mutations were introduced, these struck a fixed proportion of all types of players, who then chose randomly any strategy of the game. Thus well-performing strategies were 'eaten away' by mutations at every period. In this model the populations are finite, and one can in principle 'follow' agents through time in the selection process. "We consider a learning process that couples an aspiration-based rule for abandoning existing strategies with an imitation process for choosing new ones."[405] After each period, with a certain probability, an agent is allowed to 'learn'. He then checks his average realized reward against his personal benchmark, or *aspiration level*. If the former is above the latter, nothing changes. If it is below, there is a certain probability that he randomly selects another member of the population, whose strategy he *imitates*. This aspect of the selection mechanism will only redistribute agents within the strategy types already in use in the population. With the reverse probability, he 'mutates', that is he randomly chooses any strategy of the game. Thus both mutation and imitation find their place in the selection process, although they differ only in the range of the strategies to which a player may switch.

[403] Binmore and Samuelson (1994c), p.1.
[404] Binmore and Samuelson (1994c), p.3.
[405] Binmore and Samuelson (1994c), p.6.

There is an important technical advantage to such a specification. One would wish to study the dynamics of this model when mutation rates are low, i.e., examine the limit of the process when the mutation rates tends to zero. But if only mutations provide for changes in strategies, this means that one would need to wait for a large number of simultaneous mutations, which are very low probability events, to provoke long-run changes in equilibria. With the 'musical chairs' model one mutation, coupled with a sequence of imitations, which are much more likely, is sufficient to push the system from one equilibrium to another in the long run.

The model's results regarding equilibrium selection are rather weak. First, in games with only two strategies, the model selects equilibria different from its predecessors in that field.[406] And second, even within the model, the equilibrium selected can sometimes vary depending on the sequence in which the different limits are taken. Although Binmore's skill with words serves him well, he seems hard-pressed to motivate his claims that the evolutionary approach to equilibrium selection is really much better than the "more orthodox approach to the equilibrium selection problem [...that] invent[s] refinements of the Nash equilibrium concept."[407]

> Evolutionary game theory offers the promise of progress on the problem of equilibrium selection. At the same time, it is capable of reproducing the worst features of the equilibrium refinement literature, creating an ever-growing menagerie of conflicting and uninterpreted results. To achieve the former rather than the latter outcome, we think that evolutionary models need to be provided with micro-foundations which identify the links between the dynamics of the model and the underlying choice behavior.[408]

Binmore thinks that the construction of a 'bank' of refinements to Nash equilibrium, from which one could then make choices based on certain criteria in order to analyse a given situation, is arbitrary. But a similar procedure for dynamical systems and evolutionary selection processes is natural and "cannot be neglected". Different models starring unrealistically intelligent players are replaced by different models featuring unrealistically stupid ones, with the notable difference that the extent and exact operation of their stupidity is

[406]These are the models of Young (1993) and Kandori, Mailath and Rob (1993).
[407]Binmore and Samuelson (1994c), p.1.
[408]Binmore and Samuelson (1994c), pp.32-3.

modelled precisely.

For Binmore, it is important to study precisely *which* microfoundations lead to which types of equilibria, contructing a form of catalogue that could potentially be of service for the broad classification of 'real-world' problems. "I do not [...] believe that abstract evolutionary arguments can ever be *decisive* in an equilibrium selection debate. The best they can do is to *suggest* what data to look for in the historical record."[409] This reaches back to Binmore's (1987b) arguments for game theorists to develop and study many types of equilibrating mechanisms, whose 'usefulness' as tools to be applied would depend on the environment in which the game is played. "The point is that the same formal game might receive different analyses depending on the environment from which it has been abstracted: i.e., that *the analysis of a game may require more information than is classically built into the formal definition of a game.*"[410] Aumann would be displeased by a position that claims that the game itself does not contain enough information to allow it to be properly analysed. He would consider that the analyst did not do his job properly. Yet there are similarities between the foundational proposals of Binmore and Aumann. Both look for what, in the end, drives the results they observe; in order that there remain no doubts, when a given concept is used, regarding what need be implied by it. Although, as Binmore is not one to accept a floor without trying to find out what is under it, the paper opens the door to another level of questions: "However, we hope ultimately to dispense with the need to make arbitrary choices in the construction of the model by treating the learning process itself as being determined by evolutionary processes."[411] Although the authors give exploratory answers to this question, they will not be reviewed here. For this paper, the important point is that the question was asked. "Quite apart from the value of such assertions as "there exists in us a categorical imperative" one can still ask: what does such an assertion say of the man who asserts it?"[412]

[409]Binmore (1993), p.126.
[410]Binmore (1987b), p.183, my it.
[411]Binmore and Samuelson (1994c), p.4.
[412]Nietzsche (1886), p.92.

2.6.2 ... AND BACK AGAIN

Evolution and equilibrium selection are cornerstones for many of Binmore's constructions. His approach to political philosophy is centered on these concepts. Questions surrounding *social contracts* and social reform are identified with an equilibrium selection problem in what he calls the "Game of Life".

> I identify the set of all the commonly understood coordinating conventions operated by a society with its *social contract*. It serves only to coordinate behavior on an equilibrium in the Game of Life. The survival of the social contract does not therefore depend on its being backed up by some external enforcement mechanism. In a well-ordered society, each citizen honors the social contract because it is in his own self-interest to do so, provided that enough of his fellow citizens do the same.[413]

Indeed, Binmore sees "the advance of human mental horizons in terms of the evolution of progressively more elaborate *equilibrium selection devices*."[414] Rawls' (1972) concept of the *original position*, a hypothetical round of bargaining behind a *veil of ignorance*, where the bargainers' future roles in society are hidden from them, is reinterpreted as an equilibrium selection mechanism already developed in humans through the action of social evolution. "The original position is to be modeled as a natural device that evolved to help us coordinate on equilibria in some of the games we play."[415]

> In my model, the rules of the artificial game M [the "morality game" spawned by the "game of life" G; the latter whose "rules are determined by the laws of physics and biology; by geographical and demographic facts, by technological and physiological constraints; and by whatever else sets unbreakable bounds on our freedom of action." (p.4.)] require that Adam and Eve pretend that each round of G is preceeded by an episode during which each player can call for the veil of ignorance to be lowered. They then bargain in the original position about which equilibrium is to be operated when the veil is lifted. The rules of M then require that the actions specified by this equilibrium be implemented in future play -unless Adam or Eve demand that the veil of ignorance be lowered again for the agreement to be renegotiated. An equilibrium of the natural game G is said

[413] Binmore (1998), p.5.

[414] Binmore (1998), p184.

[415] Binmore (1998), p.10. Even leadership is viewed as an equilibrium selection device. "Whether leaders know what they are doing better than their followers or not, they can be very useful to a society as a coordinating device for solving the equilibrium selection problem in games for which the traditional methods are too slow or uncertain." (Binmore (1998), p.335)

> to be *fair* if its play would never give a player reason to appeal to the device of the original position under the rules of the morality game *M*.[416]

Therefore:

> My own interest in the original position is entirely pragmatic. It derives from the belief that rough-and-ready versions of the device are *already* built into the set of coordinating devices that humans use to get along with each other. In advocating the use of the device of the original position as an instrument for reform, I therefore make no grand metaphysical claims about its merits. All that is suggested is that we try to improve the effectiveness with which citizens in a society cooperate by adapting a coordination device that is familiar from use on a daily basis in our everyday life to larger-scale coodination problems for which social evolution has so far failed to generate anything very satisfactory.[417]

His aims in his two volumes on the social contract are political and not simply philosophical, although they do not contain any specific policy recommendations. The books are "an attempt to provide some logical underpinnings for the species of bourgeois liberalism that I am calling whiggery. Such logical underpinnings are to be found in the theory of games."[418]

> Is utilitarianism fair? Does justice require the use of the proportional bargaining solution? I care a great deal about such questions. But it would be premature to ask them at this stage because I think that justice and fairness are concepts without an *a priori* meaning. [...] I believe that such ethical concepts are human artifacts rather than Platonic forms whose true definition can be found by adopting the posture of Rodin's thinker and waiting for inspiration.[419]

Evolutionary pressure, apart from permeating the style and the intuition of the books, is needed to support a version of a theory of utility comparisons accross individuals.

> I believe [...] that the forces of biological and social evolution have [...] equipped us with the capacity to make interpersonal comparisons of well-being that a theory of the Good requires. But I am not prepared to treat this conclusion as axiomatic [... instead] I ask how and why such a phenomenon might have evolved.[420]

[416] Binmore (1998), p.11.
[417] Binmore (1993), p.334.
[418] Binmore (1993), pp.5-6.
[419] Binmore (1998), pp.86-7.
[420] Binmore (1998), pp.146-7.

This is done through *empathy*; the capacity to understand how another human being in a certain situation may 'feel' about it, based on reflections on how we ourselves would respond to it.

> I see the emphatic preferences held by individuals in a particular society as an artifact of their upbringing. As children mature, they are assimilated to the culture in which they grow up largely as a consequence of their natural disposition to imitate those around them. One of the social phenomena they will observe is the use of the device of the original position in achieving fair compromises. They are, of course, no more likely to recognize the device of the original position for what it is than we are when we use it in deciding such matters as who should wash how many dishes. Instead, they simply copy the behavior patterns of those they see using the device. An internal algorithm then distils this behavior into a preference-belief model against which they then test alternative patterns of behavior. The preferences in this model will be emphatic preferences -the inputs required when the device of the original position is employed.[421]

But in these books, it is not dynamical systems which are seen as fundamental representations of the games people play, but rather the *folk theorem* of repeated games. That is, the role of evolution is intuitive, hypothetical; it is not present in the analysis. Binmore takes for granted the existence of empathy-based interpersonal comparisons of utility, and argues in favor of them with evolutionary arguments. Long-run evolutionary explanations provide him with the 'parameters' of his model, which is an indefinitely repeated bargaining game played by rational players. The 'indefinite' length of play is considered to be a smaller time scale than the evolutionary time scale. Binmore sees the folk theorem's abundance of possible equilibria favorably:

> We feel that there must be many workable compromise arrangements between the two extremes in which Adam or Eve is entrusted with the entire power of the state. This view gains support if the game of life is modeled as a repeated game [...] What is important is that a whole spectrum of equilibria becomes available as a possible source of social contracts.[422]

"In brief, I think that the reason rules for moral or prudent conduct survive in our society is that they provide suitable maxims for sustaining equilibria in repeated games."[423] Binmore

[421]Binmore (1998), p.22.
[422]Binmore (1993), p.120.
[423]Binmore (1993), p.128.

sees social reform as a bargaining process that moves society from one equilibrium to another in the cooperative payoff region of an indefinitely repeated game; that is, from one 'social contract' to another.

> Insect species have to wait for chance to shift a society from one equilibrium to a more cooperative alternative. Societies operating the less efficient equilibrium will then eventually disappear if they are competing for the same ressources. Human societies are more resilient, since our capacity to imitate makes it possible for one society to borrow the cultural innovations of another. However, my guess is that Nature has made us even more flexible. I think it likely that we are genetically programmed with algorithms that help to immunize our societies against competition from innovative rivals. Such algorithms actively seek out Pareto-improving equilibria as these become available through changes in the environment in which we live.
>
> [...] My guess is that the fairness norms that seem universal in human societies have evolved primarily for this purpose. Their principal role is to single out one of the many equilibria typically available as Pareto-improvements on the *status quo* without the necessity for damaging and potentially destabilizing internal conflict.[424]

This short *compte-rendu* of Binmore's two thick volumes dealing with moral philosophy does not do them justice. The goal, in a manner similar to what was done for the mathematics of incompleteness, was to give an idea of the importance that equilibrium selection and evolution have in his general conceptions about the *world in general*, and not simply about the foundations of game theory. Readers may also have noticed other parallels with Binmore's work on foundations, for example the frequent use of the word 'algorithm'. As another example, consider Binmore discussing why social contracts would generally try to prevent 'cheating' with minimal rather than maximal punishment:

> The reason for discounting the possibility that real social contracts will specify maximal punishments is that, unlike the ideally rational players of traditional game theory, real people cannot guarantee that they won't occasionally stray from the equilibrium path as a result of careless inattention or foolish miscalculation. No matter how good our intentions, we must therefore sometimes expect to be punished ourselves.[425]

There is a unity of thought and purpose in *all* the numerous and diverse works of Binmore that is truly impressive. No conviction stands isolated, and their forms change to adapt

[424] Binmore (1998), p.209.
[425] Binmore (1998), p.307.

to different situations. Although at the beginning of this chapter we saw that Binmore presented himself as the lonesome Hercules fighting the nine-headed hydra, we could reverse this image and see him as the monster, whose heart and guts are in the land of game theory, but whose heads are everywhere else, chewing away at any Hercules he can set his teeth around.

CHAPTER III

THE PATHS MEET

> Things repeat themselves. One thing contains the clue to another.
> 'Abd-al-Ḥamîd[426]

> Consequently, there are two conceptions of mathematics, two mentalities, in evidence. After all that has been said up to this point, I do not see any reason for changing mine.
> Jacques Hadamard[427]

This chapter examines the issue on which Binmore and Aumann clash publicly; backward induction. The authors bring the full weight of their experience in foundations to this debate. This chapter gives a final illustration of the authors' aims, ideas and intuitions, and of the basic antagonism between these. We have here a prime example of a '*dialogue de sourds*'.

The paper by Aumann that sets off his and Binmore's public confrontations is entitled *Backward Induction and Common Knowledge of Rationality* and contains two results. The first and most important, his Theorem A, proves that if common knowledge of rationality (CKR) obtains in an extensive-form game of perfect information, then the backward inductive outcome occurs.[428] His second result, Theorem B, states that for every game of perfect information, there exists a knowledge system for that game that contains common knowledge; that is, CKR is possible in any game of perfect information. Note, once again, how he proceeds in his foundational venture. Backward induction is a widely used concept

[426] Cited in Khaldûn (1967), p.206.
[427] Hadamard (1905), p.1084.
[428] That is, any state of the world in which there is CKR is also one in which the strategies used by players are their backward inductive strategies.

in game theory. Aumann shows that if we are *willing* to *assume* CKR in the form with which he endows it, then backwards induction is *valid* as a technique helping us solve games of perfect information.

> Backward induction, the oldest idea in game theory, has maintained its centrality to this day. [... It] has a compelling logic of its own, especially in perfect information games (like chess). The last player, who must choose between leaves of the game tree, makes a choice that maximizes his payoff; taking this as given, the previous player makes a choice maximizing his payoff; and so on, until the beginning of the game is reached. Nothing seems simpler or more natural.
>
> Yet it is precisely this logic that has come under increasing recent scrutiny. [...] Indeed, *it is not obvious just what assumptions on the rationality of the players would justify it.* [...] In extensive games, there are serious difficulties in formulating the relevant concepts (knowledge, rationality, etc.)[429]

Aumann has no qualms about justifying backward induction through CKR, yet personally, he does not believe that common knowledge of rationality is 'possible' in real-world situations.[430]

> CKR is not "justified"; it does not happen. But that does not mean that CKR is unimportant. It is still very important to know what CKR says and what it implies, and to understand the connection between it and backward induction. [...] It is important to know what happens in the ideal state, although ideal states don't exist.[431]

According to some models which precede Aumann (1995), the assumption of CKR is contradictory in extensive form games. Basu (1990), for example, proves that his definition of rationality in an extensive form game leads to contradiction. For Aumann:

> The idea that CKR is self-contradictory was due to an inadequate model. If you build your model carefully and correctly, then CKR is not self-contradictory. Admittedly, I had to think for three years before coming up with the right model. It is a very very confusing business, it is very subtle, it takes a lot of thought.[432]

[429] Aumann (1995), pp.6-7, my it.

[430] Remember, in this connection, Aumann (1992), which argued that extremely small departures from the assumption of common knowledge of rationality could result in important changes in equilibria.

[431] Aumann (1998c), pp.207-8.

[432] Aumann (1998c), p.208.

Aumann admits to having spent a great amount of time and energy carefully crafting his result. He set out to show that, contrary to current 'fashion',[433] *certain* types of rationality and knowledge could not only be used coherently in extensive-form games but also yield backward induction, the tool that underlies most solutions to such games.

> The structure is not very elaborate. It has three building blocks: rationality, knowledge, and commonality of knowledge. You first examine each one separately, define it carefully; then you put the three together, and you get backward induction. The key [...] is to keep the ideas separate, not to confuse them. That is what mathematics is about, that's where mathematics helps.[434]

Aumann's model is again one of *belief systems*. What differs from Aumann (1987a) is that rationality is not defined as *Bayesian* rationality.[435] Rather, a player is rational at a given node if there does not exist a strategy which he *knows* would give him a higher payoff from that point on. Although in Aumann (1996) he provides a proof of his Theorem A with Bayesian rationality, he chooses to use the more unusual definition of Aumann (1995):

> For several reasons: One is that it yields a stronger result; that is, the result with Aumann (1995)-rationality implies that with Bayes rationality, but not the other way around. Another is that this strength is obtained at no extra cost. In fact, it is cheaper, because working with Bayes rationality requires the explicit introduction of probabilities, which just complicate the system without serving any essential purpose. Finally, justifying the standard Bayesian framework requires an axiom system such as that of Savage (1954). While we ourselves have no problem with this, there are others who do; so we felt it preferable to avoid implicitely assuming axioms that really have nothing to do with the matter at hand.[436]

Pragmatic model-building, a very easygoing view towards controversial axiom systems, dedication to mathematical rigour and generality; we have seen all these themes in Aumann earlier.

Some concepts seen in the first chapter, framed in terms of purely static strategic-form games, need to be modified to fit into the framework of extensive games. New concepts,

[433] There are many papers that, for different reasons, criticize backward induction (see Aumann (1995) for references to some of these). Aumann (1995) is one of the few who takes a stand in favor of it.

[434] Aumann (1998c), p.208.

[435] A player is Bayesian rational if, given his information as expressed by his information partition, he chooses the strategy that maximises his expected payoff.

[436] Aumann (1996), p.139.

such as the *inductive choice at node v*, an important part of the proof of Aumann's (1995) Theorem A, must be defined. This is the action that maximizes a player's payoff given that the inductive choices will be made at all the nodes succeding v. "In particular, when there are no vertices after v -when the action at v directly determines the outcome- then the inductive choice is simply the action maximizing i's payoff."[437] Theorem A will be proved through mathematical induction, and this inductive definition of inductive choices offers by itself the proof's first step.[438] What then needs to be proven is that, supposing that the inductive outcome prevails at all nodes w that follow node v, common knowledge of rationality implies the inductive choice at node v. The proof is based on the same intuition that guides the 'normal' use of backward induction; one starts with the end nodes of the game and works upwards towards the initial node.

Binmore, on the other hand, is among the first of the many authors who casts a stone at backward induction. He chooses a verbal, intuitive stone. "In 1987, I wrote a paper (Binmore (1987b)) that questioned the rationality of the backward induction principle in games of perfect information. Since that time, a small literature has grown up in which [...] numerous others have attempted with varying success to treat the issues formally."[439] He warns that his "remarks on interpretation just seem like waffle to formalists",[440] yet he insists that "all the *analytical issues* relating to backward induction lie entirely on the surface".[441] His "purpose is to question the *significance* of this and other results of the formalist genre."[442]

Binmore (1987b) is heavily critical of backward induction. Already, Binmore (1984) annouced his main concern: "Both the power of the equilibrium ideas under study in producing results and their weakness in being difficult to justify, lie in the fact that they take into account potential behavior at information sets which will not be reached in equilibrium."[443] Illustrating his ideas with a hundred-legged Centipede Game, reproduced in Figure 3.1,[444]

[437]Aumann (1995), p.8.
[438]That is, the assertion of Theorem A is true by definition when the game tree has only one node.
[439]Binmore (1996), p.1. See that article for references, many of which are also cited in Aumann (1995).
[440]Binmore (1996), p.17.
[441]BInmore (1996), p.2.
[442]Binmore (1996), p.1.
[443]Binmore (1984), p.51.
[444]The payoffs are different in Binmore (1987b, p.195) but the game is equivalent.

Binmore says that:

> It is not disputed that the result of the play of this game by rational players will be that I plays "down" at the first node. What is to be contested is a statement that is often made about subgame-perfect equilibrium strategies [...]. It is said that these represent rational plans of action for the players *under all contingencies*. The inference is that, *if* player II were to find herself called upon for a decision at node 50, then she *would* play "down". Admittedly, it makes good sense if player II explains her arrival at node 50 with some version of the Selten trembling-hand hypothesis. But is this a good explanation? Should she really attribute her arrival at node 50 to 25 uncorrelated random errors on the part of player I or should she look for some less unlikely explanation?[445]

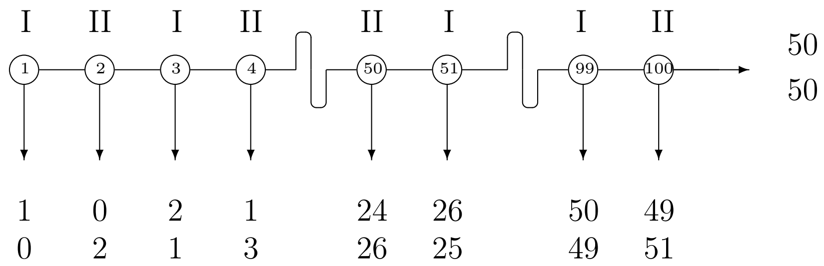

Figure 3.1

A hundred-legged Centipede Game in extensive form. At every node, a player's inductive choice is to play 'down', although, were the players to 'stay in' for a longer time, they would be much better off.

What follows is a reinterpretation of Binmore's criticism. In a game in extensive form, each player chooses a strategy, and any combination of strategies, one for each player, determines a path through the game tree to a terminal node. If all strategies are equilibrium strategies, then the path thus described is an equilibrium path. Such a path determines which nodes can or cannot be reached if particular equilibrium strategies are adhered to.[446] Since a strategy specifies an action for each information set where a player might be called upon to play, behavior off the equilibrium path is specified by equilibrium strategies. In a Nash equilibrium, the calculation of expected payoffs which *determine* which strategies are optimal assumes that opponents will stick to their strategies. Thus, although actions at

[445] Binmore 1987b), p. 196.

[446] With mixed strategies or games of imperfect information, nodes that could be reached have positive probability, while nodes that cannot be attained by any sequence of realisations of random choices have a probability of zero.

nodes off the equilibrium path are given by equilibrium strategies, one may ask why these strategies should be adhered to were a deviation to be observed. "Selten's trembling hand hypothesis"[447] would have players assign such unexpected behavior to random and uncorrelated 'mistakes' by their opponents, whose status as 'rational' would remain unchanged. Thus, here is game theorists' *homo ludens*, who "not only takes it to be common knowledge that his fellows are prodigies like himself, but continues to hold beliefs whatever evidence to the contrary he may observe."[448] A player should stick to his equilibrium strategy in such a case, as the effects of such errors were already discounted in the caluculation of the payoffs associated with the use of that strategy *before* the start of play. For Binmore:

> It is important to emphasize that such a question [concerning what players should *think* about out-of-equilibrium behavior] cannot be answered *in the abstract*. The answer depens on the environment in which the game is played. To insist on one answer rather than the other is therefore to make a judgment about the nature of this environment. My own judgment about the environment in which the game of chess is normally played would lead me to attribute player I's repeated decision to use a dominated action to some *systematic* error on his part, perhaps relating to some misunderstanding about the rules or payoffs.[449]

Binmore's views on backward induction are intimately linked to his general views on rationality. According to him, his opinions imply that:

> It is necessary to have a backround theory of a rational player that is sufficiently well-specified to allow a computer model to be constructed and sufficiently flexible to admit variation of its parameters. The first consideration necessitates an algorithmic (or constructive) view of rationality rather than an ideal (or axiomatic) view. The second necessitates serious consideration of theories of imperfect rationality even by those who are unwilling to be persuaded that perfect rationality is an unattainable ideal.[450]

We see that when Binmore and Aumann think of backward induction, they have in mind very different objects. While Binmore sees backward induction as the most powerful example[451] of an unduly formalized eductive theory that eschews the study of 'realistic'

[447] The idea which underlies Selten's refinement of the Nash equilibrium termed 'trembling-hand perfect'.
[448] Binmore and Samuelson (1994b), pp.45-6.
[449] Binmore (1987b), p.196.
[450] Binmore (1987a), p.198.
[451] Both in the sense of its conceptual frailties and in that of its importance in game theory; it is thus a very important 'target'.

decision processes, Aumann sees the need to clarify and render more precise one of traditional game theory's most important and successful tools. These issues and others will be further discussed through a more detailed examination of the actual debates.

Citing Binmore (1987b), Binmore begins his note in a 1996 issue of *Games and Economic Behavior*[452] with the following sentence: "It now seems to be generally accepted that rational players would not use their backward induction strategies *if* there were to be a deviation from the backward induction path."[453] He then goes on to present Aumann's position as though it were a claim to the contrary: "But Aumann (1995) has recently offered a formal defense of the proposition that prior common knowledge of rationality implies that play will *nevertheless* necessarily follow the backward induction path."[454] Aumann understands this as confusion surrounding the *nature* of his claims. To him the two propositions are not opposites: "Apparently, Binmore sees some kind of inconsistency or contradiction between this [the first sentence of Binmore's note] and the results of Aumann (1995)."[455] Aumann states that "we agree wholeheartedly with Binmore's first sentence".[456] Indeed, he sees no problem with players playing off the backward induction path at *any* point in the game, if that were to correspond to what they believe to be their best interest. Rationality *per se* is simply the maximisation of gains, and it can accomodate many types of non-backward inductive histories of play, but "the assumption of CKR gives us -and the players- additional information"[457] with which to predict the outcome of play. He continues, "our results concern a situation with common knowledge of rationality, *not* just rationality. Binmore continuously confounds these concepts, using them almost interchangeably."[458]

The trouble between them here stems from the '*nevertheless*' highlighted in the citation

[452]Binmore (1996)

[453]Binmore (1996), p.135, my it. That *if* is very important for understanding the rift between Binmore's and Aumann's positions in this debate.

[454]Binmore (1996), p.135, my it.

[455]Aumann (1996c), p.139.

[456]Aumann (1996c), p.139.

[457]Aumann (1996c), pp.141-2.

[458]Aumann (1996c) pp.139-41. The same point comes up in Aumann's response to Binmore and Samuelson: "The fact of the matter is that there is a big difference between rationality and CKR, a point that Binmore and Samuelson consistently slur over. CKR is a very strong assumption, much stronger that simple rationality, and in some situations may just be too much to ask for. It is this, and this only, that accounts for the strangeness of the BI outcome in some PI games." (Aumann (1996a), pp.130-1)

from Binmore above. Aumann's (1995) Theorem A states that if all players are rational, and if all players *know*[459] that the others are rational, and if they all know that the others know that they are rational and so on, then the backward induction path of play is the only one that *can* be observed. Any other strategy choice is *irrational* for any of the players involved, and since all players are rational *beyond doubt*, they will not use any such strategies. "Under common knowledge of rationality, vertices off the backward induction path *cannot* be reached; and when CKR does not obtain, the results [of his Theorem A] do not apply."[460] Deviations from the equilibrium path, with the rethinking Binmore perceives them as imposing, simply do not happen in Aumann's framework. Binmore wants to find a way for agents to be 'surprised', to have to rethink their strategy choices when facing an opponent who was *a priori* thought to be rational, but who revealed his irrationality by his choice of actions. In Aumann's formalism, 'surprises' of this kind are impossible; the agent's knowledge is always absolutely true.

It is interesting to note that Binmore and Samuelson's (1996) attack on Aumann proclaims in the first sentence that "Binmore (1987b) explains why backward induction in games cannot be justified by assuming only that there is common knowledge of rationality before the game begins."[461] That is, they directly contest the results of Aumann's paper, not the concepts and definitions used. Aumann (1996a) rejects this claim as an unsubstantiated assertion that cannot be countered by argument. Sniffing an imprecision, he proposes that "to challenge the result that CKR implies BI, one must either find a logical flaw in my demonstration, or *challenge my formulations of the underlying concepts - rationality and knowledge*.[462] Although the main ideas of Binmore's note in GEB in 1996 are essentially the same as in Binmore and Samuelson (1996), in the former Binmore seems to have adopted, in form, Aumann's second suggestion. Indeed, Binmore's (1996) purpose is, in principle, to "question the definition of rationality that [Aumann (1995)] employs."[463]

The first two sentences of Binmore's (1996) second paragraph land the two authors deeper in misunderstanding. Binmore says that "what keeps a rational player on the [back-

[459]They know "this for sure, without a shadow of a doubt" (Aumann (1996c), p.141).
[460]Aumann (1995), p.18.
[461]Binmore and Samuelson (1996), p.111.
[462]Aumann (1996a), p.130, my it.
[463]Binmore (1996), p.135.

wards induction] equilibrium path is his evaluation of what would happen if he were to deviate. But *if he were to deviate, he would behave irrationally*".[464] Again, for Aumann, this proposition is an odd one, because if a player behaves irrationally in the course of the game, how could he then have been rational, and *known* with certainty to be so by the other players, in the first place? But, in any case, Binmore pursues, suppose players 'known' to be rational deviate from the 'rationally prescribed' backwards induction path.[465] He then proposes that "other players would [...] be foolish if they were not to take [...] evidence of irrationality into account in planning their responses to the deviation".[466] Binmore then argues that Aumann's:

> Conclusions say nothing whatever about what players would do if vertices of the game tree off the backward induction path were to be reached. But if nothing can be said about what would happen *off* the backward induction path, then it seems obvious that nothing can be said about the rationality of remaining *on* the backward induction path. How else do we assess the cleverness of taking an action than by considering what would have happened if one of the alternative actions had been taken?[467]

To Binmore this point is the cornerstone of his argument, i.e., that Aumann's "formalism does not allow counterfactuals to be expressed."[468]

A *conditional* statement is a statement of the form 'if p then q'. Mathematics usually deals with one kind of conditional, 'p implies q' ($p \Rightarrow q$), formally written $q \vee \neg p$ (q or not p). That is, $p \Rightarrow q$ is true in all cases except when p is true and q is false. This is a much more narrow interpretation of conditional reasoning than that which is concerned with statements of the form: 'If weathermen were competent, they would forecast rain for tomorrow'. In such cases the relation between propositions p (the antecedent) and q (the consequent) is not one of mathematical implication. As Aumann says, "in everyday discourse, a conditional usually expresses a substantive relation between the antecedant p and the consequent q." (Aumann (1995), p.14.) Such conditionals are named *substantive* by Aumann, whereas Binmore, using a term more common among philosophers, calls them

[464]Binmore (1996), p.135, my it.
[465]I.e., the one deduced by *assuming* CKR.
[466]Binmore (1996), p.135.
[467]Binmore (1996), pp.135-6.
[468]Binmore (1996), p.136.

subjunctive. A *counterfactual* conditional is a substantive (subjunctive) conditional with a false antecedant. The example just given is thus a counterfactual, since it is commonly known that weathermen have no idea what they are talking about.

Before expanding on Binmore's view with regards to counterfactuals, let us note that Aumann seems to consider these passages as utterly obscure. He admits to not understanding the line from Binmore's note that was quoted (concerning counterfactuals) at the end of the next-to-last paragraph. Also, what he gathers from that long quotation in the same paragraph is that Binmore seems to be casting doubt on his comprehension of the definition of a strategy, one of game theory's most fundamental concepts.[469] Aumann's exasperation is palpable. He muses that Binmore is unable "to distinguish between the subjunctive and indicative moods."[470] He then cites a passage from Aumann (1995), in which he had written that:

> "Making a decision means choosing among alternatives. Thus one *must* consider hypothetical situations -what would happen if one did something different from what one actually does [...] In [...] games, you must consider what other people would do if you did something different from what you actually do".[471]

Having demonstrated that an intuitive presentation of strategic thinking was indeed present in his 1995 paper, Aumann states the matter more precisely:

> "A strategy of a player is defined as a function that assigns an action to *each* of his vertices, reached or not; the rationality of a player is defined in terms of his rationality at each of his vertices; and this, in turn, is defined in terms of what he knows about the other players' strategies."[472]

On this point, Aumann concludes, "Binmore is 180° off the mark."[473]

[469] From Aumann's rebukes it seems clear that he understood the remarks at least partly in that sense.

[470] Aumann (1996c), p.140. He writes in a footnote (previously cited in the Introduction): "Our intention could not have been made clearer. The indicative mood is heavily stressed: "are actually reached." "Are"; not "would", not "were to be". For the case that somebody might still misunderstand, the word *"actually"* is thrown in. And if, by some stretch of the imagination, *somebody* might *still* misunderstand, our very next sentence (not cited in Binmore) would surely clear thing up: "Under common knowledge of rationality (CKR), vertices off the backward induction path *cannot* be reached; and when CKR does not obtain, the results do not apply."" (Aumann (1996c), p.140)

[471] Aumann (1995), p.14.

[472] Aumann (1996c), p.140.

[473] Aumann (1996c), p.140.

Although it seems doubtful that Binmore was implying that Aumann did not comprehend how to construct strategies, he *was* opposing Aumann's *method* for doing so. In Aumann (1995), when a player is considering what to do[474] at a certain non-initial node x in the game tree, he takes that node to be the initial node of the subgame formed by cutting off from the original tree node x and all nodes that could possibly be reached from x. "Each choice must be rational "in its own right" -a player may not rely on what happened previously to get him "off the hook"".[475] With this method, in a game of perfect information, there are no 'beliefs' present when deciding what to do at some new subgame. The histories of play before that particular node are irrelevant. Formally, they are non-existent. More specifically, Aumann's Theorem A is purely forward-looking in structure; players at a given node v assume something about others' behavior at nodes after v (which was itself determined in the previous steps), but no reference at all is made to nodes before v. Thus although Aumann interprets his model as specifying all knowledge before the game actually starts,[476] formally a player at v trying to decide what action to choose for that node, need not know anything about his 'past' actions since he is in the process of determining which strategy is indeed optimal. Once this deduction is made, *supposing* that other players stick to their inductive strategy, a player has his optimal strategy. This strategy specifies actions at nodes not reached in equilibrium, and these actions were decided upon by a procedure which supposed that these nodes were 'reached', but not really during 'play'. Thus here we seem to have a counterfactual, and a delicate one at that. It really jostles our interpretation of timing in extensive games, of how and when players contruct their strategies. Aumann (1995), in his interpretation, seems to stick to safer waters. "Substantive conditionals are not part of our formal apparatus, but they are important in interpreting four key concepts [...]: strategy, conditional payoff, rationality at a vertex, and rationality."[477]

> A player's strategy specifies what he does at each of his vertices, *if* reached. His conditional payoff at a vertex v signifies what he gets *if* v is reached. For him to be rational at v means that he cannot knowingly increase his payoff if

[474]Before the start of the game, while choosing his strategy.

[475]Aumann (1995), p.12.

[476]"The formalism, which does not refer explicitely to this matter, is subject to different interpretations. We adopt the most straightforward, that all knowledge refers to the start of play: For a player to "know" something means that he knows it before any actions are taken." (Aumann (1995), p.13)

[477]Aumann (1995), p.15.

> v is reached. And for him to be rational means that for each of his vertices v, he cannot knowingly increase his payoff if v is reached. All these "if"s are substantive.[478]

Binmore complicates this explanation immensely by adding another 'if' to all the questions above concerning rationality, namely, '*if* the inductive hypothesis, that at all nodes after v the inductive choices are made, still holds'.

To Binmore, this counterfactual can be a potential source of contradiction. The logic behind it is found, if not outright faulty, at least worthy of serious study. Binmore (1987b), roughly, presents the following argument.[479] A backward inductive argument, once constructed, gives us a chain of statements, one for each node, of the form: "if my opponents are rational, I should do...". But if a player's node, say k, is rendered unreachable by an equilibrium action of an opponent at a node preceding k, then the statement attached to the kth node has a false antecedant, namely, the player's opponents are *not* rational. Thus, the equilibrium strategy "tells us *nothing* about what a rational player should do if node [k] were *actually* reached".[480] However, he goes further. He finds an inconsistency *within* the deduction of the backward inductive strategy itself. Suppose a player, at a node k, given rational choices by his opponents at nodes after k, chooses his inductive action at k. Suppose that this player now wishes to choose a strategy for some node j that precedes k, and that the inductive action at j[481] renders k unreachable. Then he has just falsified the hypothesis that he used at k, namely, that rational play reached k, and his choice at k is then based on a "spurious deduction".[482] Binmore says that "the immediate point [of "the convoluted logic underlying this argument"] is that there is nothing intrinsically irrational or illogical in [a player's] deviating from her subgame-perfect strategy at node [k]."[483]

To Aumann this type of argument, when carried out in a game with prior CKR, is downright fallacious. "The error in the argument is that it mixes the conclusion of the proof [fn: That vertices off the backward induction path cannot be reached under CKR.]

[478] Aumann (1995), p.15.
[479] See Binmore (1987b), p.197.
[480] Binmore (1987b), p.197.
[481] Or some action taken by another player at some node between j and k.
[482] Binmore (1987b), p.197.
[483] Binmore (1987b), p.197.

into the proof itself."[484] He characterises Binmore's argument thus:

> Starting with CKR, we prove that P1 goes down at the first vertex. Now, we say, lets try it again. Must P1 *really* go down at the first vertex? Lets suppose not -i.e., that she goes across. But then we have a contradiction to CKR [, so] we must abandon CKR. But then there is no longer any reason to go down at the first vertex.[485]

But Binmore's questions operate at a wholly different level. Interpreting them can be quite thorny. For example, there is a certain ambiguity regarding the temporal framework within which he embeds his arguments on (counterfactually) irrational behavior. Although Binmore's comments always show a concern for the *sequence* in which knowledge is gained and the *process* through which decisions are made, it is not always easy to know if and when he is referring to "unexpected" events in a dynamic or static setting.

In the first case, which seems to be the most common interpretation, players *revise* their strategy choices when confronted with an *actual* irrationality. That is, players *adapt* during play. In the second interpretation, players would have to consider counterfactual occurences *while* deducing a strategy for the whole game, i.e., before the start of the game. In the first, how can we speak of an individual player's strategy in the traditional game theoretic sense,[486] since all actions *observed during play* affect his choice of future actions in a way he neither did nor attempted to predict? CKR would be a common prior belief that could possibly be called into question at the first move. In the second context, agents would be faced with a bizarre task, if considered from a traditional point of view. Suppose a player is pondering the different strategies available to him before a game of perfect information with CKR begins. Looking forward to a certain node x, he somehow constructs a theory or belief structure to explain the types of "deviations" from rationality by the other players that could get him there. Speaking of deviations implies that there was an *equilibrium path* that was not respected. But at this point the player is supposed to be choosing the actions which will *construct* the equilibrium path for the game under consideration.

This is related to Binmore's distinction between eductive and evolutive processes. The

[484] Aumann (1996c), p.143.
[485] Aumann (1996c)p.143.
[486] I.e., as a complete plan.

first interpretation is close to evolutive processes, in which "strategic plans at unreachable information sets are [...] seen as "ghosts of failed competitors"."[487] That is, out-of-equilibrium strategies are 'tried' by boundedly rational players and survive in the game proportionally to their success. The second interpretation is closer to eductive processes, where "we shall regard strategic plans for unreachable information sets as "ghosts of failed hypotheses"."[488] In this setup the full complexity of Binmore's questions has to be tackled. No 'revised' equation can be final, they could only lead to further questions about possible infringements of the last revisions. How this can be done is another question. But as we have seen, Binmore's hopes are squarely with the evolutionary branch of game theory.

Again we see that Binmore's critique would render Aumann's approach to game theory's foundational questions devoid of meaning. And again, that may be the whole point. This helps in understanding why Binmore's attempts to make players evaluate counterfactuals of the type '*suppose players I knew to be rational were to behave irrationally*' appear perturbing and a little strange, if not incomprehensible, to Aumann. Indeed, to him, once common *knowledge* of rationality is *assumed*, it is senseless to have players "discover" in the course of play that their opponents may not be rational, which is what Binmore seems to be aiming for. That Binmore's aims were far-reaching is revealed in one sentence of his note: "A formal model that neglects what *would* happen if a rational player *were* to deviate from rational play must therefore be missing something important, no matter how elaborately it is analysed."[489] Here, again, is Binmore's opposition to the *type* of formalization of rationality and knowledge in game theory that Aumann helped to develop.

At the end of the debate, the only issue that seems to be settled is that the gap between both protagonists is unbridgeable. Aumann's epistemic approach, rich in results and 'too big' for Binmore to stop, continues on its way, although with the limits on the complexity of the behavior it can express being better understood. Binmore returns to evolutionary games and to philosophy, a consecrated maverick and trouble-maker, no longer to weaken the epistemic approach through direct criticism but through the construction of a possibly more attractive alternative. Interestingly, after this face-to-face debate with

[487]Binmore (1984), p.53.
[488]Binmore (1984), p.53.
[489]Binmore (1996), p.135.

Aumann, Binmore will not publish further purely methodological papers on foundations. The lines are drawn and fairly clear, and time will tell the importance that this clash of unusual vigor and contrast has had on the development of game theory.

CONCLUSION

This thesis attempted to provide a case study in the history of science. In Binmore and Aumann, we met two scholars who, even if working in a fairly small subfield of science, are 'in different worlds' when it comes to their ideas on the ends, the tools, the procedure and the 'meaning' of their activity. Knowledge, rationality, learning, equilibrium; even when they work with the same definitions of these terms, which, as we have seen, is not often, their interpretations of what these 'mean' are different. They also appraise game theory's long relationship with mathematics differently. Yet there is enough room in the discipline for both of them; Aumann's stature was never in doubt and the research program he started with *Agreeing to Disagree* is now thriving. Binmore recently established and led a large and successful research center[490] at the University College London. Also, without doubt, game theory is richer for having witnessed their clash.

This leads to two major types of questions that were not addressed in this thesis. The first, stemming directly from above, is: What have game theorists learned from this debate? How have they assimilated its many messages? What channels did these communications take? Did they read the papers, or did they attend conferences at which Binmore, Aumann, or both were present? In brief, this thesis concentrated solely on Binmore and Aumann, abstracting from the environment in which they work. But the reactions of a field as a whole are always crucial in scholarly disciplines.

Second, even the portraits of Binmore and Aumann are incomplete. Here, more biographical information would be needed to understand how they *lived* through the debate. What was its importance to them?[491] What do they think it meant for game theory? What was the extent of their correspondence at the time? What was the tone of that correspondence? In brief, as this thesis has mostly studied the models that both authors chose to construct or to criticize, an extension of it would need to look more deeply at and interract

[490] The ESRC Center for Economic Learning and Social Evolution.

[491] While this seems reasonably clear in the case of Binmore, it is much less so for Aumann.

with the authors themselves. The first part of what is left of this conclusion attempts to begin to fill that gap. It discusses a few issues that have to do with the 'personalities' of Binmore and Aumann. In the last part, more general issues are addressed.

We have painted the portrait of Aumann as of a pragmatic, 'down-to-earth' model-builder. He also seems very much the realist in 'life'. Here he discusses a committee of which he was a member, whose purpose was to try to devise rules that would minimize strategic voting in the election of fellows of the Econometric Society. He contrasts his frame of mind to that of Kenneth Arrow:

> I discussed the matter with Arrow, expecting to find a sympathetic ear from the founder of social choice theory. He did express interest, but little real enthusiasm for the issue. He himself does not vote strategically; somehow he feels that it is unbecoming that, as scholars and gentlefolk, the Fellows of the Society should engage in such practices. *I have assured him that I do; I play the game by its rules and I see nothing even remotely immoral or unethical about it.* But whereas intellectually he recognizes that some people do vote strategically, emotionally there is something in him that rejects this. He does not want to bother with that kind of thing and expressed surprise at my involvement. As an economist he recognizes the importance of incentives; but as a humanist, he cannot get terribly excited about them in a practical context.[492]

The impression is one of a very calm man, with a fatalistic, yet also somehow optimistic, outlook on life. What good are emotional roadblocks when they don't change the world, and when they make you 'miss out' on some interesting events? What good is a guilty conscience that is spawned by a gap between the necessity to participate in the world and the way we would wish the world to be?

> Incidentally, there is a man whom I love and respect very much and with whom I have worked closely and extensively, but who disagrees with me [...], and that is Mike Maschler. [...] He wants to know what the "truth" is, and, if it is difficult to calculate, he doesn't care, he has to find the truth. I don't have this notion of truth. *It is really odd to see him wanting to find the "right" bargaining set.* He has this idea that some of the bargaining sets are wrong and some of them are right.[493]

In the debate looked at in this thesis, Aumann's position asks: "Why are you stopping at the assumptions of the model? That's just the beginning, wait until it's all done, then we

[492] Aumann (1987b), p.138, my it.
[493] Aumann (1998c), p.202, my it.

shall see." "Why are you bugging *me*? If you don't like what I do, go do something else! There is plenty of room for everybody." "If you do not believe that this model is realistic, fine. But please look at it from another angle, for example, is it not pretty?"[494] For Aumann, there are always many ways to look at a situation.

> What will "really" happen? Which solution concept is "right"? None of them; they are indicators, not predictions. Different solution concepts are like different indicators of an economy; [...] different maps; [...] different batting statistics; [...] accounts of the same event by different people or different media; different projections of the same three-dimensional object [...]. They depict or illuminate the situation from different angles; each one stresses certain aspects at the expense of others.[495]

"All this may sound very slippery and unsatisfactory. There are no firm predictions, no falsifiability. If our theory apprears not to work, *we don't lose any sleep.* "Rationality is just one of the relevant factors", we say blandly; "here something else is at work.""[496]

Aumann's vision of game theory is very inclusive. Abstract theorists and empirical workers are not in opposition, in fact, they are complimentary.

> At some level the theory has to be several stages in advance of empirics. So we could be doing very complicated theory, let us say equilibrium refinement à la Mertens with cohomology and all that, and at the same time be doing empirical studies and empirical engineering like Milgrom was discussing [...]. In some sense it wouldn't be possible for [Milgrom] to work out these very practical rules without knowing something like Mertens' cohomology refinement; perhaps not necessarily that, but something of similar depth. You have to swim in it. [...] These things are all hitched together; they are part of a *milieu.*[497]

In a similar vein, Felix Klein has said:

> I hope that what I have here said concerning the use of mathematics in the applied sciences will not be interpreted as in any way prejudicial to the cultivation of abstract mathematics as a pure science. Apart from the fact that pure mathematics cannot be supplanted by anything else as a means for developing the

[494] "We know that, in bringing up our children, we must accept each one for what *he* is, for the good that is in *him*, and must not force him into somebody else's mold. The sciences are the children of our minds". (Aumann (1985), p.36)
[495] Aumann (1987c), p.464.
[496] Aumann (1985), p.37, my it.
[497] Aumann (1987b), p.191.

> purely logical powers of the mind, there must be considered here as elsewhere the necessity of the presence of a few individuals in each country developed in a far higher degreee than the rest, for the purpose of keeping up and gradually raising the *general* standard. Even a slight raising of the general level can be accomplished only when some few minds have progressed far ahead of the average.[498]

Aumann sometimes seems like a kid in a toy store when talking about game theory. He never ceases to marvel at what he sees; he craves beauty and *accepts* mystery. Of game theory's relationship to the 'real world', he says: "In fact, I find it somewhat surprising that our disciplines have any relation at all to real behavior. [...] There is apparently some kind of generalized invisible hand at work."[499] And commenting on the fact that *stable sets* often lead to insights about coalition structures, even though, as other cooperative solution concepts, they are defined in terms of payoff *imputations*,[500] he writes: "It is a matter of some mystery that this particular definition should lead so frequently and consistently to this kind of interpretation."[501] For Aumann, mathematics, and game theory by extension, can be seen as art forms.

> Mathematics at its best possesses great beauty and harmony. The great theorems of, for example, analytic number theory are reminiscent of Baroque architecture or Baroque music, both in their intricacy and in their underlying structure and drive. Other sides of mathematics are reminiscent of modern art in their simplicity, spareness and elegance.[502]

Mathematics for him is more than a tool or a 'language'; it is a way of life, a path to fulfillment and self-improvement. "Mathematics imposes a discipline of thought; it forces one to think clearly."[503] When asked "Would you say that Arrow has a mathematical bent of mind?", Aumann replies: "Absolutely. He is an extraordinarily *clear* thinker."[504]

Whereas Aumann, who is a religious man, seems peaceful and content, Binmore displays more angst and restlessness. Nearing the end of his career, his tone seems worried.

[498] Klein (1911), p.964.
[499] Aumann (1985), p.36.
[500] These make no reference to coalition structures.
[501] Aumann (1985), p.59.
[502] Aumann (1985), p.42.
[503] Aumann (1987b), p.135.
[504] Aumann (1987b), p.135, my it.

> My increasing ill health will probably force me to retire in the not-too-distant future. Insofar as I am remembered, I hope it will not be because I am counted among those who wasted their talents disputing intricate subtleties about goat's wool and moonshine in the water. Still less do I want to be counted among those who hindered, condemned, forbade, and scoffed at those who did something worthwhile. Most likely, I shall not be remembered by name at all.[505]

His sometimes aggressive manners seem to hide a fragile ego. "In spite of all evidence to the contrary, I still feel inadequate when faced by any new challenge. But it now begins to look as though I might make it to retirement without ever being found out."[506]

Binmore's reflections on his career have a tinge of sourness to them. He delivered some mean punches, but they came back to him as good as he gave them. It seems that his confrontations are more important to him than they are to Aumann. Aumann has gotten into controversies during his career, yet in the short introductions that precede the chapters of his *Collected Works*, they take up very little space; they are not in emphasis, no bitterness remains. Binmore's clashes with established wisdom have left their marks. He recalls when: "economists interrupted my later seminars on bargaining on more than one occasion to ask why I thought the subject might be of interest to them!"[507] The same happened with his experimental work: "I recall a distinguished member of a seminar audience expressing surprise that I didn't know that "economics is not an experimental science."[508]

> Unthinking hostility [...] has marred the pleasure that I otherwise would have found in my experimental work. In my years as a mathematician, I never knew what it was to have a paper rejected, but my experimental papers continue to provoke accusations of dishonesty or incompetence from referees, although all have stood the test of time once a journal could be found to publish them.[509]

He considers his attack on 'Bayesianism' to "have been a total failure".[510] "I gave up trying to press this point after being more or less laughed off the platform for suggesting to an

[505]Binmore (1999b), p.137.
[506]Binmore (1999b), p.120.
[507]Binmore (1999b), p.122.
[508]Binmore (1999b), p.127.
[509]Binmore (1999b), p.127.
[510]Binmore (1999b), p.130.

audience of financial economists that the reason that real people don't act like Bayesians in experimental laboratories was because there is something wrong with the theory which says that they should."[511] Of his algorithmic view of rationality, he says that "I have found few economists willing to believe that the work of Gödel or Turing might be relevant to their subject".[512] He sees his salvation through an evolutionary framework; he will be made immortal through his contribution to the formation of a group of thinkers possessing a theoretical edge on their competitors that will allow them to thrive and multiply. "I have great hopes that the memes created by the group of microeconomists and game theorists of which I am a part will survive as long as civilized societies persist."[513]

There is an ambivalence in Binmore's 'emotional' relationship with mathematics and, in fact, with his career as a scholar. Binmore was "born in humble circumstances during the London Blitz".[514] "Our home had few books, and my parents paid little heed to my schooling".[515] He sees his ending up in university as an exogenous phenomenon with which he had little to do:

> It was neither my need to ask questions nor my capacity to understand the answers that took me to a university. Without the educational conveyor belt created by the Labour Government that won a landslide victory after the Second World War, it would never have occurred to me that I had what it takes to be a successful scholar. I was quite taken aback when I came out at the top of the examination list after studying Mathematics for a year at London's Imperial College.[516]

Becoming an academic was a decision taken after an interview that could have led to "a lucrative job in business",[517] his preferred choice of career. That job hunt:

> Finally brought me into the grand office of a big boss who was reduced to a sweating wreck when our conversation was interrupted by a telephone call from an even bigger boss. The alternative was to read for a doctorate. My thesis topic was in classical analysis in the style of the great pure mathematicians, Hardy

[511] Binmore (1999b), p.131.
[512] Binmore (1999b), p.132.
[513] Binmore (1999b), p.137.
[514] Binmore (1999b), p.119.
[515] Binmore (1999b), p.119.
[516] Binmore (1999b), pp.119-20.
[517] Binmore (1999b), p.120.

> and Littlewood. Eventually, I solved one of the problems they left behind, but my eventual interest in game theory had already surfaced.[518]

What was this seed that was planted in Binmore when he did not even know what game theory was? In part, gambling:

> As a small boy, I was a tireless inventor and player of games. In my late teens, I discovered poker. As a graduate student, I supplemented my meager grant by playing regularly with students who could afford to lose vastly more than I. As was inevitable, the time came when I was wiped out altogether. [...] From this disaster, I learned that a friend is someone who will lend you money when you need it. I seem incapable of learning the other obvious lesson -don't play if you can't afford to lose.[519]

And poker is what eventually interested him when, having to read von Neumann and Morgenstern's (1944) book to prepare a course, he got around to game theory.

> The amount of bluffing he claims to be optimal in his second model is so enormous that I felt von Neumann must have got his mathematics wrong, but by the time I eventually convinced myself that he was right, I found that I was hopelessly hooked.[520]

He describes his conversion to economics as "a typical example of a random walk."[521] "It was only when taking stock of my career in 1980 after finishing a term as Chairman of the LSE Statistics Department, that I realized that I had written no mathematical papers for some years. Instead, I was writing papers about the economics of bargaining."[522] As is the case with Aumann, there is a passion in Binmore for what he does. But where it lies, and where it came from, is really not clear.

We hope that this thesis has pointed to the importance of looking at particular scholars' work in a detailed and open-minded manner. In mathematical disciplines, this means getting acquainted with the formal aspects of the models under study. Much of the intuitions that drive theorists can only be understood through the models they choose to construct. As it is through their mathematics that much of their creativity expresses itself, this should not

[518] Binmore (1999b), p.120.
[519] Binmore (1999b), p.120.
[520] Binmore (1999b), p.122.
[521] Binmore (1999b), p.121.
[522] Binmore (1999b), p.120.

be surprising. The formal and interpretative sections of a paper form a whole that needs to be taken in as such.

An 'individualized' approach eases the task of avoiding to 'peg' authors into prefabricated holes. However, it is also clear that theorists do not work in a vacuum and that ideas from the past can shed light on a modern thinker's positions. In the case of game theory, given the backround of its practitioners, one is naturally led to the history of mathematics. It seems also, however, that the recurrence of 'attitudes' or 'types of minds' in history merits as much reflexion as the recurrence of well-formed and articulated ideas. This idea of 'basic types' of personalities was forcefully impressed upon me by the psychoanalyst C. G. Jung's fascinating autobiography.[523] There, he refers to a small wooden figurine he had carved as a child. Although it had been of great importance to him at that time, he had lost all memory of it after his early childhood. Later in his life:

> Alors, du brouillard de l'enfance, cette fraction du souvenir surgit à nouveau dans une immédiate clarté, quand occupé à préparer [... un livre qu'il s'apprêtait à publier], j'appris l'existence de caches de pierres d'âmes près d'Arlesheim et des *churingas* des Australiens. Je découvris brusquement que je m'étais fait une idée bien précise de ces pierres bien que je n'en eusse jamais vu la moindre reproduction. [...] J'avais l'impression que cette image ne m'était pas inconnue et c'est alors que me revint le souvenir d'un plumier jaunâtre et d'un petit bonhomme. Ce petit bonhomme était un petit dieu caché de l'antiquité [...].
>
> Avec le retour de ce souvenir, j'acquis pour la première fois la conviction qu'il existe des composantes archaïques de l'âme qui ne peuvent avoir pénétré dans l'âme individuelle à partir d'aucune tradition.[524]

Or, as he says elsewhere:

> Bien que nous ayons, nous autres hommes, notre propre vie personnelle, nous n'en sommes pas moins par ailleurs, dans une large mesure, les représentants, les victimes et les promoteurs d'un esprit collectif, dont l'existence se compte en siècles. Nous pouvons, une vie durant, penser que nous suivons nos propres idées sans découvrir jamais que nous n'avons été que des figurants sur la scène du théâtre universel. Car il y a des faits que nous ignorons et qui pourtant influencent notre vie, et ce d'autant plus qu'ils sont inconscients.[525]

[523] Jung (1961).

[524] Jung (1961), pp.42-3.

[525] Jung (1961), p.114. Is not the parallel here between Nietzsche's quote from the introduction and this one remarkable? Jung read and appreciated (as well as pitied) Nietzsche, yet he came across idea of 'ancient types', at least as recounted in his autobiography, mostly through personal experience.

Of course, none of this grand scheme was made explicit in this thesis; it always remained at a suggestive level. I am convinced that in these ideas there is something important, and that they have something relevant to say about Binmore and Aumann, and about scientific discourse in general. Yet I have been unable to do anything more than to put expressions from Binmore and Aumann side by side with those of other scientists, or to point to different debates that in some way 'resemble' this one. Ludwig Wittgenstein made of such 'pointing' one of the main ideas of his *Tractacus*. Referring to the signs that represent the basic undefined objects of our world (which to Wittgenstein are the building blocks of our logical world), he writes:

> What signs fail to express, their application shows. What signs slur over, their application says clearly.
>
> The meaning of primitive signs can be explained by means of elucidations. Elucidations are propositions that contain the primitive signs. So they can only be understood if the meanings of those signs are already known.[526]

Propositions are erected with these atomic logical objects, according to certain logical rules. Yet:

> Propositions cannot represent logical form: it is mirrored in them.
>
> What finds its reflection in language, language cannot represent.
>
> What expresses *itself* in language, *we* cannot express by means of language.
>
> Propositions *show* the logical form of reality. They display it.[527]

Maybe, as Wittgenstein claims of the logical structure of our language, we are also prisoners of certain broad philosophical outlooks on the world in general, so completely intermeshed with our personalities that speaking *of* them is nonsense.[528] To seek them out one would have to point to them. To express them as fully as possible would mean to point with as many fingers as possible to as many cases as possible in as many social situations as

---

[526]Wittgenstein (1961), p.16.

[527]Wittgenstein (1961), p.31.

[528]In justice to Jung, if I should use his thoughts to buttress such fuzzy ideas, I sould mention that he does warn me in the first few pages of his book about the dangers of generalizing, of losing touch with the wealth of telling the 'story' of an individual path through life: "Ce que l'on est selon son intuition intérieure [...], on ne peut l'exprimer qu'au moyen d'un mythe. Celui-ci est plus individuel et exprime la vie plus exactement que ne le fait la science. Cette dernière travaille avec des notions trop moyennes, trop générales, pour pouvoir donner une juste idée de la richesse multiple et subjective d'une vie individuelle." (p.21.)

possible. The task would require readings history, social and natural sciences, psychology and philosophy, as well as 'life experience'.[529] Hayek's (1975) *Two Types of Mind*, Jung's (1950) *Types psychologiques*, Poincaré's logicists and geometers;[530] these disparate works (and how many others?) have something in common. Hopefully the story of Binmore and Aumann's paths and contrasts, as told in this thesis, also reaches towards that undefinable something.

[529]Even game theory may have something to say here. Rubinstein's (2000) *Economics and Language* is very suggestive.

[530]See the quotation in the introduction, pp.5-6.

REFERENCES

Abreu, D. and A. Rubinstein (1988), "The Structure of Nash Equilibrium in Repeated Games With Finite Automata", *Econometrica* 56, 1259-82.

Anderlini, L. (1990), "Communication, Computability and Common Interest Games", Working paper, St. John's College, Cambridge.

Aumann, R. (1960), "Spaces of Measurable Transformations", *Bulletin of the American Mathematical Society* 66, 301-304.

Aumann, R. (1963), "On Choosing a Function at Random", in *Ergodic Theory*, Academic Press, 1-20.

Aumann, R. (1964), "Markets With a Continuum of Traders, *Econometrica* 32, 39-50.

Aumann, R. (1966), "Existence of Competitive Equilibria in Markets with a Continuum of Traders", *Econometrica* 34, 1-17.

Aumann, R. (1967), "A Survey of Cooperative Games Without Side Payments", in *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, Princeton, Princeton University Press, 3-27.

Aumann, R. (1974), "Subjectivity and Correlation In Randomized Strategies", *Journal of Mathematical Economics* 1, 67-96.

Aumann, R. (1976), "Agreeing to Disagree", *Annals of Statistics* 4, 1236-39.

Aumann, R. (1981), "Survey of Repeated Games", in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Vol. 4 of Gesellschaft, Recht, Wirtschaft, Wissenschaftsverlag, Bibliographisches Institut, Mannheim, 11-42.

Aumann, R. (1985), "What is Game Theory Trying to Accomplish?", in Arrow and Honkapohja, Eds., *Frontiers of Economics*, Oxford, Basil Blackwell, 28-88.

Aumann, R. (1987a), "Correlated Equilibrium as an Expression of Bayesian Rationality", *Econometrica* 55, 1-18.

Aumann, R. (1987b), "Economic Theory and Mathematical Method: An Interview", in Feiwer, G.R., Ed., *Arrow and the Ascent of Modern Economic Theory*, London, Macmillan, 306-16.

Aumann, R. (1987c), "Game Theory", in Eatwell J., M. Milgate and P. Newman, Eds., *The New Palgrave: A Dictionary of Economics*, MacMillan Press.

Aumann, R. (1992), "Irrationality in Game Theory", in Dasputa, Gale, Hart and Maskin, Eds., *Economic Analysis of Markets and Games; Essays in Honor of Frank Hahn*, Cambridge, MIT Press, 214-27.

Aumann, R. (1995), "Backward Induction and Common Knowledge of Rationality", *Games and Economic Behavior* 8, 6-19.

Aumann, R. (1996a), "Reply to Binmore and Samuelson", in Arrow *et al.*, Eds, *The Rational Foundations of Economic Behavior: Proceedings of the IEA Conference Held in Turin, Italy*, New York, St-Martin's Press, 130-1.

Aumann, R. (1996b), "Reply to Margalit and Yaari", in Arrow *et al.*, Eds, *The Rational Foundations of Economic Behavior: Proceedings of the IEA Conference Held in Turin, Italy*, New York, St-Martin's Press, 106-7.

Aumann, R. (1996c), "Reply to Binmore", *Games and Economic Behavior* 17, 137-143.

Aumann, R. (1997), "Rationality and Bounded Rationality", *Games and Economic Behavior* 21, 2-14.

Aumann, R. (1998a), "Common Priors: A Reply to Gul", *Econometrica* 66, 929-938.

Aumann, R. (1998b), "On the Centipede Game", *Games and Economic Behavior* 23, 97-105.

Aumann, R. (1998c), "On the State of the Art in Game Theory: An Interview With Robert Aumann", *Games and Economic Behavior* 24, 181-210.

Aumann, R. (1999a), "Interactive Epistemology I: Knowledge", *International Journal of Game Theory* 28, 263-300.

Aumann, R. (1999b), "Interactive Epistemology II: Probability", *International Journal of Game Theory* 28, 301-314.

Aumann, R. (2000), *Collected Papers*, 2 volumes, Cambridge, MIT Press.

Aumann, R. and F. J. Anscombe (1963), "A Definition of Subjective Probability", *Annals of Mathematical Statistics* 34, 199-205.

Aumann, R. and A. Brandenburger (1995), "Epistemic Conditions For Nash Equilibrium", *Econometrica* 63(5) (September), 1161-80.

Aumann, R. and J. Dreze (1975), "Cooperative Games with Coalition Structures", *International Journal of Game Theory* 4, 217-37.

Aumann, R. and R. B. Myerson (1988), "Endogenous Formation of Links between Players and of Coalitions: An Application of the Shapley value", in Roth, A. E., ed., *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, Cambridge, Cambridge University Press, 175-191.

Aumann, R. and S. Sorin (1989), "Cooperation and Bounded Recall", *Games and Economic Behavior* 1, 5-39.

Battigalli, P. and G. Bonanno (1999), "Recent Results on Belief, Knowledge and the Epistemic Foundations of Game Theory", *Research in Economics* 53, 149-225.

Basu, K. (1990), "On the Non-existence of a Rationality Definition for Extensive Games", *International Journal of Game Theory* 19, 33-44

Bell, E. T. (1937), *Men of Mathematics*, New York, Simon and Schuster.

Binmore, K. (1976), "Social Choice and Parties", *Review of Economic Studies* 43, 459-64.

Binmore, K. (1984), "Equilibria In Extensive Games", *The Economic Journal* 95, 51-59.

Binmore, K. (1987a), "Experimental Economics", *European Economic Review* 31, 257-264.

Binmore, K. (1987b), "Modeling Rational Players I", *Economics and Philosophy* 3, 179-214.

Binmore, K. (1988), "Modeling Rational Players II", *Economics and Philosophy* 4, 9-55.

Binmore, K. (1989), "Social Contract I: Harsanyi and Rawls", *The Economic Journal* 99 (Conference), 84-102.

Binmore, K. (1990a), Ed., *Essays on the Foundations of Game Theory*, Oxford, Basil Blackwell.

Binmore, K. (1990b), "Aims and Scope of Game Theory", in Binmore (1990a).

Binmore, K. (1990c), "Preface", in Binmore (1990a).

Binmore, K. (1991a), "DeBayesing Game Theory", Center for Research on Economic and Social Theory Working Paper 16-1991, University of Michigan.

Binmore, K. (1991b), "Evolution and Convention: Social Contract IV", University College London Discussion Paper 91-09.

Binmore, K. (1992a), "Foundations of Game Theory", in J.-J. Laffont, ed., *Advances in Economic Theory*, Sixth World Congress of the Econometric Society, Cambridge, Cambridge University Press.

Binmore, K. (1992b), *Fun and Games*, Lexington, D.C. Heath and Company.

Binmore, K. (1993), *Game Theory and the Social Contract, Volume 1: Playing Fair*, Cambridge, MIT Press.

Binmore, K. (1996), "A Note on Backward Induction", *Games and Economic Behavior* 17, 135-7.

Binmore, K. (1997), "Rationality and Backward Induction", *Journal of Economic Methodology* 4, 23-41.

Binmore, K. (1998), *Game Theory and the Social Contract, Volume 2: Just Playing*, Cambridge, MIT Press.

Binmore, K. (1999a), "Why Experiment in Economics?", *The Economic Journal* 109, F16-24.

Binmore, K. (1999b), "Goats' wool", in Heertje, A, ed., *The Makers of Modern Economics* vol. IV, Cheltenham, Edward Elgar.

Binmore, K. (2001), "Evolutionary Social Theory: Reply to Robert Sugden", *The Economic Journal* 111, F244-48.

Binmore, K. and A. Brandenburger (1990), "Common Knowledge and Game Theory", in Binmore (1990a).

Binmore, K. and P. Dasgupta, eds (1987), *Economics of Bargaining*, Cambridge, Cambridge University Press.

Binmore, K. and M. Herrero (1988a), "Matching and Bargaining in Dynamic Markets", *Review of Economic Studies* 55, 17-32.

Binmore, K. and M. Herrero (1988b), "Security Equilibrium", *Review of Economic Studies* 55, 33-48.

Binmore, K., Kirman, A. and P. Tani (1993), "Introduction", in Binmore, Kirman and Tani eds, *Frontiers of Game Theory*, Cambridge, MIT Press.

Binmore, K., P. Morgan, A. Shaked and J. Sutton (1991), "Do People Exploit Their Bargaining Power? An Experimental Study", *Games and Economic Behavior* 3, 295-322.

Binmore, K. and L. Samuelson (1992), "Evolutionary Stability in Repeated Games Played With Finite Automata", *Journal of Economic Theory* 57, 278-305.

Binmore, K. and L. Samuelson (1994a), "Drift", *European Economic Review* 38, 859-67.

Binmore, K. and L. Samuelson (1994b), "An Economist's Perspective on the Evolution of Norms", *Journal of Institutional and Theoretical Economics* 150, 45-63.

Binmore, K. and L. Samuelson (1994c), "Muddling Through: Noisy Equilibrium Selection", ELSE working paper 036.

Binmore, K. and L. Samuelson (1996), "Rationalizing Backward Induction?", in Arrow *et al.*, Eds, *The Rational Foundations of Economic Behavior: Proceedings of the IEA Conference Held in Turin, Italy*, New York, St-Martin's Press, 111-24.

Binmore, K. and L. Samuelson (2001), "Evolution and Mixed Strategies", *Games and Economic Behavior* 34, 200-226.

Binmore, K., L. Samuelson and R. Vaughan (1995), "Musical chairs: Modeling Noisy Evolution", *Games and Economic Behavior* 11, 1-35.

Binmore, K., A. Shaked and J. Sutton (1985), "Testing Noncooperative Theory: a Preliminary Study", *American Economic Review* 75, 1178-80.

Binmore, K., A. Shaked and J. Sutton (1988), "A Further Test of Noncooperative Bargaining Theory: Reply", *American Economic Review* 78, 837-839.

Binmore, K., A. Shaked and J. Sutton (1989), "An Outside Option Experiment", *Quarterly Journal of Economics* 104, 753-70.

Binmore, K. and H.-S. Shin (1992), "Algorithmic Knowledge and Game Theory", in Bicchieri, C. and M.-L. Dalla-Chiara eds, *Knowledge, Belief, and Strategic Interaction*, Cambridge, Cambridge University Press.

Binmore, K., C. Proulx, L. Samuelson and J. Swierzbinski (1998), "Hard Bargains and Lost Opportunities", *The Economic Journal* 108, 1279-98.

Binmore, K., J. Swierzbinski and C. Proulx (2001), "Does Minimax Work? An Experimental Study", *The Economic Journal* 111, 445-464.

Burke, E. (1909), *Reflections on the French Revolution: A Letter To a Noble Lord*, The Harvard Classics, vol. 24, New York, F. P. Collier and Son.

Cassou-Nogués, P. (2001), *Hilbert*, Paris, Les Belles Lettres.

Courant, R., and H. Robbins (1941), *What is Mathematics?*, New York, Oxford University Press.

Debreu, G. (1992), "Random Walk and Life Philosophy", in Szenberg, M., *Eminent Economists: Their Life Philosophies*, New York, Cambridge University Press, 107-14.

De Finnetti, B. (1974), *Theory of Probability: A Critical Introductory Treatment*, Chichester, John Wiley & Sons.

Dieudonné, J. (1982), "Mathématiques vides et mathématiques significatives", in Guérard, F. and G. Lelièvre, eds., *Penser les mathématiques*, Orléans, Éditions du Seuil.

Ewald, W. B., ed., (1996), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford, Clarendon Press.

Gale,J, K. Binmore and L. Samuelson (1995), "Learning To Be Imperfect: The Ultimatum Game", in *Games and Economic Behavior* 8, 56-90.

Hadamard, J. (1905), "Letter to E. Borel", reprinted in Ewald (1996), 1084-5.

Halmos, P.R. (1985), *I Want To Be A Mathematician: An Automathography*, New York, Springer-Verlag.

Harsanyi, J.C. (1968a), "Games With Incomplete Information Played By "Bayesian" Players" I, *Management Science* 14, 159-82.

Harsanyi, J.C. (1968b), "Games With Incomplete Information Played By "Bayesian" Players" II, *Management Science* 14, 320-34.

Harsanyi, J.C. (1968c), "Games With Incomplete Information Played By "Bayesian" Players" III, *Management Science* 14, 486-502.

Harsanyi, J.C. (1973), "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points", *International Journal of Game Theory* 2, 1-23.

Hayek, F. A. (1975), "Two Types of Mind", *Encounter* 45.

Hilbert, D. (1900), "Mathematical Problems", extracts from the address to the International Congress of Mathematicians in Paris, 1900, reprinted in Ewald (1996), vol.2, 1093-1105.

Hilbert, D. (1920), "Göttingen Lectures", extracts reprinted in Ewald (1996), vol.2, 943-946.

Hilbert, D. (1925), "On the Infinite", in Benacerraf, P. and H. Putnam, *Philosophy of Mathematics*, 2nd edition, Cambridge, Cambridge University Press, 1983.

Hume, D. (1978), *A Treatise of Human Nature*, Oxford, Clarendon Press.

Jung, C. G. (1950), *Types psychologiques*, Genève, Librairie de l'Université.

Jung, C. G. (1961), *Souvenirs, rêves et pensées*, Paris, Gallimard.

Kalai, E. and Smorodinsky M., "Other Solutions to the Nash's Bargaining Problem", *Econo-*

metrica 43, 513-18.

Kandory, M., Mailath, G. and R. Rob (1993), "Learning, Mutation, and Long Run Equilibria in Games", *Econometrica* 61, 29-56.

Khaldûn, I. (1967), *The Muqaddimah*, Princeton, Princeton University Press.

Kleene, S.C. (1950), *Introduction to Metamathematics*, Princeton, D. Van Nostrand.

Klein, C. F. (1895), "The Arithmetization of Mathematics", reprinted in Ewald (1996), pp.965-71.

Klein, C. F. (1911), "On the Mathematical Character of Space-Intuition and the Relation of Pure Mathematics to the Applied Sciences", reprinted in Ewald (1996), pp.958-65.

Kohlberg, E, and J. Mertens (1986), "On the Strategic Stability of Equilibria", *Econometrica* 54, 1003-37.

Kreps, D., and R. Wilson (1982), "Sequential Equilibria", *Econometrica* 50, 863-94.

La Fontaine (1984), "La Besace" in *Fables*, Paris, MVE.

Lebesgue, H. (1905), "Letter to E. Borel", reprinted in Ewald (1996), vol.2, 1080-3.

Lewis, D.K. (1969), *Convention*, Cambridge, Harvard University Press.

Myerson, R. (1991), *Game Theory*, Cambridge, Harvard University Press. Nash, J. (1950). "The Bargaining Problem", *Econometrica* 18, 155-62.

Nietzsche (1886), *Beyond Good and Evil*, Harmondsworth, Penguin Classics.

Osborne, M. and A. Rubinstein (1995), *A Course in Game Theory*, Cambridge, MIT Press.

Poincaré, J. H. (1900), "Intuition and Logic in Mathematics", reprinted in Ewald (1996), vol.2, 1012-20.

Poincaré, J. H. (1905), "Mathematics and Logic: I", reprinted in Ewald (1996), vol.2, 1022-38.

Poincaré, J. H. (1906), "Mathematics and Logic: III", reprinted in Ewald (1996), vol.2, 1052-71.

Poincaré, J. H. (1956), "Mathematical Creation", in Newman, J.R., ed. (1956), *The World of Mathematics*, Vol. 4, New York, Simon and Schuster, 2041-50.

Radner, R. (1968), "Competitive equilibrium under uncertainty", *Econometrica* 36, 31-58.

Radner, R. (1972), "Existence of equilibrium plans, prices, and price expectations in a sequence of markets", *Econometrica* 40, 289-303.

Rawls, J. (1972), *A Theory of Justice*, Oxford, Oxford University Press.

Reid, C. (1970), *Hilbert*, Berlin, Springer-Verlag.

Rubinstein, A. (1986), "Finite Automata Play the Repeated Prisoners' Dilemma", *Journal of Economic Theory* 39, 83-96.

Rubinstein, A. (2000). *Economics and Language*, Cambridge, Cambridge University Press.

Savage, L.J. (1954), *The Foundations of Statistics*, New York, Dover Publications.

Schumpeter, J. A. (1943), *Capitalism, Socialism and Democracy*, London, Routledge.

Selten, R. (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games", *International Journal of Game Theory* 4, 25-55.

Shubik, M (1982), *Game Theory in the Social Sciences: Concepts and Solutions*, Cambridge, MIT Press.

Simon, H. A. (1997), *An Empirically Based Microeconomics: Raffaele Mattioli Lectures*, Cambridge, Cambridge University Press.

Spinoza, B. (1985), "Ethics", In *Collected Works of Spinoza*, vol.I, Princeton, Princeton University Press.

Van Damme, E. (1987), *Stability and Perfection of Nash Equilibria*, New York, Springer Verlag.

von Neumann, J. (1983), "The Formalist Foundations of Mathematics", in Benacerraf, P. and H. Putnam, eds, *Philosophy of Mathematics*, 2nd edition, Cambridge, Cambridge University Press, 1983.

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton, Princeton University Press.

Wittgenstein, L. (1961), *Tractacus Logico-Philosophicus*, London, Routledge.

Young, P. (1993), "The Evolution of Conventions", *Econometrica* 61, 57-84.