Game Theory and Economic Analysis

A quiet revolution in economics

Edited by Christian Schmidt

London and New York

First published in French in 1995 as
Théorie des jeux et analyse économique 50 ans après (special issue of
Revue d'Economie Politique, 1995, no. 4, pp. 529–733)
by Éditions Dalloz (Paris)

This edition published 2002
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2004.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Game theory and economic analysis / [edited by] Christian Schmidt.
p. cm. – (Routledge advances in game theory; 001)
Includes bibliographical references and index.
1. Game theory. 2. Economics. I. Schmidt, Christian. II.
Series.

HB144.G3727 2002
330′.01′5193 – dc21 2001056890

ISBN 0-203-16740-6 Master e-book ISBN

ISBN 0-203-26226-3 (Adobe eReader Format)
ISBN 0–415–25987–8 (Print Edition)

© 1995 Éditions Dalloz
English edition: editorial matter and selection © 2002 Christian
Schmidt; individual chapters © the contributors

Contents

Contributors

Sébastien Cochinard. LESOD, University of Laon, France.

Claude d'Aspremont. CORE, Catholic University of Louvain, France.

Mary Ann Dimand. Albion College, Michigan, USA.

Robert W. Dimand. Brock University, Canada.

The late **Louis-André Gérard-Varet**. Universities of Aix-Marseilles II and III, France.

Hervé Moulin. Rice University, Texas, USA.

Jean-Pierre Ponssard. Laboratoire d'Econométrie, Ecole Polytechnique, Paris, France.

Jean-Louis Rullière. University of Lyons Lumière 2, France.

Christian Schmidt. University of Paris-Dauphine, Paris, France.

Sylvain Sorin. Laboratoire d'Econométrie, Ecole Polytechnique, Paris, France.

Sébastien Steinmetz. INRA, France.

Hervé Tanguy. INRA, France.

Bernard Walliser. HESS, Ecole Nationale des Ponts et Chaussées, France.

Introduction

Christian Schmidt

Game theory has already observed the passage of its fiftieth birthday; that is, if one accepts the conventional chronology which places its birth at the publication of *Theory of Games and Economic Behavior* (*TGEB*) by Von Neumann and Morgenstern (1944). This anniversary evidently did not escape the notice of the Academy of Stockholm, which in 1994 awarded the Nobel Prize in Economic Sciences to three game theorists, Nash, Harsanyi, and Selten. A look back at its brief history brings out several troubling similarities with economic science, in places where one might not expect to find them.

Game theory was invented in order to satisfy a mathematical curiosity. The difficulty at the outset was to find a theoretical solution to the problems posed by uncertainty in games of chance. The example of checkers interested Zermelo (1913), and then the first complete mathematical formulation of strategies for games "in which chance (hasard) and the ability of the players plays a role" was sketched out by Borel (1924), who was himself co-author of a treatise on bridge. Nothing about this singular and rather marginal branch of mathematics would at this time have suggested its later encounter with economics.[1] The analogy between economic activity and what goes on in casinos was only suggested much later, in a far different economic environment than that which these two mathematicians would have been able to observe.

One could say that J. Von Neumann was the person who both conferred a sense of scientific legitimacy upon this mathematical construction, and whose work would lead to the connection with economic analysis.[2] The principal stages were as follows:

- *1928*: Von Neumann demonstrates his *minimax theory*. This demonstration occurs within the framework of a category of two-person zero-sum games in which, to use Borel's terminology, chance (hasard) plays no part, at least no explicit part, and in which the results depend solely upon the reason of the players, not upon their ability. "Strategic games" lend themselves naturally to an economic interpretation (Von Neumann 1928)
- *1937*: Pursuing his topological work on the application of the fixed-point theorem, Von Neumann discovers the existence of a connection between

the *minimax* problem in game theory and the saddle point problem as an equilibrium in economic theory (Von Neumann 1937)

- *1940*: Von Neumann chooses the economist O. Morgenstern to assist him in the composition of what would become the first treatise of game theory. The title of their work is explicit: the theoretical understanding of games is presented as relevant to the analysis of economic behavior.

However seductive it may seem, this saga is nonetheless deceptive. To look a little closer, the bonds that connect Von Neumann's mathematical thought to economic theory are more fragile, and partially contingent. The applicability of strategic games, in the sense of the 1928 article, is obviously not limited to the domain of economics. The connection between the *minimax* theorem and the saddle point is the result of a property of convexity, independent of any economic interpretation of it that might be given. The reasons for Von Neumann's collaboration with Morgenstern go beyond the realm of science. Finally and above all, their work together did not in fact culminate in the announced fusion of game mathematics and the analysis of economic situations. Two-thirds of *Theory of Games and Economic Behavior* are devoted to zero-sum games, and non-zero-sum games are handled with recourse to the device of the "fictitious player." As for Böhm-Bawerk's famous example of the horse market, it represents a particular economic situation that offers only a fragile support for the theoretical result it illustrates. One need only change the numerical givens in the auction market bearing on substitutable but indivisible goods (the horses), and one can demonstrate that the "core" of the allocations is empty (cf. Moulin, this volume: Chapter 4).

Contemporaries were not fooled. As evidenced by the long articles that followed the publication of this massive work, economists did not respond to Von Neumann's and Morgenstern's hopes (cf. Dimand and Dimand, this volume: Chapter 1). Indeed, over the course of twenty years, game theory would remain above all, with only a few exceptions, either an object of study for a small group of mathematicians, or a research tool for military strategists. The first category, working with Kuhn and Tucker at Princeton, would refine, deepen, and generalize the formal properties of the theory left behind by Von Neumann. The second category, which benefited from substantial military funding, worked – particularly in connection with the Rand Corporation – to apply these concepts to new strategic realities by linking them to operational research. A last group of applied mathematicians working around the University of Michigan tried to create a bridge between the statistical approach of decision-making theory and the new theory of games through experimental tests. Among them, emerged the names of Thomson and Raiffa.

But the most suggestive aspect of this history is probably the behavior of Von Neumann himself. Working with the Manhattan project, and having left Princeton, he looked skeptically upon applications of game theory to economics. Shortly before his premature death in 1957, he formulated a critical

judgment which went beyond a simple statement of facts. According to him, there were more than just empirical difficulties standing in the way of the development of such applications. The application of game theory to economics posed a more fundamental problem due to the distance separating several major concepts articulated in *Theory of Games and Economic Behavior* (rules of the game, game solution, coalition, etc.) from the categories constructed by economic analysis.[3] Whatever the case, the small group of economists who persisted in working on games found themselves faced with serious difficulties. In particular, they had to free themselves from the hypothesis of the transferability of utilities: they had to introduce a dynamic into what had been an essentially static treatment of the interactions between the players, and they had to abandon the unrealistic framework of complete information.

A third point of view on the relations between game theory and economic theory would modify matters further. The publication of Nash's profoundly innovative articles in the early 1950s quickly refreshed the thinking of those few economists who had been seduced by game theory, and thereafter they directed their energies towards retrospective reconstructions. Shubik rediscovered in Cournot's work the premises of Nash's concept of equilibrium (Shubik 1955). Harsanyi compared Nash's model of negotiation with economic analyses beginning with Zeuthen and continuing with Hicks (Harsanyi 1956). Similarities came to light between the problematic of competition laid out by Edgeworth and the laws of the market (Shubik 1959). The way was now open for further comparisons. The question could be asked, for instance, whether Shapley's solution did not simply develop, in axiomatic form, several of the ideas suggested by Edgeworth in his youthful utilitarian phase.[4] Those works are to be considered as a starting point for a kind of archaeology. In the train of these discoveries, a hypothesis took shape. An economic game theory perhaps preceded the mathematical theory elaborated by Von Neumann (Schmidt 1990). It is surely not by chance that several of the problems posed by the application of game theory to economics were resolved in the 1960s by the very scholars who had been the most active in researching the economic roots of game theory. One thinks particularly of the work of Shubik, Harsanyi, Shapley, and Aumann.

In the light of these new developments, the role of the Hungarian mathematical genius in this affair appears more complex. While he remains the undeniable intermediary between the mathematics of games and economics, it is necessary also to recognize that he has contributed, through the orientation he gave to his theory (zero-sum games with two players, extension to n players and, only finally, to non-zero-sum games through several fictions), to eclipsing the old strategic approach to economic problems, a tradition illustrated by often isolated economists going back to the nineteenth century. It is true that the tradition always remained hopelessly foreign to his economist collaborator Morgenstern, who was educated in a quite different economic discipline, namely the Austrian school.

At the end of the 1970s, the connections between game theory and economics entered a new phase. The game theory approach had progressively invaded the majority of sectors of economic analysis. Such was the case first of all with industrial economy, which was renewed by the contribution of games. Insurance economics, then monetary economics and financial economics and a part of international economics, all, one by one, were marked by this development. The economy of well-being and of justice have been affected, and more recently the economics of law. It would be difficult today to imagine a course in micro-economics that did not refer to game theory. And at the same time, proportionally fewer and fewer pure mathematicians have been working on game theory; which obviously does not mean that all the mathematical resources applicable to game theory have already been exploited.[5]

The results of the pioneering work of the few economists invoked above have begun to bear fruit. Other, deeper, factors explain this double metamorphosis, of which only one will be mentioned here. In the course of its development, game theory has revealed characteristics that are opposite to those it was initially considered to possess. Far from representing a straitjacket whose application to the analysis of real phenomena imposed a recourse to extremely restrictive hypotheses, it has shown itself, quite to the contrary, to be a rigorous but sufficiently supple language, able to adapt itself to the particular constraints of the situations being studied. In exchange for this flexibility, game theory seems to have lost its original unity. The diversity of game solution concepts and the plurality of equilibria-definitions susceptible to being associated to a single category of games provide the most significant illustrations of this, to say nothing of the ever-increasing number of game types that enrich the theory. The question today is whether the name "game theory" should remain in the singular, or become "game theories" in the plural. This tendency towards fragmentation represents a handicap in the eyes of the mathematician. But for the economist it offers an advantage, to the degree that it brings game theory closer to the economist's more familiar environment: for the plurality of situations and the diversity of perspectives are both the daily bread of economists.

This particular evolution of game theory contradicts the prophesy of its principal founder. The relations between game theory and economic science is in the process of reversing itself. Economics is today no longer the domain of application for a mathematical theory. It has become the engine of development for a branch of knowledge. Indeed, a growing amount of cutting-edge research in game theory is the work of economists or of mathematicians who have converted to economics. The result has been to place the discipline of economics in an extremely unfamiliar position, and to give a reorientation to its developments (renaissance of micro-economics, expansion of experimental economics, new insights in evolutionary economics, first steps in cognitive economics). The first three chapters of the history have been laid out, but it is not over, and no doubt still holds surprises in store.

The ambition for this special edition is to present an image of the many facets characterizing the variety of current contributions of game theory to economics. The contents reflect several major evolutions observed in this domain.

In the middle of the 1980s, the majority of contributions would have dealt with non-cooperative games. What was called "Nash's research program" (Binmore and Dasgupta 1986, 1987; Binmore 1996) dominated the field. The pendulum has now swung back in the other direction and there is a growing interest in cooperative games. The abstract distinction between these two game categories is now clarified. This does not prevent it from seeming unsatisfying, both from the point of view of the classification of the realms of study of theory, as well as from that of their appropriateness to the economic phenomena being studied (Schmidt 2001). It has long been recognized that the analysis of negotiation could adopt one or other point of view. Industrial economics, on the other hand, had up to the present privileged non-cooperative games; but now it makes reference to cooperative games in order to provide a theoretical substratum to the study of coalitions. In the opposite sense, public economics took up the question of the allocation of resources in terms of cooperative games; now, it has begun to discover the fecundity of non-cooperative games, when it extends that line of inquiry through the analysis of the mechanisms of incentive that allow it to be put into practice (cf. the "theory of implementation"). The complementary nature of these developments must not make us forget the existence of a no-bridge between these two approaches. The current efforts of several theoreticians consists in attempting to join them, through various rather unorthodox means (Roth's semistable partitions, Greenberg's theory of social situations, etc.: cf. Cochinard, this volume: Chapter 5).

The subjects of game theory are the players, and not a supposedly omniscient modeler. Only recently have all the consequences of this seemingly banal observation come to light. How ought one to treat the information possessed by the players before and during the game, and how ought one to represent the knowledge they use to interpret it? This question leads to an enlargement of the disciplines involved. The initial dialogue between mathematics and economics which accompanied the first formulation of the theory is coupled with a taking into consideration of the cognitive dimension, which necessarily involves theories of information, logic, and a part of psychology. Thus the definition of a player cannot be reduced to the identification of a group of strategies, as once thought, but requires the construction of a system of information which is associated with him. Thus game theory requires a deeper investigation of the field of epistemic logic (Aumann 1999). If this layer of semantics in game theory enlarges its perspectives, it also holds in store various logical surprises about the foundations of the knowledge it transmits.

As for the new openness towards experimental psychology, it enriches its domain while complicating the game theoretician's methodological task. Making judgments turns out to be delicate when the experimental results

contradict the logical results of the theory, as is the case, for example, with the centipede game.[6] The heart of the difficulty lies in reconciling two different conceptions of the use of game theory. Either one sees it as a storehouse of models for clarifying the economic phenomena one wishes to explain, or one considers it a support for experimentation on interactive behavior in situations close to those studied by economists (cf. Rullière and Walliser, this volume: Chapter 7).

The origin of this volume was a special issue of the *Revue d'Economie Politique* devoted to game theory and published in 1995. From this basis, several papers have been revised and enlarged, some dropped and others added. The chapters that make up this collection fall into two categories. Some lay out in a non-technical way the panorama of a particular branch of the theory, of the evolution of one of its concepts, or of a problem that runs through its development. Others are original contributions bearing on a domain of specific research that, nonetheless, is significant for the field as a whole. All attempt to show how the present situation derives directly or by default from the work of Von Neumann and Morgenstern. The order of arrangement follows the historical chronology of the problem, and its degree of generality in game theory. The contributions are distributed in three parts respectively devoted to historical insight, theoretical content, and applications.

The chapter by R. W. Dimand and M. A. Dimand traces the prehistory, the history, and what one might call the "posthistory" of *TGEB*. In particular, they draw on Léonard's research in shedding light on the role played by Morgenstern. Their presentation leads one to the conviction that, even if the intellectual quality of *TGEB* was assessed favorably, the majority of economists immediately after the war, even in the USA, remained impervious to its message for economic science.

C. Schmidt raises the question of the continuity of game theory between *TGEB* and Nash's contributions during the 1950s. He first captures the aim of the research program contained in *TGEB* and then tries to reconstruct a complete Nash program from his few available papers. Their confrontation shows that Nash, starting from a generalization of Von Neumann's main theorems (1950), quickly developed a quite different framework for studying non-cooperative games, which culminated in his bargaining approach to cooperation (1953). According to this view, Nash obviously appears as a turning point in the recent history of game theory. However, this investigation also reveals an actual gap between the respective programs of Von Neumann and Morgenstern, on one side, and Nash on the other side. Such a gap opens up a domain that remains hardly explored by game theorists until today.

S. Sorin looks at players' strategic use of information. His first concern is to isolate the historic origins of the question which, via Von Neumann and Morgenstern, may be traced back to Borel and Possel. He shows how mixed strategies were conceived of at this period as a strategic use of chance (hasard). He then studies the incidence of the revelation of the players'

strategies (both true and false) regarding the unfolding of the game, starting with the example of poker, which, abundantly treated in *TGEB*, sheds light on the possibilities for manipulating information in a bluff. Finally he extends his field of inquiry to contemporary research on the analysis of signals, of credibility, and of reputation, showing that all these are extensions of the strategic recourse to uncertainty.

H. Moulin offers a state of the question on cooperative games and at the same time develops a personal thesis on its role and its place in the literature of games. Considered as a sort of "second best" by Von Neumann and Morgenstern, cooperative games flourished in the 1960s, with the studies on the heart of an economy, before becoming once again the poor relation of the family. Moulin rejects the interpretation that would see cooperative games as a second-rate domain of research. He maintains, on the contrary, that the models of cooperative games lead back to a different conception of rationality whose origin lies in a grand tradition of liberal political philosophy. After having reviewed the problems posed by the application of the concept of the core to the analysis of economic and social phenomena (economies whose core is empty, economies whose core contains a very high number of optimal allocations), he emphasizes the recent renewal of the normative treatment of cooperative games through the comparison and elaboration of axiomatics that are able to illuminate social choices by integrating, in an analytic manner, equity in the allocation of resources and in the distribution of goods.

In an extension of Moulin's text, S. Cochinard takes on the question of the organization and functioning of coalitions. He especially underlines the fact that coalitions present the theoretician with two distinct but linked questions: how is a coalition formed (external aspect)? and how are its gains shared between the members of the coalition (internal aspect)? The examination of the relation between these two problems orients this chapter. He states first of all that this distinction does not exist in the traditional approach to this question via cooperative games (Von Neumann and Morgenstern's solution, Shapley's solution, Aumann and Maschler's solution, etc.). He reviews the different formulae proposed, and shows that none of them responds to the first problem, which requires an endogenous analysis of the formation of coalitions. Next he explores several approaches to the endogenization of coalitions in a game in following the notion of coalition structure due to Aumann and Drèze (1974). Two conclusions emerge from this study: the very meaning of a coalition varies so widely from one model to the next that there results a great variety of responses to the proposed question; and a convergence is traced out in the results obtained between the approach to the problem via cooperative games and the approach via non-cooperative games. Such an observation suggests another look at the borderline between these two components of game theory.

C. Schmidt considers the connections that persist between the mathematical game theory conceived by Von Neumann and the vast domain assigned to him by researchers today. To illustrate his topic, he analyzes the incidence

of the information a player holds regarding the other players in the definition of rational strategy. He shows first how this question led Von Neumann to formulate two hardly compatible propositions. On the one hand, each player chooses his strategy in complete ignorance of the strategies chosen by the other players; on the other hand the strict determination of the values of the game requires that players' expectations of the others are quite perfect (Von Neumann 1928, 1969), thanks to auxiliary construction, Von Neumann and Morgenstern succeed in making them consistent in *TGEB*. Thus he explains how the suggestions formulated by Von Neumann and Morgenstern came to be at the origin of such heterodox projects as Howard's theory of metagames and Schelling's idea of focal points. Finally, he examines the extensions that might be given them. Metagames lead to a more general analysis of each player's subjective representations of the game, and focal points lead to an innovative approach to the coordination of players' expectations.

The chapter by J.-L. Rullière and B. Walliser bears on the apprehension of the problem of the coordination of strategic choices between independent players. The two authors maintain that game theory has evolved on this question. It started from a strictly hypothetical-deductive approach that supposed in each player the faculty to mentally simulate the reactions of others, while today game theory insists on the players' handling of received information in the course of the development of the game, and on the effects of apprenticeship it can engender. This way of proceeding succeeds in integrating temporality into the process, but raises other difficulties. The authors emphasize in conclusion the epistemological consequences of this transformation of game theory, which caused it to lose its status as a speculative theory and to draw closer to the sciences of observation.

With the chapter by C. d'Aspremont and L.-A. Gérard-Varet, one encounters original research on more particular points of game theory. The two authors examine a few possible developments of non-cooperative games leading to an illumination of incentive mechanisms that satisfy a criterion of collective efficiency. They introduce a general incomplete information model characterized by a Bayesian game. This model permits a mediator who knows the players' utility configuration, the structure of their beliefs, and a result function, to identify the balanced transfers that satisfy a paretian criterion of collective efficiency. Next they analyze the problem of each player's revelation of his private information, which permits them to reduce equilibrium constraints to incentive constraints. In comparing the conclusions yielded by their model with the results obtained by other methods, they are able to specify the domains in which their research may be applied (public oversight, relation between producers and consumers of public goods, judgment procedures, and insurance contracts). While they confirm that collectively efficient incentive mechanisms exist when the phenomena of moral hasard and of anti-selection manifest themselves, the meeting of individual incentives and of collective efficiency is far from being always guaranteed, on account of the different nature of the content of their information.

J.-P. Ponssard, S. Steinmetz, and H. Tanguy's contribution is devoted to an analysis of strategic problems raised by coordination inside firms. The question is investigated through pure coordination team games, where the players have exactly the same payoff functions. Such a general framing is successively applied to two different situations. The firm is supposed to be completely integrated in the first case and decentralized in the second case. The main interest of the exercise is to associate the definition of a precise policy profile to each Nash equilibrium identified, which gives rise to relevant interpretations according to the structural hypotheses chosen. This theoretical approach is supplemented by the interpretation of some experimental results. Finally, the chapter shows a direction where game theory can provide fruitful insights on problems as crucial as the dual coordination decentralization for firms' management.

Notes

1 Borel, however, pointed out the economic application of his tentative theory of games from the very beginning (Borel 1921) and even sketched out a model of price adjustment in a later publication (Borel 1938).
2 This interpretation of Von Neumann's role as an interface between mathematical research and economic theory is buttressed and developed in Dore (1989).
3 See in particular J. Von Neumann, "The impact of recent developments in science on the economy and on economics," (1955) (Von Neumann 1963: Vol. 6). This original diagnostic by Von Neumann was interpreted by Mirowski as the culmination of a process of realizing the unsuitability of the minimax theory to the economic preoccupations manifested in *TGEB* (Mirowski 1992). We prefer to think that this position, which Von Neumann took for the most part before the work on *TGEB*, was based on the obstacles encountered in the application of the method adopted in *TGEB* for the analysis of economic interactions.
4 Provided the value of Shapley is interpreted as the result of putting into play normative principles guiding an equitable allocation, and provided one does not limit Edgeworth's utilitarian work to *Mathematical Psychics* (1881) but goes back to his earlier works.
5 The possibilities offered by "calculability" in the form of Turing machines only began to be explored in a systematic manner by extending the suggestions of Binmore (1987). On finite automata equilibria see Piccione (1992) and Piccione and Rubinstein (1993).
6 Here it is a question of non-cooperative two-player games which unfold according to finite sequences known in advance by the players. The players alternate turns. With each sequence, the total payments are augmented by a coefficient k but their sharing-out between the two players is reversed, so that the possible gain for each player is always less than for the turn immediately following his choice. The logical solution suggested by backward induction would have the first player stop at the first move. But experimental results show, on the contrary, that hardly any player stops at the first move and that very few follow the game to its end (MacKelvey and Palfrey 1992). Indeed, Aumann has demonstrated that when rationality is common knowledge among the players and the game of perfect information, players' rationality logically implies backward induction (Aumann 1995). And so what? The lesson to be drawn from these counterfactuals results remains far from clear (Schmidt 2001).

References

Aumann, R. J. (1994), "Notes on interactive epistemology," mimeograph copy.

Aumann, R. J. (1999), "Interactive epistemology: I and II," *International Journal of Game Theory*, 28, 265–319.

Binmore, K. (1987), "Modeling rational players," *Economics and Philosophy*, 3 and 4.

Binmore, K. (1996), Introduction to *Essays on Game Theory*, J. F. Nash, Cheltenham, Edward Elgar.

Binmore, K. and Dasgupta, P. (1986), eds, *Economic Organizations as Games*, Oxford, Basil Blackwell.

Binmore, K. and Dasgupta, P. (1987), *The Economics of Bargaining*, Oxford, Basil Blackwell.

Borel, E. (1924), "Sur les jeux où interviennent le hasard et l'habilité des joueurs," reproduced as note IV in *Eléments de la théorie des probabilités*, Paris, Librairie Scientifique, Hermann. (NB: the English translation [*Elements of the Theory of Probability*, trans. John E. Freund, Englewood Cliffs, NJ, Prentice-Hall, 1965], is based on the 1950 edition of Borel's work, and therefore does not contain this essay.)

Borel, E. (1938), *Applications aux jeux de hasard*, Paris, Gauthier-Villars.

Borel, E. and Cheron, A. (1940), *Théorie mathématique du bridge à la portée de tous*, Paris, Gauthier-Villars.

Dore, M. (1989), ed., *John Von Neumann and Modern Economics*, Oxford, Clarendon.

Edgeworth, F. Y. (1877), *New and Old Methods of Ethics*, Oxford, James Parker.

Edgeworth, F. Y. (1881), *Mathematical Psychics*, London, Kegan Paul.

Harsanyi, J. C. (1956), "Approaches to the bargaining problem before and after the theory of games: a critical discussion of Zeuthen's, Hick's and Nash's theories," *Econometrica*, 24.

MacKelvey, R. D. and Palfrey, T. R. (1992), "An experimental study of the centipede game," *Econometrica*, 60.

Mirowski, P. (1992), "What were Von Neumann and Morgenstern trying to accomplish?," in Weintraub, E. R., ed., *Toward a History of Game Theory*, Durham, NC, Duke University Press.

Neumann, J. Von (1928), "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100. English translation (1959), "On the theory of games of strategy," in *Contributions to the Theory of Games*, Vol. 4, Tucker, A. W. and Luce, R. D., eds, Princeton, NJ, Princeton University Press, pp. 13–42.

Neumann, J. Von (1937), "Über ein Ökomisches Gleichungssystem and eine Verallgemeinerung des Bronwerschen Fixpunktazes," in *Ergebnisse eins, Mathematisches Kollokium*, 8.

Neumann, J. Von (1963), "The impact of recent developments in science on the economy and economics," (1955) in Taub, A. H., ed., *The Collected Works of Von Neumann*, New York, Pergamon, Vol. 6.

Neumann, J. Von and Morgenstern, O. (1944), *Theory of Games and Economic Behavior*, Princeton, NJ, Princeton Economic Press.

Piccione, M. (1992), "Finite automata equilibria with discounting," *Journal of Economic Theory*, 56, 180–93.

Piccione, M. and Rubinstein, A. (1993) "Finite automata equilibria play a repeated extensive game," *Journal of Economic Theory*, 9, 160–8.

Schmidt, C. (1990), "Game theory and economics: an historical survey," *Revue d'Economie Politique*, 5.

Schmidt, C. (2001), *La théorie des jeux: Essai d'interprétation*, Paris, PUF.
Shubik, M. (1955), "A comparison of treatments of the duopoly problem," *Econometrica*, 23.
Shubik, M., (1959), "Edgeworth market games," in Tucker, A. W., and Luce, R. D., eds, *Contributions to the Theory of Games*, Vol. 4, Princeton, NJ, Princeton University Press.
Zermelo, E. (1913), "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels," in *Proceedings of the Fifth International Congress of Mathematicians*.

Part I

Historical insight

1 Von Neumann and Morgenstern in historical perspective

Robert W. Dimand and Mary Ann Dimand

Introduction

John Von Neumann's and Oskar Morgenstern's *Theory of Games and Economic Behavior* (*TGEB*) (1944) made great advances in the analysis of strategic games and in the axiomatization of measurable utility theory, and drew the attention of economists and other social scientists to these subjects. In the interwar period, several papers and monographs on strategic games had been published, including work by Von Neumann (1928) and Morgenstern (1935) as well as by Émile Borel (1921, 1924, 1927, 1938), Jean Ville (1938), René de Possel (1936), and Hugo Steinhaus (1925), but these were known only to a small community of Continental European mathematicians. Von Neumann and Morgenstern thrust strategic games above the horizon of the economics profession. Their work was the basis for postwar research in game theory, initially as a specialized field with applications to military strategy and statistical decision theory, but eventually permeating industrial organization and public choice and influencing macroeconomics and international trade.

The initial impact of the *Theory of Games* was not based on direct readership of the work. The mathematical training of the typical, or even fairly extraordinary, economist of the time was no preparation for comprehending over six hundred pages of formal reasoning by an economist of the calibre of John Von Neumann, even though Von Neumann and Morgenstern provided much more narration of the analysis than Von Neumann would have offered to an audience of mathematicians. Apart from its effect on Abraham Wald and a few other contributors to *Annals of Mathematics*, the impact of the *Theory of Games* was mediated through the efforts of a small group of eminent and soon-to-be-eminent scholars who read and digested the work, and wrote major review articles. The amount of space accorded these reviews and review articles by journal editors was extraordinary, recalling the controversy following the publication of Keynes's *General Theory*, but there was an important difference. Economists might find the *General Theory* a difficult book, but they read it (until recent years). Apart from the handful of young

Presented at a joint session of the American Economic Association and History of Economics Society, Boston, January 3, 1994.

mathematicians and mathematically inclined economists specializing in the new field of game theory, most economists had to rely on Leonid Hurwicz or Herbert Simon, Richard Stone or Abraham Wald, or another reviewer for a sense of what Von Neumann and Morgenstern had achieved and proposed.

The background

Strategic games have long prehistory. The notion of war as a zero-sum (or constant-sum) game between two players goes back at least to *The Art of War* written by Sun Tzu in the third century BCE or earlier (*Sunzi bingfa*; see Cleary 1988, which also translates eleven classical Chinese commentaries on the work). Emerson Niou and Peter Ordeshcok (1990) credit Sun Tzu with anticipations of dominant and mixed strategies and, with weaker textual support, understanding of minimax strategy. The historical setting for Von Neumann and Morgenstern's *Theory of Games and Economic Behavior* consisted, however, of two sets of writings closer to them in time and place. Several economists, notably Cournot, Edgeworth, Böhm-Bawerk, and Zeuthen, had considered the strategic interaction of market participants (see Schmidt 1990). Between the two world wars, a number of Continental European mathematicians interested in probability theory took the step from games of pure chance to games of strategy. A third strand of work on strategic games, the mathematical models of war and peace devised by Lanchester (1916) and Richardson (1919), remained apart until the 1950s.

Émile Borel (1924) started from Joseph Bertrand's (1889) discussion of the difficulty of finding an optimal pure strategy for the game of chemin de fer. In a series of papers, Borel (1921, 1924, 1927) formulated the concepts of randomization through mixed strategies, which were also defined, elimination of bad (dominated) strategies, and the solution of a strategic game. He found minimax mixed strategy solutions for specific games with finite numbers of pure strategies. He did not, however, prove that two-person zero-sum games would have minimax solutions in general. He initially conjectured that games with larger finite numbers of possible pure strategies would not have minimax solutions, not noticing that this contradicted his conjecture that games with a continuum of strategies would have minimax solutions. Borel expressed his belief that the theory of psychological games would have economic and military applications (see Dimand and Dimand 1992).

John Von Neumann (1928a) stated the minimax theorem for two-person zero-sum games with finite numbers of pure strategies and constructed the first valid proof of the theorem, using a topological approach based on Brouwer's fixed-point theorem. He noted in his paper that his theorem and proof solved a problem posed by Borel, to whom he sent a copy of the paper. Borel published a communication of Von Neumann's result in the proceedings of the Academie des Sciences (Von Neumann 1928b). Von Neumann learned of Borel's work on the subject after completing a preliminary version, but he already knew Zermelo's (1913) proof that the game of

chess has a solution, having corrected an error in the Zermelo paper in correspondence in 1927 (Kuhn and Tucker 1958: 105).

Von Neumann's 1928 minimax paper was acclaimed by René de Possel (1936). Borel explored psychological games further in one number of his vast treatise on probability (Borel 1938). In this work, he analyzed a military allocation game as Colonel Blotto, and his student and collaborator Jean Ville, citing Von Neumann, provided the first elementary, nontopological proof of the minimax theorem and extended the theorem to games with a continuum of strategies (see Dimand and Dimand 1996). Von Neumann and Morgenstern (1944) referred to Borel's (1938) discussion of poker and bluffing and to Ville's minimax proof, which they revised to make it more elementary. Their book did not cite Borel's earlier papers.

Von Neumann continued to display an occasional interest in the mathematics of games during the 1930s. In April 1937, the mathematics section of the *Science News Letter* reported a talk given by Von Neumann at Princeton about such games as stone–scissors–paper and a simplified version of poker. In November 1939 he listed the "theory of games" as a possible topic for his lectures as a visiting professor at the University of Washington the following summer, and mentioned having unpublished material on poker (Leonard 1992: 50; Urs Rellstab, in Weintraub 1992: 90). Most importantly, he cited his 1928a article in his famous paper on general economic equilibrium, published in 1937 in the 1935–6 proceedings of Karl Menger's seminar, noting that "The question whether our problem has a solution is oddly connected with that of a problem occurring in the Theory of Games dealt with elsewhere" (Baumol and Goldfeld 1968: 302n). Even before meeting Oskar Morgenstern in Princeton, Von Neumann was aware that his minimax theorem was relevant to economic theory.

Morgenstern brought to the *Theory of Games* the other stream of work recognized in retrospect as analysis of games: the economic contributions of Cournot on duopoly, and especially Eugen von Böhm-Bawerk on bargaining in a horse market. Böhm-Bawerk was cited five times in Von Neumann and Morgenstern (1944), more often than anyone else except the mathematician Birkhoff.

The treatment of Morgenstern in the literature has been rather curious. He has been credited with encouraging Von Neumann to write on game theory, with the Sherlock Holmes–Moriarty example of Morgenstern (1928, 1935b) and with having "*accidentally* discovered Borel's volume (1938) containing the elementary minimax proof by Ville" (Leonard 1992: 58; Leonard's emphasis). To Philip Mirowski (1992: 130) "the early Oskar Morgenstern looked more or less like a typical Austrian economist of the fourth generation," while Leonard (1992: 52) noted that Morgenstern "remained personally incapable of taking the theoretical steps that he himself envisioned . . . in his continuous agitation for mathematical rigor, he was ultimately calling for a theoretical approach in which thinkers of his own kind would have increasingly little place." These remarks occur in a conference volume (Weintraub

1992) on the occasion of the donation of the Morgenstern papers to the Duke University Library. They do not do justice to the economist who was co-author not only to Von Neumann on game theory but also to Clive Granger on the spectral analysis of stock prices (two articles in Schotter 1976: 329–86; and a book, Granger and Morgenstern 1970) and John Kemeny and G. L. Thompson on mathematical models of expanding Von Neumann economies (three papers Schotter (1976, 73–133) and a book, Morgenstern and Thompson 1976), contributions not cited in the 1992 conference volume.

One early work in particular identifies Morgenstern as a most atypical Austrian economist. The *Encyclopedia of Social Sciences*, commissioning articles by the outstanding experts in their fields, such as Wesley Mitchell on business cycles, Marc Bloch on the feudal system and Simon Kuznets on national income, reached to Vienna to assign a long article on mathematical economics (within the article on economics) to Oskar Morgenstern (1931). This article is listed in the bibliography of Morgenstern's writings in Schotter (1976), but has otherwise been neglected. Although Morgenstern was an economist, not a mathematician, and was very conscious of the contrast between his mathematical training and ability and that of Von Neumann and Wald, he was well acquainted with the existing body of mathematical economics, and his mathematical knowledge was distinguished for the economics profession of his time.

Morgenstern (1931: 366) offered a strikingly heretical reinterpretation of Austrian economics and its founder Carl Menger: "Although Menger did not employ mathematical symbols he is listed by Irving Fisher in his bibliography of mathematical economics and quite properly so, for Menger resorts to mathematical methods of reasoning. This is true also of many later representatives of the Austrian school." He rejected objections to the use of mathematics in economics that "tend to identify mathematics with infinitesimal calculus and overlook the existence of such branches of mathematics as are adapted to dealing with qualities and discrete quantities; moreover mathematics is no more to be identified with the 'mechanical' than ordinary logic" (1931: 364). The application of discrete mathematics to economics is not the only development anticipated by Morgenstern in 1931, for he also criticized Gustav Cassel, who "took over Walras' equations in a simplified form, but in his presentation there are more equations than unknowns; that is, the conditions of equilibrium are overdetermined" (1931: 367). This preceded similar criticisms of Cassel by Neisser in 1932, by Stackelberg and by Zeuthen, the last two in 1933 in the *Zeitschrift für Nationalökonomie*, edited by Morgenstern. Interesting for his knowledge of earlier work are Morgenstern's brief discussions of Cournot (1838), "even at present considered a masterpiece of mathematical economic reasoning," and of Edgeworth, who "originated the idea of the contract curve, which presents the indeterminateness of conditions of exchange between two individuals; it should be said, however, that Menger before him treated the same notion in a non-mathematical form"

(1931: 365, 368). This point about the contract curve was also made in Morgenstern's (1927) obituary of his acquaintance Edgeworth, in which he made the unkept promise that "The substance of Edgeworth's work will be analyzed at another occasion" (Schotter 1976: 478, 480).

What is noteworthy about these early articles by Morgenstern is his eye for what would be of lasting interest in the application of mathematics to economics: Edgeworth's contract curve, the inadequacy of Cassel's attempted proof of existence of general equilibrium, discrete mathematics. Morgenstern was not attracted by more chimerical approaches to economics dressed up in mathematical garb such as business cycle forecasting based on fixed periodicities, Major Douglas's A + B theorem of social credit, or F. Creedy's (1934) *Econometrica* paper explaining economic fluctuations by rigid analogy to Newton's laws of mechanics (assuming, for example, that a constant times the rate of acceleration of spending equals the unspent balance of income, as an analogy to Newton's third law). Morgenstern's first book was an attack on mechanical business cycle forecasts (Morgenstern 1928).

In the 1930s, Morgenstern attended the mathematical colloquium of Karl Menger (son of the economist) and was tutored in mathematics by Abraham Wald, whom Morgenstern, on Menger's recommendation, had hired at the Austrian Institute for Business Cycle Research. Such an attempt at keeping up with the frontier in mathematical economics was highly unusual for an economist of the time. Morgenstern presented his paper on "Perfect foresight and economic equilibrium" (1935b), expounding the problem of strategic interaction, illustrated by Professor Moriarty's pursuit of Sherlock Holmes (1928: 98, 1935b: 173–4; Von Neumann and Morgenstern 1953: 176–8) and citing articles by Menger and Wald, in Menger's colloquium. At the presentation, the Czech mathematician Eduard Cech drew Morgenstern's attention to Von Neumann (1928a) on game theory (Morgenstern 1976: 806). Morgenstern did not, however, meet Von Neumann in Vienna, because Menger and Wald accepted Von Neumann's paper on general equilibrium (in Baumol and Goldfeld 1968) for the proceedings without Von Neumann presenting it in the seminar.

Morgenstern took a particular interest in the work of Schlesinger, Wald, and Von Neumann on the existence of general equilibrium with nonnegative prices (the Walrasian method of counting equations and unknowns failed to count the nonnegativity constraints). After Wald presented his two technical papers on the subject (translated in Baumol and Goldfeld 1968), "In view of the significance of this work and the restricted character of the publication, I persuaded Wald to write an expository article" (Morgenstern 1951: 494). A translation of Wald's expository article was published in *Econometrica* in 1951 as a companion piece to Morgenstern's memorial article. Morgenstern's review article on Hicks extensively cited the Wald and Von Neumann papers from Menger's colloquium in attacking Hicks for attempting to prove the existence of equilibrium by counting equations and unknowns (Morgenstern 1941: 192–9), the first presentation of this research in English, although

carrying the unfulfilled promise that "The discussion of the work done by the two mathematicians, J. Von Neumann and A. Wald, will be reserved for another occasion when more space is available for a presentation of the fundamental principles involved" (1941: 197n).

After meeting John Von Neumann at Princeton, Morgenstern engaged him in the long and fruitful conversation about games that initially was expected to produce a long paper of up to fifty pages for submission to the *Journal of Political Economy*, then a pamphlet of perhaps a hundred pages, then a short book, and finally a book of well over six hundred pages (see Morgenstern 1976). The extended conversation engaged Von Neumann, who did not lack other interests from quantum mechanics to computing, in careful exposition and the exploration of many cases and conditions. The resulting long book full of mathematical notation was not regarded as a commercial proposition by the publisher. Just as Irving Fisher's *Making of Index Numbers* (1922) required the financial support of the monetary heretics Foster and Catchings to be published, the *Theory of Games and Economic Behavior* required a subsidy to the Princeton University Press of $4,000 of Rockefeller money. This source of funding may be related to Morgenstern having directed one of the European business cycle institutes supported by the Rockefeller Foundation. Mirowski (1991: 240) finds another motivation for the subsidy, but his claim that "J. D. Rockefeller . . . at that time happened to be Chief of Naval Operations" is mistaken (and would have surprised Admiral King). Without the extended conversation between Morgenstern and Von Neumann, there would have been no *Theory of Games and Economic Behavior*.

The achievement

To examine psychological games as exhaustively as possible, Von Neumann and Morgenstern elected to use a method of axiom, definition, and successive refinement. This, a novel approach in economics, led them to deal more carefully and explicitly with such issues as the definition of "solution" and a game's information structure and timing than had previous authors. It also led them, aware as they were of the St Petersburg paradox, to consider how to model a player's payoff – another question which had previously been finessed rather than pondered. This motivated their demarcation of conditions under which a Von Neumann–Morgenstern utility function exists, a subsidiary innovation which captured the economics profession earlier than game theory *per se.*

Borel, Von Neumann (1928a, 1928b) and Ville had not questioned whether minimax strategy gave "the" solution to a game. Early game-theoretic writers blithely employed solution concepts which seemed appropriate to the problems they analyzed, whether the issue was some game of chance (Waldegrave, Borel) or the outcomes of voting rules (most notably C. L. Dodgson). Writers of works in economics, on the other hand, often tended (and tend) to equate solution with competitive market clearance, although models of monopoly,

oligopoly and collusion had been discussed frequently, informally since Adam Smith and more formally beginning with Cournot.

Von Neumann and Morgenstern were the first writers to define a concept of static economic equilibrium that did not depend on limiting the form of interaction modeled to perfect competition, or indeed to markets. Von Neumann and Morgenstern specified that

> A set S of elements (imputations) is a solution when it possesses these two properties:
>
> No y contained in S is dominated by an x contained in S.
>
> Every y not contained in S is dominated by some x contained in S.
>
> (Von Neumann and Morgenstern 1947: 40)

Unlike previous treatments of equilibrium, such as the general competitive equilibrium of Walras, Pareto, and Fisher, Von Neumann and Morgenstern's definition of equilibrium did not depend on any particular "rules of the game," although any application of the concept is model-dependent. When bidding was not part of the strategy space he considered, Borel assumed that a game had been solved when players maximized their minimum probability of winning. For Walras, an equilibrium allocation was feasible, and such that consumers maximized utility subject to their budget constraints and producers profit maximized. Von Neumann and Morgenstern's "solution" depended on dominance – on players ruling out strategies which would definitely disadvantage them. The application of "dominance" depends on the objectives of players and the rules of the game played: this definition of solution applies to problems of individual optimization, cooperative games, games of tiddlywinks, and games of politics.

Von Neumann and Morgenstern stressed that where the game permitted and where individuals could benefit from it, coalition formation was crucial to the concept of a solution. Hurwicz (1945: 517) noted that H. von Stackelberg had remarked in 1932 on the possibility of duopolists forming a coalition "[b]ut no rigorous theory is developed for such situations (although an outline of possible developments is given). This is where the *Theory of Games* has made real progress." Considering coalition as an alternative move was analogous to the concerns of Coase (1937) in considering that the formation of coalitions (organizations) might be more efficient than market contracts, although there is little reason to believe either author had read Coase's article. They stated explicitly that their concept of solution was in no sense an optimum, and that it was not in general unique.

Their explicit consideration of information partitions in games (that is, possibly imperfect information), combined with a definition of solution which did not depend on optimality and in which various coalitions might form, delivered multiple equilibria in most games. While writers on market structure such as Stackelberg were interested in explaining and rationalizing multiple equilibria, and Edgeworth emphasized the indeterminacy of

bilateral exchange, recognition of the possibility of multiple equilibria was rare among economists in general. Keynes's *General Theory*, which was general in the sense of considering all states from which there was no tendency for agents to move, had examined multiple equilibria, though in a less systematic form than *The Theory of Games*. Keynes argued that the classical full employment equilibrium was only the limiting case of a range of possible equilibrium levels of employment. It has been observed that, unlike many current game theorists, Von Neumann and Morgenstern were attracted rather than disturbed by a multiplicity of equilibria (Shubik 1992: Mirowski 1992).

Minimax strategies as player objectives stemmed naturally from Von Neumann and Morgenstern's emphasis on zero-sum games, which arose from the concern with gambling by precursors in game theory. In such games A's loss is B's gain, the situation is one of complete conflict, and maximizing the minimum payoff one can achieve if one's opponent plays in a hostile fashion is quite reasonable. Solutions derived from a minimax objective were a subset of solutions as defined by Von Neumann and Morgenstern. These sorts of equilibria, used for much of a book which concentrated on normal form representation and one-time play, were brilliantly critiqued by Daniel Ellsberg (1956). Why, asked Ellsberg, wouldn't a player be willing to take a little more risk for the chance of greater gain? What if a player had some priors on how her opponent was likely to play which indicated the possibility of greater gains by non-minimax strategy?

Among other things, Ellsberg was implicitly targeting a concept tacit in Von Neumann and Morgenstern's book: the assumption of large numbers as a way to deal with behaviour under uncertainty. Von Neumann and Morgenstern meticulously confined themselves to the consideration of games to be played once when they specified that their analysis was static. But in a game of imperfect information to be played once, where players are not obliged to divulge their strategies, it is not clear why they would use a maximin strategy unless they were facing a large pool of potential opponents who might behave in all sorts of ways. In particular, where a mixed strategy is part of an equilibrium, the idea of random play in a one-time game is a problem. It is easy enough to interpret random play by one's opponents on the basis of each opponent coming from a large pool of potential players of different types. It is less easy, however, to rationalize a player's decision to play a mixed strategy in a one-time game unless one assumes the player wishes to tell her opponent her strategy before using it.

A game of imperfect information, such as those in which players move simultaneously, partakes of an uncertainty (noted by Borel) which depends on the play of one's opponents. Indeed, there is such psychological uncertainty about any game which does not have a unique equilibrium in pure strategies. Von Neumann and Morgenstern, whose emphasis was on choice problems with a high degree of interdependence between agents, were chiefly concerned with games in which there was uncertainty. Unlike their predecessors, they were worried about simply taking an expectation of

monetary payoffs (in the case of gambling games) or probabilities of winning (in the case of elections). Aware of the St Petersburg paradox, but also of the advantages of using expected money payoffs, they discussed the conditions legitimizing a Von Neumann–Morgenstern utility function as an apologia for using expected (utility) payoff in a player's criterion. In so doing, they both acknowledged and finessed the problems of measurability and observability which have remained bugbears of experimental games.

A source of uncertainty to the player of a game is that *he cannot know how an opponent values money payoffs* – whether an opponent takes satisfaction in altruism or in revenge, apart from her valuation of augmented income. Shubik's (1992) description of "McCarthy's revenge rule" is an amusing example. This is at least equally a problem to an experimental game theorist, whether an academic or a Williamsonian entrepreneur. It is potentially a great problem in analyzing games, one which Von Neumann and Morgenstern assumed away by positing that individual choice obeyed the axioms which allow the use of expected utility. Game theorists have differed about the importance of the axiomatization of (individually) measurable utility in the *Theory of Games and Economic Behavior*. Some have seen it as essential, others as a desideratum. In a way, it was both. Von Neumann and Morgenstern in effect said, "There is a chasm in our sidewalk; under the following circumstances it does not exist" and stepped over it. Although a number of thinkers had analyzed problems which would later become subjects of game theory, Von Neumann and Morgenstern originally, sometimes in a very game-theoretic style, systematized the questions asked in this branch of choice theory. It was they who first described games as a class, who first delimited a game's information structure, drew a game tree, and defined a solution to a game. Whatever one might think of the Von Neumann–Morgenstern utility function and its role in their book, it must be acknowledged that they looked a substantial difficulty in the face before ignoring it.

The impact

Journal editors allocated surprising amounts of space to reviews of the *Theory of Games*. Jacob Marschak (1946) took nineteen pages in the *Journal of Political Economy*, Leonid Hurwicz (1945) seventeen pages in the *American Economic Review*, Richard Stone (1948) seventeen pages in the *Economic Journal*, E. Justman (1949) eighteen pages in the *Revue d'Économie Politique*, G. K. Chacko (1950) seventeen pages in the *Indian Journal of Economics*, while Carl Kaysen's more skeptical account of "A revolution in economic theory?" (1946) not only occupied fifteen pages of the *Review of Economic Studies* but began on page 1 of the journal's 1946–7 volume, unusual prominence for a review article. G. T. Guilbaud's review in *Économique appliquée* (1949) was longer still, taking forty-five journal pages (twenty-nine in translation). Shorter reviews of four to eight pages appeared in economics journals in Switzerland (Anderson 1949), Denmark (Leunbach 1948), and Sweden

(Ruist 1949). Given normal publishing lags and the need for the reviewers to master 625 pages of technical prose, reviews began to appear quite soon after publication. The first review, in the *American Journal of Sociology*, was by Herbert Simon (1945), who heard about the *Theory of Games* before its publication and within weeks of its appearance "spent most of my 1944 Christmas vacation (days and some nights) reading it" (Simon 1991: 108, 114).

The length of the review articles, and the tone of most of them, expressed excitement and enthusiasm. They introduced such concepts as pure and mixed strategies, randomization, the solution to a game, and the minimax theorem to an audience of economists uneasy with mathematical reasoning and used to thinking about competitive equilibrium rather than strategic interaction. Herbert Simon (1991: 326) recalls that "In 1950, it was still difficult to get a paper published in the *American Economic Review* if it contained equations (diagrams were more acceptable)." Hurwicz's review article, reprinted in the American Economic Association *Readings in Price Theory* and in James Newman's *The World of Mathematics* (1956), eschewed equations, as did the other reviews. This was necessary to make the work accessible to the bulk of the economics profession at a time when a calculus course was not generally required for a doctorate in economics in the United States, and even Keynes's *General Theory* had recently been dismissed as unreadably mathematical by G. D. H. Cole, Reader in Economics at Oxford (M. Cole 1971), and Stephen Leacock, Dow Professor of Economics and Political Science at McGill: Leacock "opened the book but, unfortunately, at one of the few pages with algebraic equations. He thereupon threw it down and, in disgust, as he walked away, said: 'Goldenberg, this is the end of John Maynard Keynes' " (Carl Goldenberg, in Collard 1975: 49).

The barrier to comprehension by economists of the time presented by mathematical expression is illustrated by the response to Von Neumann's paper on general equilibrium in the proceedings of the Menger colloquium. Nicholas Kaldor (1989: viii), to whom Von Neumann sent an off-print, recalled that "Unfortunately the paper was quite beyond me except for the beginning," while Richard Goodwin (1989: 125) "alas, reported back to Schumpeter that it was no more than a piece of mathematical ingenuity." J. R. Hicks (1966: 80n) recalled "from personal recollection, that [Von Neumann] had these things in mind in September 1933, when I met him with Kaidor in Budapest. Of course I did not understand what he was saying!"

The prominence and enthusiasm of this wave of major review articles achieved little in stimulating work on game theory among economists. The economics profession as a whole displayed nothing comparable to the interest and activity generated among mathematics and economics graduate students at Princeton. Even the reviewers themselves wrote little more on game theory, apart from Wald, whose links with Von Neumann and Morgenstern and work extending game theory to statistical decisions predated his review, and Guilbaud (1952, 1960, 1968). Kaysen wrote a paper in 1952 on choice of

strategy under uncertainty, Hurwicz (1953) reflected on "What has happened to the theory of games?", and Stone discussed his original review article in the Royal Economic Society's centenary volume, but otherwise they pursued other interests.

Oskar Morgenstern recorded in his diary (quoted by Mirowski 1991: 239 n. 13) both the hostility of economists when he discussed game theory in seminars (in contrast to the praise of most published reviews) and his impression that they had not read the book. "None of them has read *The Theory of Games*" at Harvard in 1945, "Allais opposed . . . Nobody has even seen the book" in Paris in June 1947, "Röpke even said later that game theory was Viennese coffeehouse gossip" in December 1947, and in Rotterdam in 1950 "They had heard of game theory, but Tinbergen, Frisch, etc. wanted to know nothing about it because it disturbs them." The seminars were at least scheduled and attended, even if without enthusiasm.

At Princeton, Morgenstern's interests were not shared by his colleagues in the economics department and the view that "this new mathematical bag of tricks was of little relevance to economics . . . was put forward in particular by Jacob Viner whose favourite comment on the subject was that if game theory could not even solve the game of chess, how could it be of use in the study of economic life, which is considerably more complex than chess" (Shubik 1992: 153). Viner's attitude was especially unfortunate, for his hostility to mathematical formalism blinded him to the closeness of game theory to his own thought on strategy. In a lecture to the American Philosophical Society in November 1945, published in January 1946, Viner analyzed "The implications of the atomic bomb for international relations." He considered the choice of a strategy on the assumption that the other side will respond by inflicting as much damage as it can: surprise was worthless if the attacked country could still respond with nuclear weapons (Freedman 1981: 28, 42–3; Kaplan 1983: 27). Viner, however, "never was much of a mathematician" (Kaplan 1983: 14) and appears never to have connected his reflections on military strategy to the game theory that he derided.

Aversion to mathematics and failure to read a long, technical book cannot entirely account for the limited response of economists to the *Theory of Games*. The failure of the *Theory of Games* to affect the mainstream of the discipline in the first decades after its publication is shown most clearly by the Cowles Commission for Research in Economics, located at the University of Chicago from 1939 until it moved to Yale as the Cowles Foundation in 1955. Cowles stood out as the centre of mathematical economics, and its research staff would not be disconcerted by the hundreds of pages of mathematical notation used by Von Neumann and Morgenstern. The back cover of the paperback edition (Von Neumann and Morgenstern 1967) quotes effusive praise from four reviews (identified by journal, not reviewer). Three of these reviews were written by members of Cowles: Hurwicz (then at the University of Illinois), Marschak, who was director of research at Cowles, and Simon, then teaching at the Illinois Institute of Technology where he attended the

topology course given by his Illinois Tech colleague, Karl Menger. The Hurwicz and Marschak review articles were reprinted together in 1946 as Cowles Commission Paper no. 13, and were endorsed by Von Neumann and Morgenstern, who recommended these reviews to the "economically interested reader" in the preface to their second edition. At Cowles, if anywhere, game theory could be expected to be taken up by economists.

The list of Cowles Commission and Foundation Papers (reprints) and Discussion Papers in the *Cowles Fiftieth Anniversary* volume (Arrow *et al.* 1991: 109–84) shows what happened. Cowles Commission Paper no. 40 in 1950 by Kenneth Arrow, David Blackwell, and M. A. Girschick concerned Bayes and minimax solutions of sequential decision problems, following Wald's investigation of minimax solutions to statistical decision problems. Cowles Commission Paper no. 75 in 1953 was "Three papers on recent developments in mathematical economics and econometrics" from the *Papers and Proceedings* of the American Economic Association and, together with Tjalling Koopmans on activity analysis and Robert Strotz on cardinal utility, included Hurwicz's reflections on what had become of game theory. Otherwise, there is nothing related to game theory until Martin Shubik, who had been a graduate student in Morgenstern's seminar at Princeton, began appearing in the list with Cowles Foundation Paper no. 164 in 1961. Similarly, among the discussion papers, the only reference to game theory before Shubik's arrival at Cowles was in 1952, when Martin Beckmann considered "The problem of musical chairs and an equivalent 2-person game" (*Discussion Paper* no. 2044) and Leo Tornqvist examined "Some game theoretic points of view on scientific research" (no. 2056). Philip Mirowski (1991: 239) reports finding no papers on game theory among Cowles Discussion Papers 101 to 151, dated April 1947 to April 1950, but, according to the list in the *Cowles Fiftieth Anniversary* volume, the lowest-numbered discussion paper in those years was no. 201 in 1947 (the numbering of the economics discussion papers jumped from 299 for the last in 1950 to 2001 for the first in 1951 because the statistics series had begun with no. 301 and the mathematics series with no. 401, both in 1947). In the Cowles Monograph series, Monograph no. 13, a conference volume on activity analysis in 1951, includes a paper on "iterative solutions of games by fictitious play" by G. W. Brown, who the previous year had collaborated with Von Neumann on "Solution of games by differential equations" in the first volume of Princeton *Contributions to the Theory of Games* (Kuhn and Tucker 1950).

This prolonged paucity of game theory in the publications and discussion papers of the Cowles staff, after the initial laudatory reviews, is startling, given that the Cowles Commission held seven seminars on the theory of games from January to April 1949 (Debreu, in Arrow *et al.* 1991: 30). Instead of this seminar series leading the Cowles researchers into game theory, what caught their attention was Marschak's discussion in one of the seminars of Von Neumann and Morgenstern's axiomatic version of cardinal utility (unique up to a positive linear transformation), notably in an appendix added

to the 1947 second edition. The Von Neumann and Morgenstern theory of measurable utility struck a familiar note, following as it did a long history of controversy over ordinal versus cardinal utility, unlike strategic interaction, reduction of a game-tree to the strategic form of a game, or the stable-set solution of the coalitional form of a game. The Cowles economists were attracted by a new twist to something familiar. The titles of Cowles Commission Discussion Papers nos 226, 2002, 2012, 2021, 2039, 2083, 2105, and 2106 refer to measurable utility. The axiomatic approach of Von Neumann and Morgenstern may also have influenced the axiomatic approach to social choice and general equilibrium theory adopted by such Cowles economists as Arrow and Debreu. Whitehead and Russell had attempted an axiomatization of the foundations of mathematics decades before in their *Principia Mathematica*, and Kolmogorov (1933) had axiomatized the mathematical theory of probability, but economists had not followed their example.

Applied mathematicians responded more strongly to game theory. Copeland (1945) considered that "posterity may regard [the *Theory of Games*] as one of the major scientific achievements of the first half of the twentieth century." The early substantive responses and contributions, as distinct from expository and evaluative reviews, appeared in the Princeton-based *Annals of Mathematics* or in the *Proceedings of the National Academy of Sciences*. Despite publication lags, the 1945 volume carried three game-theoretic articles. Two of them were by Abraham Wald, then at Columbia (initially as Hotelling's research associate) but spending much of his time at the summer home of his wife's family in New Jersey and at nearby Princeton, attending Morgenstern's games seminar and lecturing (Morgenstern 1951 in Schotter 1976: 496–7). Wald (1945a) treated statistical decision as a game against nature, in an examination of statistical decision functions that minimized the maximum risk leading to Wald (1950). The shaping of statistical decision theory, through influence on Wald, was the greatest immediate consequence of the *Theory of Games*. Wald (1947) also provided a non-technical exposition of Von Neumann and Morgenstern's book for readers of the *Review of Economic Statistics* (as it was then named), and lectured on game theory in Paris and Rome on the trip on which he died (Morgenstern 1951 in Schotter 1976: 497). Wald (1945b) extended the minimax theorem for zero-sum two-person to certain cases of a continuum of strategies while Kaplanski (1945) explored the role of pure and mixed strategies in zero-sum two-person games. Between 1950 and 1959, four volumes of *Contributions to the Theory of Games*, edited by H. W. Kuhn and A. W. Tucker and then by M. Drescher, Tucker and P. Wolfe and by R. D. Luce and Tucker, appeared in the series of *Annals of Mathematics Studies* sponsored by the *Annals* through the Princeton University Press. This series published much of the most important work in game theory in that decade. John Nash's paper on "Noncooperative Games," a cornerstone of the next stage of game theory after Von Neumann and Morgenstern (1944), and Julia Bowman Robinson's "An Iterative Method of Solving a Game" both appeared in the *Annals of Mathematics* in

1951. Loomis (1946), Dines (1947) and Nash (1950) were published by the National Academy of Sciences. The economics profession, apart from the handful already specialized in game theory, are unlikely to have looked at the *Annals of Mathematics* or the *Naval Research Logistics Review*, founded in 1954 and coedited by Morgenstern, although the more technically inclined economists would encounter game theoretic articles by Nash, Shubik, and others in *Econometrica*.

The economics profession as a whole in the late 1940s and the 1950s did not take up the interest in game theory encouraged by the book reviewers and shared by Princeton's mathematics department and the strategists at the RAND Corporation and Office of Naval Research. The sheer size of the *Theory of Games* and the mass of mathematical notation, which turned out on closer study to be much more accessible than, say, Von Neumann's 1928 article, impressed the reviewers who had committed themselves to reading the book, rather as readers of other difficult books, such as Keynes's *General Theory* or Marx's *Capital*, develop a vested interest in the importance of what they struggled through. Other economists, unbound by promises to any book review editor and hostile to mathematics, were repelled by these same features of the book. Acceptance by mainstream economists was also not helped by the sharply critical, and even condescending, attitude of Von Neumann and Morgenstern to such eminent works of more conventional economic theory as Hicks's *Value and Capital* (Morgenstern 1941, in Schotter 1976: 185–217) or Samuelson's *Foundations* (Von Neumann quoted in Morgenstern's diary, Mirowski 1991: 239n; cf. Mirowski 1992: 134 on Von Neumann declining to review Samuelson's book). Economists did not regard eminence in another science as a guarantee of soundness in economics, as with Frederick Soddy, the Oxford Nobel laureate in chemistry and monetary heretic. Paul Samuelson (1989: 115–16) listed great mathematicians whose economic writings were undistinguished. The research staff and associates of the Cowles Commission, the outstanding concentration of economists who would not be put off by mathematical formalism, produced an initial flurry of reviews, but the only aspect of Von Neumann and Morgenstern (1947) to capture their lasting attention was the theory of measurable utility.

The community of scholars who responded to the challenge of game theory were the applied mathematicians, notably at Princeton. From their work, a later generation of economists would take the game theory that they applied to industrial organisation, microeconomic theory, macroeconomic policy coordination, and international trade negotiations. The initial long, effusive reviews of Von Neumann and Morgenstern (1944) in economics journals was followed by prolonged neglect by the bulk of the economics profession, but the long-run influence of game theory on the discipline of economics has been great, and the modern field of game theory stems from Von Neumann and Morgenstern. Some landmark works in economics, such as Cournot, were influential only after long delay. Others, such as Keynes's

Treatise on Money, received great attention upon publication, but then faded from the discipline's consciousness. The *Theory of Games* is highly unusual in having faded from the mainstream of economics after being greeted by enthusiastic review articles, but eventually having its intellectual descendants reshape economics.

References

Anderson, O. (1949), "Theorie der Glücksspiele und ökonomisches Verhalten," *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 85: 46–53.

Arrow, K. J. *et al*. (1991), *Cowles Fiftieth Anniversary*. New Haven, CT: Cowles Foundation for Research in Economics at Yale University.

Baumol, W. J. and Goldfeld S. (1968), *Precursors in Mathematical Economics*. London: London School of Economics.

Borel, É. (1921), "La théorie du jeu et les équations, intégrales à noyau symétrique gauche," *Comptes Rendus de l'Académie des Sciences* 173 (December): 1304–8. Translated by L. J. Savage as "The theory of play and integral equations with skew symmetric kernels," *Econometrica* 21: 97–100.

Borel, É. (1924), "Sur les jeux où interviennent l'hasard et l'habilité des joueurs," in *Théorie des probabilités*, Paris. Librairie Scientifique, J. Hermann. Translated by L. J. Savage as "On games that involve chance and the skill of players," *Econometrica* 21: 101–15.

Borel, É. (1927), "Sur les systèmes de formes linéaires à déterminant symétrique gauche et la théorie générale du jeu," *Comptes Rendus de l'Académie des Sciences* 184: 52–3. Translated by L. J. Savage as "On the systems of linear forms of skew symmetric determinant and the general: theory of play," *Econometrica* 21: 116–17.

Borel, É. (1938), *Applications des jeux de hasard, tome 4, fascicle 2* of *Traité du calcul des probabilités et de ses applications*, Paris: Gauthier-Villars.

Chacko, G. K. (1950), "Economic behaviour: A New Theory," *Indian Journal of Economics* 30: 349–65.

Cleary, T. (1988), *Sun Tzu, The Art of War*. Boston, MA, and Shaftesbury, Dorset: Shambala.

Coase, R. H. (1937), "The nature of the firm," *Economica* n.s. 4: 386–405.

Cole, M. I. (1971), *The Life of G. D. H. Cole*. London.

Collard, E. A. (1975), *The McGill You Knew*. Don Mills, ON: Longman.

Copeland, A. H. (1945), "John Von Neumann and Oskar Morgenstern's theory of games and economic behavior," *Bulletin of the American Mathematical Society* 51: 498–504.

Cournot, A. A. (1838), *Recherches sur les principes mathématiques de la théorie des richesses*. Translated by N. T. Bacon as *Researches into the Mathematical Principles of the Theory of Wealth*, New York: Macmillan, 1927.

Creedy, F. (1934), "On the equations of motion of business activity," *Econometrica* 2: 363–80.

De Possel, R. (1936), *Sur la théorie mathématique des jeux de hasard et de reflexion*. Paris: Hermann & Cie, Actualités scientifiques et industrielles, no. 436.

Dimand, R. W., and Dimand, M. A. (1992), "The early history of the theory of strategic games from Waldegrave to Borel," in E. R. Weintraub (1992), 15–28.

Dimand, R. W., and Dimand, M. A. (1996), "From games of pure chance to strategic

games: French probabilists and early game theory," C. Schmidt (ed.) *Uncertainty in Economic Thought*. Cheltenham: Edward Elgar, 157–68.
Dines, L. L. (1947), "On a theorem of Von Neumann," *Proceedings of the National Academy of Sciences, USA* 33: 329–31.
Dore, M. H. I., Chakravarty, S., and Goodwin, R. M., eds (1989), *John Von Neumann and Modern Economics*. Oxford: Clarendon Press.
Ellsberg, D. (1956), "The theory of the reluctant duellist," *American Economic Review* 46: 909–23.
Fisher, I. (1922), *The Making of Index Numbers*. Boston: Houghton, Mifflin for the Pollak Foundation for Economic Research.
Freedman, L. (1981), *The Evolution of Nuclear Strategy*. New York: St Martin's.
Goodwin, R. M. (1989), "Swinging along the autostrada: cyclical fluctuations along the Von Neumann ray," in Dore, Chakravarty, and Goodwin (1989), 125–40.
Granger, C. W. J., and Morgenstern O. (1970), *The Predictability of Stock Market Prices*. Lexington, MA: D. C. Heath.
Guilbaud, G. Th. (1949), "La théorie des jeux: contributions critiques à la théorie de la valeur," *Économie Appliquée* 2: 275–319. Trans. A. L. Minkes as "The theory of games," *International Economic Papers* 1: 37–65.
Guilbaud, G. Th. (1952), "Les problèmes du partage matériaux pour une enquête sur les algèbres et les arithmétiques de la répartition," *Économie Appliquée* 5: 93–137.
Guilbaud, G. Th. (1960), "Faut-il jouer au plus fin? Notes sur l'histoire de la théorie des jeux," *La Décision*, Paris: Centre National de la Recherche Scientifique, 171–82.
Guilbaud, G. Th. (1968), *Éléments de la théorie mathématiques des jeux*. Paris: Dunod.
Hicks, J. R. (1966), "Linear theory," in *Surveys of Economic Theory* Vol. 3, New York: St Martin's for American Economic Association and Royal Economic Society, 75–113.
Hurwicz, L. (1945), "The theory of economic behavior," *American Economic Review* 36: 909–25. Reprinted in *A.E.A. Readings in Price Theory*, G. Stigler and K. Boulding, eds, Chicago: Richard D. Irwin, 1952, 505–26.
Hurwicz, L. (1953), "What has happened to the theory of games?", *American Economic Review* 65: 398–405.
Justman, E. (1949), "La théorie des jeux (Une nouvelle théorie de l'équilibre économique)," *Revue d'Économie Politique*, 5–6: 909–25.
Kaldor, N. (1989), "John Von Neumann: a personal recollection," foreword to Dore, Chakravarty, and Goodwin (1989), vii–xi.
Kaplan, F. (1983), *The Wizards of Armageddon*. New York: Simon and Schuster.
Kaplanski, I. (1945), "A contribution to Von Neumann's theory of games," *Annals of Mathematics* 46: 474–9.
Kaysen, C. (1946), "A revolution in economic theory?", *Review of Economic Studies* 14: 1–15.
Kaysen, C. (1952), "The minimax rule of the theory of games, and the choices of strategies under conditions of uncertainty," *Metroeconomica*, 4: 5–14.
Kolmogorov, A. N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer Verlag. Translated by N. Morrison as *Foundations of the Theory of Probability*. New York: Chelsea Publishing, 1950.
Kuhn, H. W. and Tucker, A. W., eds (1950), *Contributions to the Theory of Games*, 1, Annals of Mathematics Studies, 28. Princeton, NJ: Princeton University Press.

Kuhn, H. W. and Tucker, A. W. (1958), "John Von Neumann's Work in the Theory of Games and Mathematical Economics," *Bulletin of the American Mathematical Society* 64: 100–22.
Lanchester, F. W. (1916), "Mathematics in warfare," in F. W. Lanchester, *Aircraft in Warfare*. Reprinted in J. R. Newman, ed., (1956), 2138–59.
Leonard, R. J. (1992), "Creating a context for game theory," in E. Roy Weintraub, ed., *Toward a History of Game Theory*, annual supplement to *History of Political Economy* 24, Durham, NC: Duke University Press, 29–76.
Leunbach, G. (1948), "Theory of games and economic behavior," *Nordisk Tidsskrift för Teknisk økonomi* 1–4: 175–8.
Loomis, L. H. (1946), "On a theorem of Von Neumann," *Proceedings of the National Academy of Sciences* 32: 213–15.
Marschak, J. (1946), "Neumann's and Morgenstern's new approach to static economics," *Journal of Political Economy* 54: 97–115.
Mirowski, P. (1991), "When games grow deadly serious: the military influence on the evolution of game theory," in C. D. Goodwin, ed., *Economics and National Security, History of Political Economy* Supplement to Vol. 23, Durham and London: Duke University Press, 227–56.
Mirowski, P. (1992), "What were Von Neumann and Morgenstern trying to accomplish?", in E. R. Weintraub, ed., 113–47.
Morgenstern, O. (1928), *Wirtschaftsprognose*. Vienna: Julius Springer-Verlag.
Morgenstern, O. (1931), "Mathematical economics," *Encyclopedia of the Social Sciences*. New York: Macmillan, Vol. 5, 364–8.
Morgenstern, O. (1935a), "The time moment in value theory," as translated in A. Schotter, ed. (1976), 151–67.
Morgenstern, O. (1935b), "Perfect foresight and economic equilibrium," as translated by F. H. Knight in A. Schotter, ed. (1976), 169–83.
Morgenstern, O. (1941), "Professor Hicks on value and capital", *Journal of Political Economy* 49(3): 361–93. As reprinted in A. Schotter ed. (1976), 185–217.
Morgenstern, O. (1951), "Abraham Wald, 1902–1950," *Econometrica* 19(4): 361–7, as reprinted in A. Schotter, ed. (1976), 493–7.
Morgenstern, O. (1976), "The collaboration of Oskar Morgenstern and John Von Neumann on the theory of games," *Journal of Economic Literature* 14: 805–16.
Morgenstern, O. and Thompson, G. L. (1976), *Mathematical Theory of Expanding and Contracting Economies*. Lexington, MA: D. C. Heath.
Nash, J. F. (1950), "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences* 36: 48–9.
Nash, J. F. (1951), "Non-cooperative games," *Annals of Mathematics* 54: 286–95.
Newman, J. R. (1956), *The World of Mathematics*. New York: Simon and Schuster.
Niou, E. M. S., and Ordeshook, P. C. (1990), "A game-theoretic interpretation of Sun Tzu's *The Art of War*," California Institute of Technology Social Science Working Paper 738.
Richardson, L. F. (1919), *The Mathematical Psychology of War*. Oxford: W. Hunt.
Robinson, J. B. (1951), "An iterative method of solving a game," *Annals of Mathematics* 54: 296–301.
Ruist, E. (1949), "Spelteori och ekonomiska problem," *Economisk Tidsskrift* 2: 112–17.
Samuelson, P. A. (1989), "A revisionist view of Von Neumann's growth model," in Dore, Chakravarty, and Goodwin (1989), 100–22.

Schmidt, C. (1990), "Game theory and economics: an historical survey," *Revue d'Économie Politique* 100(5): 589–618.
Schotter, A., ed. (1976), *Selected Economic Writings of Oskar Morgenstern*. New York: New York University Press.
Shubik, M. (1992), "Game theory at Princeton, 1949–1955: a personal reminiscence," in E. R. Weintraub, ed. (1992), 151–64.
Simon, H. (1945), "Review of *The Theory of Games and Economic Behavior*, by J. Von Neumann and O. Morgenstern," *American Journal of Sociology* 27: 558–60.
Simon, H. (1991), *Models of My Life*. New York: Basic Books.
Steinhaus, H. (1925), "Definitions for a theory of games and pursuit", *Mysl Akademicka Lvov* 1: 13–4. As translated by E. Rzymovski with an introduction by H. Kuhn, *Naval Research Logistics Quarterly* 105–8.
Stone, R. N. (1948), "The theory of games," *Economic Journal* 58: 185–201.
Ville, J. (1938), "Sur la théorie générale des jeux où intervient l'habilité des joueurs," in É. Borel (1938), 105–13.
Viner, J. (1946), "The implications of the atomic bomb for international relations," *Proceedings of the American Philosophical Society* 90(1).
Von Neumann, J. (1928a), "Zur theorie der gesellschaftsspiele," *Mathematische Annalen* 100: 295–320. Translated by S. Bargmann as "On the theory of games of strategy," in A. W. Tucker and R. D. Luce, eds, *Contributions to the Theory of Games* 4, Annals of Mathematical Studies 40, Princeton, NJ: Princeton University Press, 1959.
Von Neumann J. (1928b), "Sur la théorie des jeux," *Comptes Rendus de l'Académie des Sciences* 186(25): 1689–91.
Von Neumann J., and Morgenstern, O. (1944, 1947, 1953), *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
Von Neumann J., and Morgenstern, O. (1967), *The Theory of Games and Economic Behavior*. New York: John Wiley and Sons.
Wald, A. (1945a), "Statistical decision functions which minimize the maximum risk," *Annals of Mathematics* 46: 265–80.
Wald, A. (1945b), "Generalization of a theorem by Von Neumann concerning zero-sum two-person games," *Annals of Mathematics* 46: 281–6.
Wald, A. (1947), "*Theory of games and economic behavior* by John Von Neumann and Oskar Morgenstern," *Review of Economic Statistics* 39: 47–52.
Wald, A. (1950), *Statistical Decision Functions*. New York: John Wiley and Sons.
Weintraub, E. R., ed. (1992), *Toward a History of Game Theory. History of Political Economy*, supplement to vol. 24, Durham and London: Duke University Press.
Zermelo, E. (1913), "Uber eine Anwendung der Mengenlehre auf die theorie des Schachspiels," Proceedings, Fifth International Conference of Mathematicians, 2, 501–4. Translated by Ulrich Schwalbe and Paul Walker as "On an application of set theory to the theory of the game of chess," in Eric Rasmusen, ed., *Readings in Games and Information*. Malden, MA: Blackwell, 2001.

2 Rupture versus continuity in game theory

Nash versus Von Neumann and Morgenstern

Christian Schmidt

Introduction

The relationship between game theory and economic theory is neither simple nor one-sided. The history of this relationship started a long time before the beginning of game theory as a mathematical corpus. The views of Cournot and Bertrand on the duopoly are generally considered as its starting point. Such a specific approach to economic situations has been followed by Edgeworth in his bilateral monopoly's treatment and went on until the beginning of the second world war, with the contributions of Zeuthen and Stackelberg. On his side, Borel quickly noticed the possibility of applying some of the concepts he had elaborated for the understanding of games to economic situations (Borel 1923).

Indeed, it is tempting to consider the first edition of *Theory of Games and Economic Behavior* (*TGEB*) in 1944 as the founding act of the theory of games as well as the cornerstone of its privileged alliance with economics. This simplifying vision has its share of mythology. Historians and economists inclined towards the recent history of game theory have unearthed many enigmas which continue to shroud this rather unique vein. Of course this book develops J. Von Neumann's anterior work which would lead to the publishing of "Zur theorie der Gesellschaftsspiele" in 1928. Nevertheless, the nature of the relations between Von Neumann's work and the research undertaken in France during this same period by E. Borel and several of his students remains a question upon which little light has been cast. Also, Morgenstern's contribution to the conception and the editing of *Theory of Games and Economic Behavior* remains difficult to evaluate. References to the economic theory of utility, from the Austrian School, the Lausanne School, as well as certain examples such as the one inspired by Böhm-Bawerk's horse market can be attributed to Morgenstern without difficulty. But beyond these elements, Morgenstern's influence is not easy to discern precisely (Rellstab 1992; Schotter 1992; Leonard 1993) except in the concepts of "accepted standard of behavior" and "established social order" (Schmidt 2001). But none of the economists who participated in these early findings, like Cournot and Edgeworth, is even merely mentioned in *TGEB*.

The real impact of *TGEB*'s publication on economic theory is hard to evaluate. Most of the economic theoreticians were first puzzled and game theory emerged as a new branch of economic analysis a long time after 1944. Such an observation may sound surprising according to the current of mathematical economists who anticipated the game theory approach as previously said. We have put forward an explanation elsewhere (Schmidt 1990). Whereas the analyses of these past economists from Cournot to Zeuthen can be easily connected in retrospect to Nash's non-cooperative approach to game theory, their intellectual linkage to *TGEB*'s inspiration is almost non-existent. A complete understanding of the relationship between game theory and economic theory requires investigating the nature of the filiation from Von Neumann and Morgenstern to Nash's works.

The following contribution is exclusively devoted to a specific facet of this intricate story. Adopting a duly retrospective point of view, we follow the current of questions to which we have just now referred from their "source." In *TGEB* we find the lines of a precise research program. However, only one part of this program was actually brought to term by the two co-authors. This part of the program pertained to the following aspects: axiomatization of a theory of utility, definition of a concept of solution, demonstration of the existence of a solution in the case of a zero-sum two-person game. Different circumstances can be thought to explain how neither of the authors was able to pursue the actual accomplishment of this work. In the 1950s, Nash, one of the researchers in Von Neumann's Princeton University seminar on game theory, engaged himself in another direction of research, exploring the configuration of games called "non-cooperative games" because of the exclusion of coalitions to be framed by the players (Nash 1951). Drawing from this research, Nash would reach a general definition of an equilibrium point susceptible to being associated with all non-cooperative games and upon this basis he renewed the analysis of bargaining. The elements of a research program concerning the theory of games can be gleaned from Nash's four principal articles published between 1950 and 1953 which certain contemporary authors unhesitatingly refer to as "The Nash Program" (Binmore and Dasgupta 1986, 1987; Binmore 1996). If it is unquestionable that the ideas developed by Nash are tributaries of Von Neumann's teaching as well as of the fundamental results demonstrated in *TGEB*, the question we must ask is rather: *Can these ideas be inscribed in the continuum of the* TGEB *program, or rather, on the contrary, do they represent a rupture in its development?*

According to a first interpretation, preferred notably by Aumann, it is continuity which rules the theory of games, at least since the publication of *TGEB* (Aumann 1987b, 1992). On the contrary, according to the interpretive framework particularly upheld by Binmore, Nash's contributions introduced a discontinuity into game theory which followed the publication of *TGEB*. The retrospective evaluation of *TGEB* today closely depends on the response which will be given to this question. Furthermore, the position adopted by

Binmore logically led him to form this abrupt judgment on *TGEB*. "I have read the great classic of game theory from cover to cover, but I do not recommend this experience to others" (Binmore 1992: xxix).

We will present successively the research program elaborated by Von Neumann and Morgenstern in *TGEB* and the strong lines of the research program reconstructed from Nash's work. Their confrontation will then permit us to distinguish that which truly separates the approach adopted by Nash and his followers from the initial approach put forth by Von Neumann and Morgenstern. This juxtaposition will lead us finally to formulate several observations on the evolution of the distinctions between cooperative and non-cooperative games of more general analytical importance.

TGEB guidelines for a research program in game theory

The research program contained in *TGEB* concerning game theory has two principal dimensions. The first dimension results from a fundamental epistemological option. In order to become a scientific and mathematically pertinent object, strategic games, intuitively defined by Von Neumann as early as 1928 are described in terms of an axiomatic procedure.[1] Axiomatization constitutes an essential and logically prior basis for all investigation. Even though the demonstrated solution is only guaranteed in a limited model, as is the case for the minimax theorem in the zero-sum two-person game, this is because the definition of strategic games deduced by means of axioms necessarily affects the legitimacy of successive simplifications for the commodity of mathematical treatment of problems (*TGEB* 523: 48). Thus, the axiomatized definition of the concept of strategy permits us, for example, to justify the elaboration of all strategic games in a normal form (*TGEB*: 84).

The second dimension resides in a classification procedure for strategic games from which a research agenda can be developed. The starting point is provided by the particular case of the zero-sum two-person game. Next, it is generalized by the means of augmenting the number of the players from two to three and then from three to n. Thanks to the concept of coalition introduced with the zero-sum three-person games (*TGEB*: 220–1), it is indeed possible, under certain conditions, to reinterpret all strategic games through the initial framework of the zero-sum two-person game. This allows us to pass from zero-sum games to non-zero sum games by introducing an additional fictitious player as a mathematical device to make the sum and the amounts obtained by the players equal to zero (*TGEB*: 506–7). The introduction of the fictitious player leads us to reinterpret all n-person games as $(n+1)$-person zero-sum games. That is, bearing in mind certain precautions (*TGEB*).

In these two cases the success of the enterprise corroborates with the possibility of constructing the mathematical theory of the object, without being obliged to explain the manner by which coalitions form in the first case nor the way in which the fictional player operates in the second.

The axiomatization of strategic games and the procedure which presides

over the organization of their studies go together in the *TGEB*. This general method leads to consequences which seem paradoxical in terms of common sense. In adopting this perspective, the general theory of strategic games is found in the theory of zero-sum games, itself already present in the theory of the zero-sum two-person game. The notion of the general theory of strategic games runs itself into an impassable limit (in the *TGEB* authors' view) from the very fact that the augmentation of the number of players modifies the possibilities for coalitions and necessitates each time the elaboration of a specific theory, which can be understood as a new generalization of the theory of the zero-sum two-person game. In order to better understand the nature of this difficulty, let us examine the two dimensions of the *TGEB* program in more detail.

Axiomatization

Student of Hilbert and defender of the formalist thesis in the debate about the foundations of mathematics (Von Neumann, 1925, 1927, 1928b, 1932), Von Neumann naturally set out to orient the *TGEB* program towards an axiomatization of game theory. This was so, even if the publication of Gödel's theorem had already persuaded him during the period when he was undertaking the elaboration of *TGEB* with Morgenstern that the formalist program developed by Hilbert led to an impasse in mathematics (Mirowski 1992). In fact, however, *TGEB* uses axiomatization only twice, and this is in order to treat the question of the remaining differences. Axiomatization is that which is used in order to guarantee the logical existence of a utility function in an uncertain universe (*TGEB*: 24–8). Next, it provides a rigorous basis for a general description of strategic games. But the use of the axiomatic approach is not the same in the two cases and its contribution reveals itself to be different concerning game theory.

The axiomatic treatment of utility responds to a limited and precise object. It involves establishing the existence of at least one mapping making utility a number up to linear transformation (*TGEB*: 617). The recourse to axiomatization corresponds here to a procedure similar to its use in pure mathematics. The method followed seems to have been inspired by the tentatives of axiomatization in arithmetic.[2] This analogy depends on an interpretation of the axiomatization of utility developed by Savage. Indeed, according to Savage, the question is to find a system of axioms compatible with the formulation of utility proposed by D. Bernouilli (Savage 1954: 58–9). The axioms presented in *TGEB* satisfy this condition. However, even as Peano's axiomatic is compatible with other mathematical objects than the set of natural numbers, the axiomatic of utility in *TGEB* is also found to be satisfied by functions of utility other than that offered by Bernouilli. This property of the axiomatic of utility constitutes rather an advantage in the context in which it is formulated, limiting thus the potential for analogies with the axiomatization of arithmetic. The fruitfulness of this recourse to axiomatization is questionable

in the context of game theory inasmuch as the links between the axiomatic of utility and the conditions of rationality associated with players' behavior have not been clarified. The research program developped by *TGEB* remains vague regarding this point (*TGEB*: 33), and one would be tempted to hold that this is the missing link at the origin of numerous ulterior misinterpretations concerning the relations between the Bayesian theory of individual decision and the theory of games.[3] A discrepancy may be noticed on these grounds between J. Von Neumann's earlier formulations (Von Neumann, 1926, 1928b) and *TGEB*. In his 1928 article, Von Neumann argues that "draws" are inessential to the understanding of the games of strategy. Thus, nothing is left of "games of chance" in a game of strategy, although the standard probability theory remains thoroughly valid (Von Neumann 1928b: 20–1).

The axiomatization of the general concept of a game is a more ambitious operation. It involves not only translating the intuitive components of the idea of game strategy in light of the formalism of set theory, but also proceeding to a rigorous identification of an object for which the elaboration of the theory or theories is at the center of the program of research. The consequences of this operation have a considerable influence on the program itself. We can first observe that the group of axioms proposed in *TGEB* do not constitute a veritable axiomatic, as it verifies for only two of the conditions which are conventionally associated with all axiomatic systems, namely, independence and non-contradiction. The third condition, categoricity (or completeness), is not satisfied, as the group of axioms proposed is equally compatible with many different games (*TGEB*: 16). This axiomatization of games strategy is thus incomplete. Several interpretations account for this situation. The most radical consists in considering that the phenomenon studied in the theory of games is not axiomatizable as the categories by which it is described elude any homogenous definition.[4]

In whatever way we choose to understand this difficulty, it remains that the axiomatic approach toward games, privileged in the *TGEB* program has direct implications on the orientation of game theory. Completely separating mathematical formalism from the interpretation which seeks to account for it, permits, as previously indicated, reducing the analysis of the players' strategies to the study of the normal form of the game they play (*TGEB*: 84). Moreover, it amounts to introducing the concept of coalition without specifying except by way of literary commentaries, the signification of the notions of "agreements" and "understandings" which are intuitively associated with the idea of a coalition (*TGEB*: 222, 224). This point is important because of the capital role of the concept of coalition in the definition of the solution of an n-person game when $n > 2$ proposed by *TGEB*.

Classification, categorization, and agenda for the research

The tendency toward axiomatization had effectively rendered hopeless the search for a single theory for strategic games. By substitution *TGEB* proposes

the elaboration of theories in terms of categories of games, "utilizing" the same conceptual materials each time, by applying them from one theory to the next. This approach depends upon, first, the identification of a pertinent hierarchy of categories of games, and, second, the discovery of technical procedures permitting the translation of the essential components of the theory of one game into the language of the theory of another, more fundamental, game in the hierarchy. The classification we find in *TGEB* depends, first, upon the distinction between the zero-sum type games and the non-zero-sum games, and, second, upon the number of the players. Definitively, this involves a compromise between two options.[5]

The first option focuses the totality of the research upon the distinction between zero-sum and non-zero-sum games and its foremost aim is to establish a general theory of zero-sum games; the second option puts the emphasis, rather, on the number of players and seeks in priority to construct a general theory of two-person games: "in a general two-person game [i.e. with a variable sum] there is a certain similarity between the general two-person game and the zero-sum three-person game" (*TGEB*: 221, n. 1).

The wish to preserve the link between the solution of the game and the minimax theorem is found, no doubt, at the origin of the final choice in favor of the first option in *TGEB*, as this link is only assured for zero-sum games. In any case, this link is quite determinant in the orientation of the program. In order to measure its incidence on the conceptualization of games, let us consider the notion of cooperation. In accordance with the first option, this notion does not appear in the analysis until a third player is taken into account. The treatment of this third player is reduced to the notion of coalition which is introduced at the same occasion. The adoption of the second option would have led to an approach quite different from this notion of coalition. Nothing excludes the existence of cooperation in a two-person game, provided that it is not a zero-sum game. This situation had already been envisioned by Cournot in the framework of his analyses of the duopoly and the "concours des producteurs" (Cournot 1838; Schmidt 1992). Such cooperation between the two players, provides an occasion for an alliance which can be assimilated under certain conditions to the single-player case. But it cannot be a question of coalition according to the sense of the definition in the framework of the zero-sum three-person game, because the alliance between two players is not derived from their opposition to a third player.

In conclusion, the option retained by *TGEB* permits us to occult the distinction between cooperative and non-cooperative games. The case is not at all the same in the other option where, in the case of the two-person game, one is obliged to make the distinction between the cases where players cooperate and those in which they do not.

The approach followed in *TGEB* leads to the following three categories of games:

- *Category 1*: The zero-sum two-person game
- *Category 2*: The zero-sum $n > 2$-person games[6]
- *Category 3*: The non-zero-sum n-person games

These three categories are not homogenous. The first category is also "only one game," whereas the second and third categories contain as many different games as can be defined by the number n. Thus, each of the games belonging to category 2 requires the elaboration of a specific theory (*TGEB*: The theory of zero-sum three-person games – Chapter 5; The theory of zero-sum four-person games – Chapter 7). This formulation of categories permits us to render precise that which is to be understood by this general theory of zero-sum games in the *TGEB* research program's perspective. At first view, a contradiction does indeed exist between the objective sought, i.e. the formulation of a general theory of the zero-sum n-person game and the process of the research program which theorizes first a zero-sum two-person game (category 1), and from this a theory of the zero-sum three-person game is deduced; then a theory of the zero-sum four-person game and so on. Strictly speaking a general theory of category 2 by successive extensions of the zero-sum two-person game does not exist, at least for cooperative games, furthermore we notice that the degree of precision of the theories obtained varies inversely to the number of players, in such a way that what one gains in extension is automatically lost in comprehensiveness. Conscious of this difficulty as well as the dead-end into which the option of their choice had led them, the authors of *TGEB* fell back upon a weakened acceptance of the general theory. One must then understand this to be the evidence of a more or less general technical procedure which permits the generation of a three-person game from the same conceptual framework used in elaborating the two-person game, then a four-person game based upon the theory of a three-person game and so on. This explains the recourse to a heuristic procedure, the dummy player, which permits the reduction of the four-person game to "an essential three-person game of the players, 1, 2, 3, inflated by the addition of a 'dummy player' " (*TGEB*: 301). We are very far from the initial ambition of finding an axiomatic theory, and this justifies, with slightly different arguments, Mirowski's provocative interpretation of the *TGEB* program as being conceived as an enterprise of deconstruction of the primitive project (Mirowski 1992).[7]

A process of generalization presides over the hierarchization of these categories. This process implies successively letting go of the restrictive hypotheses contained in the preceding category (number of players and payoff functions). But the application of this process did not actually lead to the sought-after classifications in order of increasing generality. This was a serious burden to one of the *TGEB* program's most important research orientations. Indeed, games belonging to category 2 may be defined upon the basis of category 1 by simply weakening the restriction on the number of players which can be involved. But this presentation is not able to yield a more general theory on account of its solutions strongly dependent upon possible

coalitions which vary according to the number of players. The mere extension of the theory concerning the two-person zero-sum game from which the elements for a theory of $n > 2$ person zero-sum games have been attained, has thus failed to achieve a sufficient degree of generality.

This difficulty is aggravated once we follow the same path in order to pass from the category 2 to category 3. By definition, a non-zero-sum game is nothing other than a zero-sum game without the restriction concerning its payoff function. This tautology is however an insufficient justification for reducing the elaboration of the theory for the non-zero-sum game to a simple generalization by extension of the theory of zero-sum games. The authors of *TGEB* are aware that a technical device no longer suffices in passing from category 2 to category 3 (in the way it was supposed to in the initial transformation of category 1 in order to derive the theory for category 2). The principal concepts from the category 2 games, such as the characteristic functions, the domination of a coalition, and the solution of the game, must now be reexamined (*TGEB*: 505–8). In spite of their doubt about the relevance of a simple technical treatment of this matter, the authors persisted nevertheless in formulating their question in similar terms, and the method followed to attain a solution is based upon the same sort of inspiration. This consists in finding a procedure permitting the reduction of category 3 games (and consequently the general *n*-person game), into the general mode with which they treat the zero-sum game; that is, considering the non-zero-sum games as simple extensions of zero-sum games. The case is similar with regard to the introduction of the "dummy player" in order to diminish the complexity of the four-person game. The latter is considered to be essentially like a three-person game. Falling back upon a fictitious player permits them this time to apprehend all *n*-person general games as (*n*+1)-person zero-sum games. But the analogy between these two operations remains imperfect and the introduction of the fictitious player raises additional problems. It cannot be guaranteed, for example, that the compensation of gains and losses for which the fictitious player is fictitiously responsible will have no influence over the course of the game. This means that we cannot necessarily take the fictitious player for a dummy. The way in which the authors of the *TGEB* settle upon the dummy player as their answer to this question reveal the limits of this method. In order for a non-zero-sum game to be treated as an extension of the zero-sum game, the former must share certain characteristics with at least one zero-sum game, and according to this hypothesis, the fictitious player must actually be a dummy one (*TGEB*: 537–8).

The formalist rule is preserved for appearance's sake because the mathematical result is independent of the interpretation given for the fictitious player. But the project which aims at obtaining a general theory of non-zero-sum games, by extending the conceptual framework of the zero-sum games, has at least partially failed because among non-zero-sum games, only certain ones have the same characteristic function as a constant-sum game which is the required condition (*TGEB*: 537–8).

The entire research program developed by *TGEB* reposes definitively upon two hypotheses: (1) for sake of simplicity, the zero-sum two-person game constitutes the essential interpretive model for games; (2) the definition of game theory is to be sought in the elaboration of non-zero-sum games from which this general case can be derived.

These hypotheses are questioned by the general definition of a non-cooperative game (Nash 1950a, 1951). It opens the way to the treatment of cooperative games as being a larger version of the non-cooperative game in following a path which has hardly any relation to the second hypothesis (Nash 1950b, 1953). Nash's four contributions have introduced new concepts which cannot be deduced from the *TGEB*. It remains to be understood whether these new concepts lead to the formulation of a research program which could be considered today as a fully-fledged alternative to the *TGEB* program.

Nash's research program: a retrospective construction

The idea that the four articles published by Nash between 1950 and 1953 are the starting point of a research program in game theory, distinct from *TGEB*, is the result of a retrospective construction. Nash himself considered certain of his results as generalizing concepts present in the *TGEB*. But he also reached other results which he considered as part of a venture into unknown territory. The general concept of the equilibrium point is an illustration of the first case. Nash shows that it represents a generalization of the concept of a solution for the zero-sum two-person game (Nash 1950b, 1951). Other results, such as the existence of at least one equilibrium point for every non-cooperative game as well as the possibility of reducing some cooperative games to non-cooperative games, fall into the second case (Nash 1950a, 1953).

There are several versions of the Nash research program which diverge in their emphasis on different aspects of Nash's work. A first, which we can qualify as an "open version" emphasizes the specific way in which the problems inaugurated by Nash's contributions are addressed, e.g. the distinction between cooperative and non-cooperative games and the research on the logical foundations for game theory (Binmore and Dasgupta 1986, 1987). According to this version, the Nash research program provides a starting point for a criticism of game theory (Binmore 1990, 1992) and its reformulation (Binmore 1996).

A second version, which we can refer to as a "closed version," is more concerned with the solutions proposed by Nash than in the problems he raises. This version places a particular emphasis on the individualistic conception of strategic rationality in Nash's work, consistent with the Bayesian treatment of uncertainty (Harsanyi 1967, 1973, 1975, 1977). Nash's program is also considered as the first stage in a general theory of games based on the concept of an equilibrium point and on the refining of the definitions of equilibrium and rationality (Harsanyi and Selten 1988).

In other words, the reconstruction of "the Nash program" was used by some in order to shed light upon the inadequacies in *TGEB* as well as to identify the problems which confront the current developments in game theory (open version). Others reconstructed this program in order to reinforce the position that game theory is to be investigated as a generalization of the Bayesian theory of individual decision making in uncertainty (closed version).

If we try to reconstitute the actual generation of the Nash research program based on the author's own methodological choices and interests, three features emerge. First, Nash does not share J. Von Neumann's choice of mathematical options. The former considers the axiomatization from an operational point of view, as a tool which can be useful but not exclusive of other approaches (the axiomatization approach in the two-person cooperative game, Nash 1953). Second, Nash quickly showed a certain interest in the experimental part of game theory. He actively participated in the seminar organized in Santa Monica in 1952 on the general topic of "The design of experiments in decision making" (Karlish *et al.* 1954) and took a special interest in the results obtained by the experiment directed at the Rand Corporation by Dresher and Flood, better known as the prisoner's dilemma (Flood 1958).[8] Selten insists upon the importance of the work from this seminar in terms of the ulterior development of experimental game (Selten 1992). Third, Nash's reflexion developed, based on a specific problem which we can summarize in the following terms: how can we link the solution of a game to the analysis of the players' expectations? The question hides an eminently speculative dimension. This is how it led Nash to imagine a general definition of equilibrium, valid for *n*-person games, based on a player's strategic evaluations of the possible outcomes (Nash 1950b). But this also concerns equally important domains of application in game theory, of which bargaining occupies a foremost place (Nash 1950a, 1953). The renewal of the bargaining problem which resulted quickly attracted the attention of economists. Shubik was immediately sensitive to the economic potential offered by Nash's approach which did not need to be derived from zero-sum games (Mayberry *et al.* 1953). On his side, Harsanyi discerned in this approach the elements of a general solution to the economic problem of bargaining by trying to link it with Zeuthen's past work about this problem (Harsanyi 1956).

We will now examine how the combination of these three features culminated in a research program.

To cooperate or not to cooperate

The class of non-cooperative games provides a privileged space for study and resolves, at least partially, the problem which preoccupies Nash. This new awareness is explicit in "non-cooperative games" (Nash 1951). Its conclusion contains the only brief indication which can permit us to historically justify

our attributing a unity to Nash's work, thus qualifying it as a "research program" (Nash 1951: 295).[9] Binmore and Dasgupta were not mistaken about this (Binmore and Dasgupta 1987: 5). Indeed on the one hand, the class of the non-cooperative games is the interpretive domain par excellence for the equilibrium point elaborated previously in order to solve the speculative facet of the Nash problem (see Nash 1950a). On the other hand, the class of non-cooperative games suggests an outlet to investigate the other facet of the question, namely the bargaining problem. Negotiation can be considered as a preplay, analyzed as a part of an enlarged non-cooperative game.

Falling back upon non-cooperative games permits the establishing of two complementary results which give Nash's project the dimension of a veritable research program. He demonstrates that every non-cooperative game has at least one equilibrium point (Nash 1951). Then he establishes that in some cases the analysis of a cooperative game can be reduced to the search for a non-cooperative game corresponding to the chosen model of negotiation (Nash 1950a, 1953).

The distinction between cooperative and non-cooperative games becomes the cornerstone upon which the entire program will be constructed. Moreover, it is interesting to note here that Nash himself did not completely realize what this would later signify. At first glance, this distinction might seem simple. The non-cooperative game is none other, prima facie, than a game which excludes all coalitions and where every player chooses independently a rational strategy. But these considerations stem from the definition of the game in *TGEB* where the notion of coalition, as we have seen, is only introduced in order to bypass the zero-sum two-person game (category 1) to the zero-sum ($n > 2$)-person game (category 2). A deeper analysis reveals that the satisfying definition of a non-cooperative game cannot be deduced from the elimination of coalitions. We still must explain what reasons cause the rational players not to cooperate. In order to have access to this explanation, it is necessary to analyze the particularities of this type of game. The analysis shows that it is not the absence of communication between the players which makes it impossible for them to cooperate, but their inability to reach binding agreements (cf. the well-known example of the prisoner's dilemma).[10] Such a result which has no place in the conceptual framework of *TGEB* is determinant in the course of the development of game theory.

Examining Nash's reinterpretation of the *TGEB* program in terms of the distinction between cooperative and non-cooperative games may help to shed light upon the original links made between this distinction itself and the *TGEB*. In the introduction to "Non-cooperative games" (1951), Nash writes:

> Von Neumann and Morgenstern have developed a very fruitful theory of the two-person zero-sum game in their book *TGEB*. This book also contains a theory of *n*-person games of a type which we would call cooperative. The theory is based on an analysis of the interrelationship of various coalitions which can be founded by the players of the

> game. . . . Our theory in contradistinction is based on the absence of coalition.
>
> (Nash 1951: 256)

According to Nash, first of all, the theory of the zero-sum two-person game which is achieved in *TGEB* can be distinguished from another incompletely elaborated theory concerning *n*-person cooperative games. Second, the two-person zero-sum game theory is the fundamental intersection between the *TGEB* program and Nash's own research directions. More generally speaking, this theory of the two-person zero-sum game constitutes a sort of least common denominator for all research programs in game theory, in such a way that even those most determined to defend the idea that Nash's work constitutes a breach in the continuity of game theory can not deny the evidence of an affiliation in this respect. What differentiates the orientation of Nash's work as compared with the work in *TGEB* lies in the direction chosen by Nash in proceeding towards a generalization of the theory. As opposed to Von Neumann and Morgenstern who, as we have seen, opted for the quest for a general theory of *n*-person zero-sum games, Nash set out in the direction of a general theory of two-person games, easily extended to *n*-person games.

The choice of this option is confirmed by the last of Nash's four articles devoted to two-person cooperative games (Nash 1953). Finally, Nash does not present the distinction between cooperative games "à la Von Neumann and Morgenstern" and his own work on non-cooperative games as necessarily being in opposition. It is, rather, a different starting point for the investigation of the phenomena modeled by game theory. Thus understood, the distinction depends on two extreme hypotheses: the one consists in considering that coalitions are possible between the players (*TGEB*), according to the other, coalitions are impossible (Nash 1951). These two extreme hypotheses have an unquestionable heuristic virtue. However, intermediary situations can be imagined. Let us consider, for example, a two-person pure deterrence game, which corresponds to the situations where the two players mutually deter each other from attacking. At first sight, the game is to be labeled as a non-cooperative game on the grounds that every coalition between the two players is excluded ab initio. However, the agreement they reach is, in a way, self-enforcing, as in a cooperative game (Schmidt 1993).

It remains that the problems encountered and the solutions put forth in order to resolve them are not identical, depending upon whether we have chosen to begin with the hypothesis that coalitions are possible or that they are impossible.

Systems of axioms and models of interpretation

The option retained by Nash for generalizing the two-person zero-sum game sheds light upon the manner in which he proceeded, seeking to apply the

same results he found for non-cooperative games to certain cooperative games. This way does not account for the difference which separates Nash's treatment from the general method followed in the *TGEB*. Each player implements a strategy of retaliation if an agreement is not reached, referring back to a preplay based on threat. A strategy of refusing any sort of alliance cannot be reduced to a simple mathematical device even if it provides a mathematical solution to the bargaining problem. The choice of threat strategies is determinant for the payoffs of the players if the players do not cooperate. Therefore the threats can be interpreted either as a first move of the game or as strategies associated with a threat game. Such a preplay or tacit-game according to Schelling's formulation (Schelling 1960) is a construction which is very different from the one concerning the fictitious player by means of which we can extend category 2 to category 3 of *TGEB*. Incidentally, the threat game which serves as a backdrop for the treatment of the bargaining problem is almost the same as a zero-sum game, as Nash himself remarks (Nash 1953: 136). This similarity is not very surprising however as a two-person game cannot be at the same time zero-sum and cooperative. It indicates only how we can find a result already demonstrated in *TGEB* starting from non-cooperative games, i.e. the Nash option.

The position which Nash adopted regarding axiomatization is not foreign to this difference. Nash, as we have already pointed out, did not share Von Neumann's faith in Hilbert's formalist program for mathematics. For him, axiomatization and modeling must be understood as two complementary approaches ("each of them helps to clarify the other": Nash 1953: 129). This complementarity made the elaboration of the concept of threat possible, and distinct from "demand," which was successfully developed by Harsanyi in the case of his theory of bargaining (Harsanyi 1956, 1973, 1977). If we hold ourselves to the axiomatic approach followed in the second part of Nash's 1953 article, the notion of threat strategy has no place at the level of abstraction where this reasoning is situated. Due to the construction of the model of bargaining in the first part of this chapter, threat strategies must have been introduced in the analysis and have so enriched game theory.

More generally, the position which Nash followed permits him to benefit from the logical distinction between a system of axioms (syntax of the theory) and a model of interpretation (semantics of the theory). The solution of the bargaining problem showed by a system of axioms proposed in the second part of the article has a larger domain of interpretation than the only model which is described in the first part (Nash 1953). It is indeed not necessary that the two players have a mutual benefit to cooperate in order that the bargaining problem has a solution in mathematical terms for the system of axioms. This means, for example, that applying Nash's system of axioms to the Edgeworth bartering problem leads to the no-trade point as the equilibrium solution.[11] On the other hand, the bargaining model analyzed in the first part provides information on the phenomena which the system of

axioms is not able to provide itself. As we have seen, not only are the concepts of threat and demand not made precise here, but also the axiomatic approach does not bring to light the difference between the normal and the extensive form of the game. Or, as long as both players do not necessarily move simultaneously, the game cannot be relevantly reduced to its normal form. Also on this point Nash's work creates a breach in the *TGEB* program which intends to reduce all extensive-form games to normal-form games. Thus, the question becomes an object of controversy among game theoreticians.

The comparison between the article from 1950 and the one from 1953, both devoted to treating the "bargaining problem," reveals the importance of Nash's progress between these two dates. Nash finds as early as 1950 that the manner in which *TGEB* treats the bargaining problem does not in general lead to determining a solution, except for the specific case of the zero-sum two-person game where there is no place for bargaining (Nash 1950b: 157). More precisely the bargaining problem is considered in *TGEB* as a particular case of the extension of the zero-sum game to an *n*-person zero-sum game. (*TGEB*: 608–16). This approach towards cooperative games runs into a multiplicity of possible alliances between the players. Even in posing the bargaining problem in a different way, Nash in his 1950 article remained in any case under the influence of the general method developed by Von Neumann and Morgenstern in *TGEB*.[12] Nash's idea consists in using the numerical utility derived from the *TGEB*'s axiomatic treatment of utility in order to define a two-person's "anticipation" as a combination of two one-person anticipations focusing on the specific sort of anticipation which leads to non-cooperation. In a certain way, the Nash solution for the two-person bargaining problem consists in an extension of the *TGEB* utility theory to rational expectations defined in a narrow meaning. The interpretation which he gives for this extension remains, on these grounds, a sort of offshoot from the general inspiration of *TGEB*. Thus it is with the hypothesis of symmetry that he interprets as expressing an equality in bargaining skill (Nash 1950a: 159) Such an interpretation seems to echo the conclusions in the *TGEB* concerning the bargaining problem and according to which:

> So we observe for the first time how the ability of discernment of a player . . . has a determining influence on his position in bargaining with an ally. It is therefore to be expected that problems of this type can only be settled completely when the psychological conditions refered to are properly and systematically taken into account.
>
> (*TGEB*: 616)[13]

Nash in pursuing his own project provides the elements which form the missing link between the axiomatized theory of utility from Von Neumann and Morgenstern and the analysis of players' strategically rational behaviors. One observes in the mean time that among these, we can find the condition of

independence which lays the cornerstone for the model of expected utility developed later by Savage. In any case, one can legitimately uphold that based upon this sort of argument, Nash's work in his 1950 article can still be considered as a part of the *TGEB* program's continuum.

Things change with Nash's 1953 article. His intellectual relations with Von Neumann deteriorated during the period which separated the publication of these two articles (Shubik 1992). Of course the axiomatic treatment of the bargaining problem developed in the second part of the 1953 article remains formally very close to the 1950 article and for this reason, also, close to the axiomatization of the theory of utility in *TGEB*. But this time Nash places his research in the framework of his general investigation of non-cooperative games (Nash 1953: 128–9). Most of all, he does not use the axiomatic approach in the same way. In the 1950 article, no model of interpretation of the theory is proposed and Nash is content to simply illustrate by means of two numerical examples (Nash 1950a: 160–1). In the article published in 1953, as we have seen, the axiomatic approach comes in the second part of the article as the complement of an actual model of interpretation of a bargaining game. A better adaptation to the model becomes the determining criterion for the modification of the Nash's 1950 system of axioms (Nash 1953: 137–8). One is then far from the formalist inspiration which, under the influence of positions taken by Von Neumann on the foundations of mathematics, impregnates the entire construction of *TGEB*. This change in direction which falls back upon the axiomatic approach in the 1953 article has direct consequences on the development of game theory. It makes the imperfect correspondence which characterizes the relations between a system of axioms destined to provide a logical framework for the theory on the one hand, and, on the other hand, the elaboration of a model which has the objective of establishing the bases for its interpretation. It is precisely in deepening the study of this "gap" that important progress can be made regarding the treatment of the bargaining problem (Roth 1979; Binmore 1987a). This is why the true divergence vis-à-vis the *TGEB* program occurred progressively during the period in which Nash elaborated new directions, rather than at the beginning of his work.

From one research program to another

In order to know whether the Nash program is a continuation of that of *TGEB*, we must examine whether it is possible to find the domain of investigation assigned to strategic games by *TGEB* starting off with the hypotheses which the Nash program had constructed. Roughly speaking, Nash proposes reducing cooperative games to enlarged non-cooperative games by means of his particular treatment of bargaining. Nash showed how the procedure he imagined actually operates in the case of a two-person bargaining, but his extension to *n*-person bargaining shows itself to be rather limited. It consists

of decomposing the *n*-player bargaining into as many two-person bargainings. This approach cannot lead to the same results as those obtained when considering the coalitions in *n*-person cooperative games (cf. the joint-bargaining paradox: Harsanyi 1977: 203–11). Starting from an essential non-cooperative game, Nash's procedure thus does not reach the entire range of possible cooperative games for which *TGEB* has attempted to make the theory (or theories).

Moreover, the result obtained in following this procedure is limited by the ambiguity which accompanies its interpretation. The preplay associated with the solution of the non-cooperative game constitutes indeed a mechanism designed to reinforce the players' commitments. However, if the non-cooperative game is treated as mathematically independent of this preplay, the initial announcements made by the two players during the preplay are integrated into the game itself in the form of constraints in the definitions of the players' payoffs. One may wonder if these preplay interactions are considered as exogenous or endogenous to the non-cooperative game, or at which moment the game is considered to have actually begun – which is almost the same question.[14]

The large category of cooperative games set apart by Von Neumann and Morgenstern cannot be reduced to a particular case of a non-cooperative game as studied by Nash. But the confrontation of these two research programs leads to another conclusion. The analysis of cooperation is definitively excluded from the two programs according to opposite modalities. Either, (1), cooperation is given at the beginning as being a part of the definition of a game by the simple implication of the hypothesis according to which coalitions are possible between players. If so, its study is outside game theory (*TGEB* program). Or, (2), cooperation is eliminated a priori from the definition of a game and is not reintroduced, except, indirectly, in terms of threats in the case of non-cooperation (Nash program). This gap in the approaches of both Von Neumann–Morgenstern and Nash offers researchers a vast and unexplored domain. For the reasons we have recalled, it is the category of cooperative games which we must reexamine, contrary to Binmore's emphasis on non-cooperative games reduced to a so-called "contest" since Nash's work.[15] Reinterpreted in this way, the object studied in *TGEB*, rather than its research program, remains nevertheless, fruitful for future investigation.

Notes

1 The purpose of Chapter 2 is not only to translate the description of the concept of game in the terms of the set theory, but also, more ambitiously, to provide definitions "in the spirit of the axiomatic method" in order to treat them as objects of an exact mathematical investigation (*TGEB*: 74, n. 1).

2 Von Neumann tried through various articles to develop the research that Zermelo and Fraenkel undertook in the axiomatization of set theory. In particular, he presented an axiomatized formalization of his theory of transfinite numbers,

which led him to propose an axiomatized theory of ordinal number arithmetic (Von Neumann 1923; 1925). Von Neumann developed his ideas about the role and limits of axiomatization in mathematics in his last articles devoted to the subject (Von Neumann 1928b, 1929). On this point, see Van Heijenhoort, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, Harvard University Press, 7th edn.

3 The relevance of expectations as numerical values of utility combined with probabilities is briefly discussed in *TGEB* (*TGEB*: 28–9). But the authors do not propose any interpretation of probabilities and merely mention "mathematical expectations" without deciding anything specific regarding their foundations. One can consider that the use of probabilities constitutes as well a mathematical device. On the other hand, it is worthwile noticing that the notion of mixed strategy, already found in Borel's work (Borel 1923; 1926), is introduced in *TGEB* by the means of examples such as "matching pennies" and "stones–scissors–papers," which are zero-sum two-person games for which the mixed strategy is simply understood as "statistical" strategy (*TGEB*: 144). Later, however, authors like Harsanyi developed a subjective interpretation of the probabilities used in the definition of mixed strategies. This particular interpretation led Harsanyi to hold the supposition that players individually randomize joint strategies according to a degree of statistical correlation among players' subjective randomized strategies (Harsanyi 1977: 97). In this way, Aumann definitely integrated the Bayesian rationality into game theory via the concept of "correlated equilibrium" as soon as all the priors are common to the players and that this "common prior" is common knowledge (Aumann 1987a). But such a reformulation of a game is far from the initial guidelines found in *TGEB*.

4 Von Neumann and Morgenstern recognize that their description of games does not satisfy the categoricity principle (*TGEB*: 76). They are aware of the consequences of this situation since they write in the following note: "This is an important distinction in the general logistic approach to axiomatization. Thus the axioms of Euclidean geometry describe a unique object. While those of group theory (in mathematics) or of rational mechanics (in physics) do not, since there exist many different groups and many different mechanical systems" (*TGEB*: 76, note 3).

This remark directly follows Von Neumann's comments on the lack of categoricity in the axiomatization of set theory (Von Neumann 1925, 1963: 76). More precisely it must be related to his attempt to give an axiomatical formulation to the quantic theory (Von Neumann 1932). One may perhaps understand this note of *TGEB* as the manifestation of a doubt upon the feasibility of Hilbert's program, strictly speaking (Mirowski 1992). Regarding the different texts to which we refer here, Von Neumann does not seem to reach the same radical conclusion as we have. This conclusion constitutes, however, a direct extension of his assertion related to the axiomatization of games. The lack of categoricity of the systems describing game theory may be meaningfully related to the heterogeneity of the categories used to describe the economic and social phenomena which are the true topic of game theory (Von Neumann 1963).

5 The two options were conceived by Von Neumann and Morgenstern, who wrote in the beginning of Chapter 3; "Afterwards there is a choice of dealing either with the general two-person games or with the zero-sum three-person games. It will be seen that our technique of discussion necessitates taking up the zero-sum three-person game first. After that we shall extend the theory to the zero-sum n-person game (for all $n = 1,2,3 \ldots$) and only subsequently to this will it be found convenient to investigate the general n-person game" (*TGEB*: 87–8).

6 Let us recall the key role played by the zero-sum three-person games in this construction, because this category is the cornerstone of coalition formed by the players.

7 The "deconstruction" consists less, from our point of view, in the axiomatization project itself, but in the original idea of a general theory of games.
8 Nash's comments on this experiment have been extensively discussed by Roth (1993).
9 "The writer has developed a dynamic approach to the study of cooperative games based upon reduction to non-cooperative form. One proceeds by constructing a model of pre-play negotiation so that the steps of negotiation become moves in a larger non-cooperative game. . . . This larger game is then treated in terms of the theory of this paper. . . . Thus the problem of analyzing a cooperative game becomes the problem of obtaining a suitable and convincing non-cooperative model for negotiation" (Nash 1951: 295). Let us notice that according to this program, Nash focuses on the approximate computational methods as the only way for analyzing more complex games than the simple models in this chapter. Such a suggestion must be understood as the first step in the recognition of the limits of the axiomatic approach.
10 The fact that the players are free to talk and negotiate an agreement is clearly irrelevant as long as such agreements are neither binding nor enforceable by the rules of the game, i.e. as long as the game is cooperative (Harsanyi and Selten 1988: 2–4).
11 "Bartering" is a special kind of bargaining in which contracts between traders are directly expressed in terms of quantities of the two delivered commodities (Binmore 1987: 239–56).
12 In one of his 1950 articles, Nash thanks Von Neumann and Morgenstern "who read the original form of the paper and gave helpful advice to the presentation" (Nash 1950b: 155). Furthermore he strictly follows the abstract mathematical approach advocated by Von Neumann, explicitly starting from the axiomatization of utility of *TGEB* (see axioms 1 to 5 which summarize the appendix of the axiomatic treatment of utility in *TGEB*). The actual changes are expressed by means of axioms 7 and 8, which correspond to an extension of Von Neumann and Morgenstern's system of bargaining in a non-cooperative situation.
13 As Binmore and Dasgupta rightly observe, in his 1953 article Nash discarded the irrelevant assumption concerning the equal psychological bargaining ability used in his 1950 article (Binmore and Dasgupta 1987: 6). Thus the field that Nash assigned to the bargaining theory in his 1953 article no longer coincides with that of *TGEB* since it is absolutely independent of any psychological consideration upon the players' ability to negotiate. This is a major consequence of Nash's non-cooperative approach to the bargaining problem. One can however discuss its argumentation. Indeed Nash considers that the assumption of complete information makes it meaningless to refer to the "bargaining ability" (Nash 1953: 138). In fact, one can argue that Nash opts for this approach, not because the model he develops is a game of complete information, but because this approach implicitly refers to a common knowledge rationality hypothesis.
14 Nash's own interpretation is not completely clear. While the formal negotiation model he describes is divided into four stages, only two stages are devoted to the player's decision and move. Finally, considering the second move separately, the payoff function of the demand game is determined by threats at an anterior stage and the demand game is actually taken into account from a strict game-theoretic point of view (Nash 1953: 131–2). Thus the actual status of the threat game remains arguable.
15 Binmore and Dasgupta call "contest" any non-cooperative game without preplay communication, i.e. before the beginning of the formal game. According to this definition, a contest obviously belongs to a very specific class of non-cooperative games. But the most important question is not, as Binmore and Dasgupta argue, the question of the existence of a set of non-cooperative games that are not

"contests" and that are still not studied (Binmore and Dasgupta 1986–7). The actual question aims at the investigation of coordination mechanisms in both categories of games. Nowadays, the analysis of coordination is susceptible to changing its formulation (cf. Schelling 1960: Chapter 4 and appendix C). In the theory of cooperative games developed by Von Neumann and Morgenstern in *TGEB* coordination is considered as belonging to the psychological and social fields, and for this reason, outside the theory. In the theory of non-cooperative games, its importance has evolved recently. The coordination problem has been mainly explored by the means of a refinement of rationality assumptions. But a lot of work is still to be accomplished on this question, which should begin with the integration of the analysis of the coordination within the theory of cooperative games through an analysis of players' beliefs. One would then reconsider the object of *TGEB* without following on this point the two authors' conjecture.

References

Aumann, R. J. (1987a), "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, 55, 1.

Aumann, R. J. (1987b), "Game theory," in *Game Theory: The New Palgrave, a Dictionary of Economics*, Eatwell, J., Milgate, M. and Newman, P., eds, London, Macmillan, 1–53.

Aumann, R. J. (1992), "Irrationality in game theory," in *Economic Analysis of Markets and Games*, Dasgupta, P., Gale, D., Hart, D. and Maskin, E., eds, Cambridge, MA, MIT Press.

Binmore, K. (1987a), "Nash bargaining theory: 1", in *The Economics of Bargaining*, Binmore, K. and Dasgupta, P., eds, Oxford, Basil Blackwell.

Binmore, K. (1987b), "Nash bargaining theory: 3", in *The Economics of Bargaining*, Binmore, K. and Dasgupta, P., eds, Oxford, Basil Blackwell.

Binmore, K. (1990), *Essays on the Foundations of Game Theory*, Oxford, Basil Blackwell.

Binmore, K. (1992), *Fun and Games*, Lexington, MA, Heath and Co.

Binmore, K. (ed.) (1996), *Essays on Game Theory, John F. Nash*, Cheltenham, E. Elgar.

Binmore, K. and Dasgupta, P. (1986), "Game theory: a survey," in *Economic Organizations as Games*, Binmore, K. and Dasgupta, P., eds, Oxford, Basil Blackwell, 1–48.

Binmore, K. and Dasgupta, P. (1987), "Nash bargaining theory: an introduction," in *The Economics of Bargaining*, Binmore, K. and Dasgupta, P., eds, Oxford, Basil Blackwell, 1–26.

Borel, E. (1923), "Sur les jeux de hasard où interviennent le hasard et l'habileté des joueurs," *Association Française pour l'armement des sciences*, 3.

Borel, E. (1926), "Un théorème sur les systèmes et formes linéaires à déterminant symétrique gauche," *Compte rendu de l'Académie des Sciences*, 183.

Cournot, A. (1838) [1974], *Recherches sur les Principes Mathématiques de la Richesse*, Paris, Calmann-Lévy.

Flood, M. (1958), "Some experimental games," *Research memorandum* RM 789, Santa Monica, Rand Corporation.

Harsanyi, J. C. (1956), "Approaches to the bargaining problem before and after the theory of games," *Econometrica*, 24.

Harsanyi, J. C. (1967a), "Games with incomplete information played by 'Bayesian' players: Part I," *Management Science*, 14.
Harsanyi, J. C. (1967b), "Games with incomplete information played by 'Bayesian' players: Part II," *Management Science*, 14.
Harsanyi, J. C. (1968), "Games with incomplete information played by 'Bayesian' players: Part III," *Management Science*, 14.
Harsanyi, J. C. (1973), "Games with randomly disturbed pay-offs: a new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, 2.
Harsanyi, J. C. (1975), "The tracing procedure: a Bayesian approach to defining a solution in n-person games," *International Journal of Game Theory*, 2.
Harsanyi, J. C. (1977), *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge, Cambridge University Press.
Harsanyi, J. C. and Selten, R. (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, MA, MIT Press.
Karlisch, G. K., Milnor, J. W., Nash, J. F., and Nering, E. D. (1954), "Some experimental n-person games," in *Decision Processes*, Thrall, R. M., Coombs, C. H., and Davis, R. L., eds, New York, John Wiley.
Leonard, R. J. (1993), "Dr Morgenstern and game theory," oral communication to the ACGHPE meeting, Paris.
Mayberry, J. J., Nash, J. F., and Shubik, M. (1953), "A comparison of treatment of a duopoly situation," *Econometrica*, 21.
Mirowski, P. (1992), "What were Von Neumann and Morgenstern trying to accomplish?", in *Toward a History of Game Theory*, Weintraub, E. R., ed., Durham, NC, Duke University Press 113–47.
Nash, J. F. (1950a), "The bargaining problem," *Econometrica*, 18.
Nash, J. F. (1950b), "Equilibrium points in n-person games," *Proceedings of the National Academy of Science*, 36.
Nash, J. F. (1951), "Non-cooperative games," *Annals of Mathematics*, 54.
Nash, J. F. (1953), "Two-person cooperative games," *Econometrica*, 21.
Rellstab, V. (1992), "New insights into the collaboration between John Von Neumann and Oskar Morgenstern, *On the Theory of Games and Economic Behaviour*," in *Toward a History of Game Theory*, Weintraub, E. R., ed., Durham, NC, Duke University Press, 77–93.
Roth, A. (1979), *Axiomatic Models of Bargaining*, Berlin, Springer-Verlag.
Roth, A. (1993), "The early history of experimental economics," *Journal of History of Economic Thought*, 15.
Savage, L. J. (1954), *The Foundations of Statistics*, New York, John Wiley.
Schelling, T. C. (1960), *The Strategy of Conflict*, Cambridge, MA, Harvard University Press.
Schmidt, C. (1990), "Game theory and economics: a historical survey," *Revue d'Economie Politique*, 5.
Schmidt, C. (1992), "Concurrence, concours des producteurs et modes d'organisation de la production chez Antoine Augustin Cournot," *Economie et Société*, série Oeconomia, 16, 3.
Schmidt, C. (1993), "L'homo bellicus et la coordination économique," in *Revue Économique*, 46, 3.
Schmidt, C. (2001), "From the 'standards of behaviour' to the 'theory of social situations.' A contribution of game theory to the understanding of institutions," in

Knowledge, Social Institutions and the Division of Labour, Porta, P. L., Scazzieri, R. and Skinner, A., eds, Cheltenham, E. Elgar.
Schotter, A. (1992), "Oskar Morgenstern's contribution to the development of a theory of games," in *Toward a History of Game Theory*, Weintraub, E. R., ed., Durham, NC, Duke University Press, 95–112.
Selten, R. (1992), "Other impressions from the early years," in *Toward a History of Game Theory*, Weintraub, E. R., ed., Durham, NC, Duke University Press, 258–60.
Shubik, M. (1992), "Game theory at Princeton, 1949–1955: a personal reminiscence," in *Toward a History of Game Theory*, Weintraub, E. R., ed., Durham, NC, Duke University Press, 151–63.
Van Heijenoort, J. (ed.) (1976), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Cambridge, MA, Harvard University Press, 7th edn.
Von Neumann, J. (1923), "Zur Einführung der transfiniten Zahlen," *Acta Litterarum ac scientificarum Regiae Universitatis Hungaricae*, 1.
Von Neumann, J. (1925) (English translation, 1967), "Eine Axiomatisierung der Mengenlehre," *Journal für die reine und angewandte Mathematik*, 154.
Von Neumann, J. (1927), "Zur Hiberschen Beweistheorie," *Mathematische Zeitschrift*, 26.
Von Neumann, J. (1928a), "Die Axiomatisierung der Mengenlehre," *Mathematische Zeitschrift*, 27.
Von Neumann, J. (1928b), "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100, translated under the title "On the theory of games of strategy," in *Contributions to the Theory of Games*, Tucker, A. W. and Luce, R. D., eds, Princeton, NJ, Princeton University Press, 1959, vol. 4.
Von Neumann, J. (1929), "Über eine Widerspruchsfreiheitsfrage in der axiomatischen Mengenlehre," *Journal für die reine und angewandte Mathematik*, 160.
Von Neumann, J. (1932), *Mathematische Grundlagen der Quantummechanik*, Berlin, Springer-Verlag.
Von Neumann, J. (1961–3), *Collected Works*, 6 vols, New York, Pergamon.
Von Neumann, J. and Morgenstern, O. (1944) (1953), *Theory of Games and Economic Behavior*, Princeton, NJ, Princeton University Press.

Part II

Theoretical content

3 Bluff and reputation

Sylvain Sorin

Introduction

The theory of games is basically concerned with strategic behavioral interactions as opposed to individual maximization, typically found in decision theory:

> Thus each participant attempts to maximize a function (his above-mentioned "result") of which he does not control all variables. This is certainly no maximum problem, but a peculiar and disconcerting mixture of several conflicting maximum problems. Every participant is guided by another principle and neither determines all variables which affect his interest.
>
> (Von Neumann and Morgenstern 1944: 11, from now on, quoted as *TGEB*)

To analyze a situation in terms of a game, it is crucial to determine exactly the strategy spaces of the agents and, in particular, to specify the information on which their actions are based.

We will be mainly concerned in this chapter with some aspects of the strategic use of information: chance moves and mixed strategies, poker and bluff, reputation and cooperation in perturbed games, signals and anticipations in repeated games, etc.

We stop far short of exhausting the topic "strategy and information." In particular, we deliberately omit the links between information and selection of equilibria (i.e., signaling games used as a test for equilibrium selection concepts, see Kreps and Sobel 1994), the field of cognitive and epistemic foundations of the different concepts related to knowledge and logical interactive deduction (mutual knowledge, universal belief space, rationalizability, backwards induction), as well as the procedures used to extend a game (communication mechanisms, cheap talk, etc.).

Chance, information, and mixed strategies

The presentation by de Possel (1936) differentiates between games of pure chance, games of pure deduction, and games of pure cunning. In the first category the strategic aspect is missing and they do not really belong to the theory of games but rather to decision and probability theory (in the spirit of Pascal's early works).

The second class corresponds, for example, to finite games in extensive form with perfect information, where Zermelo's theorem (or Kuhn's extension) applies. Each player has a pure optimal strategy: given publicly known information, it induces deterministic behavior. The fact that the opponents know it or not is irrelevant, the difficulty lies in finding it or computing it explicitly.

The last category covers strategic games in normal form, where the players choose their strategies simultaneously. A typical example is the two-person zero-sum game, "leaf, scissors, rock," described by the following matrix of player 1's payoff:

	L	*S*	*R*
L	0	–1	1
S	1	0	–1
R	–1	1	0

It is clear that any deterministic way of playing, if announced in advance, can only guarantee –1 (if a player uses a pure strategy and this one is known, then his choice is predictable), while one could get 1 by guessing the other's choice.

The formal introduction of mixed strategies is due to Borel (1921).[1] Borel considers two-person, zero-sum and symmetric games where there is no pure optimal strategy, as in the example above. He then shows the advantage one has in varying the way of playing (its "code" in French) and adds:

> If one wants to formulate a precise rule for varying the play, with only features of the game entering the rule, and not psychological observations on the player to whom one is opposed, that rule will necessarily be equivalent to a statement such as the following. The probability that, at a given moment of play, A adopts the code C_k to determine his conduct at that moment is p_k.
>
> (Borel 1921: 1305, 1953: 98)

The first observation of this fact seems due to Waldegrace, who "solves," in mixed strategies, a zero-sum game, in a letter quoted by Montmort (1713: 409–12).[2]

The payoffs in these games corresponding basically to probabilities of winning (hence the expression "Jeux où le gain dépend à la fois du hasard et de l'habileté des joueurs" (Games where the payoffs depend both on chance and on the skill of the players) (Borel 1921: 1304), are already an expectation of payoffs; hence the use of mixed strategies and the corresponding definition of the mixed extension of the game with expected payoff does not introduce any new conceptual problem. One should observe that the axiomatics of expected utility is not presented in the original version of *TGEB*, the appendix "The axiomatic treatment of utility" is published only in the third edition (1953) of the book.

Mixed strategies appear as a way to introduce randomness, the mechanism being such that the player himself may not necessarly know which action he is in fact using. One can think of an order sent to some agent and one should distinguish between strategic choice and effective realization. In fact this interpretation can be found in Von Neumann (1928b: 28).

A statistical approach that considers mixed strategies as frequencies, to justify the computation of the payoff as an expectation, conflicts with the fact that the game itself is not repeated, otherwise the strategy sets also would change (*TGEB*: 146–7).

Two main points are then established in *TGEB*: first the fact that the potential danger mentioned above – that your opponent guesses your strategy – is not a danger (this is a consequence of the minimax theorem), and then the fact that this danger is an unavoidable consequence of a complete theory:

> Let us now imagine that there exists a complete theory of the zero-sum two-person game which tells a player what to do, and which is absolutely convincing. If the players knew such a theory then each player would have to assume that his strategy has been "found out" by his opponent.
>
> (*TGEB*: 148)

It is interesting to note that it is precisely this aspect that induces Borel to invoke "la nécéssité de la psychologie" (the necessity for psychology) (1938: 117), when the computation of an optimal strategy appears to be too complicated.[3]

From this point of view, having as a consequence the minimax theorem, the introduction of mixed strategies corresponds to the elimination of cunning in finite two-person zero-sum games (de Possel 1936: 119–20): to announce one's strategy is not risky, to know the strategy of your opponent is not beneficial.

Note that, from a mathematical point of view, the fact of taking incertainty into account translates on one hand by the convexity of the set of mixed strategies (that remains compact) and on the other hand by the linearity of the extended payoff function (that remains continuous).

Hence, there remain as games of cunning, the games without a value, i.e.

where the maximin and minimax differ: the order in which strategies are "announced" do matter.

On the other hand, in extensive form games with an initial random move, pure strategies can induce on the terminal nodes the same distribution as a mixed strategy. If, as remarked by Von Neumann (1928b: 26) – when chance is absent from the game or even when chance has been eliminated (by taking e.g. expectation in payoffs), it will reappear in the strategies – a kind of converse is also true.

In fact, let us consider the case where one player obtains some private information described by a random variable with a non-atomic distribution (Bellman and Blackwell 1949); if his signal is x, uniform on [0,1] and he has to play "top" or "bottom" with probability (1/3, 2/3), he can use a pure strategy like: "top" if $0 \leq x \leq 1/3$ and "bottom" otherwise. Formally a probability Q on a finite set A can always be realized through an application f from a probability space Ω with a non-atomic probability P, to A, with $P\{x; f(x) = a\} = Q(a)$.

The first representation corresponds to a mixed strategy with law Q while the second describes a pure strategy f, that induces the move $f(x)$, if the signal is x. A mixed strategy thus appears either as a random choice of actions given some public information, or as a deterministic choice depending upon private signals.

For more general games, the difficulties in order to get simple representations of strategies are due to the correlation between the different players' information structures and the dependence of the payoffs on the signals.[4] An especially interesting model, where the uncertainty concerns the payoffs, has been introduced and studied by Harsanyi (1978): the appearance of mixed strategies for a player reflects the uncertainty, even small, of his opponents upon his own payoff function, that he himself knows.

Bluff and revelation

A second justification for the use of mixed strategies is related to the notion of bluff. In particular its application to the game of poker appears early in the writings of Von Neumann (1928b) and Borel (1938: 61):

> What one has to remember is that in games where a more or less free and more or less secret choice of a player occurs (bid, selection, etc.), every way of playing which is too rigid has the inconvenience of giving information to the opponent, who would then know it, which would give him the possibility of a profitable reply . . . hence, if one tried to fool the opponent, one could also fool the partner.
>
> (Borel 1938: 116–17)

One has to distinguish here two aspects corresponding to the kind of "information" transmitted to the opponent:

1 If one speaks about the move one is going to play in a simultaneous game, uncertainty is useful since, if the move was predictable, the opponent could take advantage of it; it is this unpredictability, thus a lack of correlation, that may bother the partner.
2 Another possibility corresponds to the private information transmitted through a move; here again, what will be good against an adversary – the opaqueness of the signal – will be a nuisance in a cooperative framework.

The first aspect was dealt with in the previous section and we will now consider the second.

An adequate context is that of games with incomplete information (Harsanyi: 1967–8), where the agents have private information on a parameter that modifies their payoffs. The example par excellence is poker that is deeply studied by Borel (1938: Ch. 5), then by Von Neumann and Morgenstern in *TGEB*, and also by Gillies, Mayberry, and Von Neumann (1953). Its sequential structure makes the recursive analysis easier, as in all games with "almost perfect information" (Ponssard 1975).

Let us consider a simple case where the deal can be either good (g) or bad (b) for each of the players (with a publicly known initial probability). Player 1 then bids high (H) or small (S) amount and the opponent either folds or bids and sees. The simple strategy of player 1 that consists of making an announcement according to his state (H if g, and S if b) is revealing and hence wrong. The advantage of a mixed strategy is to make this revelation fuzzy and to allow for the following events to occur:

(i) In case b, to announce H, so that the opponent will pass (bluff).

Notice that an alternative interpretation is proposed in *TGEB*: 188: if, in case b, player 1 always announces S, then the bid H would signify that player 1 has a good hand, hence player 2 will not ask to see in this case. Thus it is also to incite player 2 to see after the bid H, if the hand was g, that the strategy corresponding to (i) is used.

(ii) In case g, to bid S, in order that the opponent will bid and see.

A pure strategy that would simply follow (i) and (ii) would also be completely revealing (*TGEB* uses the term *inverted signaling*: 54). The interest of the game thus lies in the determination of the optimal quantities $Prob(H|b)$ and $Prob(H|m)$.[5] For a general study of the use of mixed strategies to reveal partially some information, see the analysis of the "splitting lemma" in Aumann and Maschler (1995).

One should remark here (cf. *TGEB*: 189) that (i) and (ii) do not have the same status. As a matter of fact, if the event expected by the use of such a strategy occurs, hence if the opponent follows the anticipated behavior, the initial private information will be revealed in case (ii) but not in case (i). As a

consequence, if this strategy (ii) is played, it will be discovered and player 2 will adapt his behavior to it; on the contrary the strategy (i) remains secret and an answer to this possibility of bluffing will be for player 2 to bluff also, namely, to bid and see after a signal H even with a bad hand (*TGEB*: 219).

One can already observe in this classroom example all the fundamental concepts for the strategic analysis of information: revealing strategies, conditional probabilities, observability of the signals, credibility of the messages . . . One has to add, for the cooperative aspect in the case of non-zero-sum games, the problems related to the transmission of information and coordination (cf. *TGEB*: 53–4).[6]

Beliefs and cooperation

The study of questions dealing with reputation effects was largely developed in response to two now well-known paradoxes: the "prisoner's dilemma" (Luce and Raiffa 1957: 94) and the "chain store paradox" (Selten 1978). In both cases the paradoxical aspect is "solved" by introducing uncertainty, but the problematics differ and we will study the one related to the second case in the next section.

The prisoner's dilemma is a two-person game that can be represented by the following matrix:

	C	*D*
C	3, 3	0, 4
D	4, 0	1, 1

This game has a single equilibrium outcome in any finite repetition, the sequence of payoffs (1, 1). This means that, repetition does not lead to a Pareto outcome, such as a sequence of (3, 3)s; and moreover it does not allow for learning phenomena (the players could realize that it is better for them both to play (*C*, *C*)), or for signaling (like playing *C* to announce his willingness to cooperate in the future).

To solve this paradox, Kreps *et al.* (1982) consider the situation where there exists a positive probability ε, that one of the players, at least, will be forced to play the "tit for tat" strategy (TfT): that is, to play *C* at stage one, and from then on to reproduce, at each stage, the previous action of the opponent. They show that, in long games (where the number of stages is greater than some bound $N(\varepsilon)$), not only equilibrium payoffs different from (1, 1) are possible – in particular the cooperative outcome (3, 3) can be approached –[7] but that any sequential equilibrium payoff (as introduced by Kreps and Wilson, 1982a) will be near this outcome.

In short, this type of uncertainty on the behavior of one of the players, small though it is, will be enough to enforce cooperation. The reputation

phenomenon grows from some initial perturbation but works as soon as the latter is present, whatever its size. In term of optimal strategies, everything proceeds as if the probability to face a cooperative opponent, meaning one using TfT, increases to one with the repetition, as soon as it is positive at the start.

The proof reveals the following properties:

1 If the strategy of the perturbed player differs from TfT, he can nevertheless mimic TfT without his opponent being able to detect any deviation.
2 As long as TfT has been played, it is better to keep on playing it (at least as long as one is far from the end of the game).

The first property is a fundamental datum of perturbed games and related reputation phenomena. The presence of an initial uncertainty conveys a qualitative change in the interpretation of the deviations. The second aspect is specific to this strategy and to this game: the use of TfT by one player implies that the payoffs of both players will be close to each other, uniformly with respect to the length of the game and, moreover, a best reply to TfT is almost always *C*.

However, if the game is perturbed by another class of strategies, the result changes drastically and cooperation is no longer ensured. In fact a result of Fudenberg and Maskin (1986) shows that any feasible and individually rational payoff can be obtained as an equilibrium payoff of the game perturbed in a way adapted to this payoff and repeated a large number of times.

Similarly, a condition like "perfection in subgames" is necessary for the result to hold. The following equilibrium strategies, due to V. Krishna, induce an average payoff near (2, 2): the non-perturbed player plays *D* during the first half of the game then *C* during the second half while the perturbed player plays like TfT and punishes (for ever) if his opponent plays *C* too early. This equilibrium is not sequential since the perturbed player would be better off by continuing to play TfT after a deviation.

Another model where repetition, uncertainty and bounded rationality induce cooperation has been proposed by Aumann and Sorin (1989). The framework is a two-person game where there exists a non-ambiguous reason to cooperate: the set of feasible payoffs has only one Pareto point (in the strong sense). These are games with "common interest," as in the following example:

	C	*D*
C	9, 9	0, 8
D	8, 0	7, 7

Assume that one cannot exclude the possibility that one of the players uses a strategy with bounded recall of size *N* (namely depending only upon the

last N moves of his opponent), then in long games all payoffs corresponding to pure strategy equilibria will be close to the Pareto outcome.

Here the introduction of some uncertainty, independent of the game itself, suffices for the equilibrium mechanism to induce a cooperative outcome. The possibility to build a reputation is there, present among other possibilities, but it is the rational behavior of the agents itself that chooses to mimic the perturbed strategy, which is the best for all.

To get the result, one uses on one hand the property that strategies with bounded recall can be identified in finite time, and on the other the fact that the maximal payoff that one can achieve against them is independent of the past; this allows one to describe the optimal behavior after a (potential) deviation. One then shows that strategies that do not induce a cooperative outcome are dominated by others that mimic strategies with bounded recall.

The existence of a pure equilibria is obtained through an explicit construction. If one deals with mixed strategies, the revelation mechanism is slower and can force the players to spend a fraction of the duration of the game trying to gather information, and this eventually leads to a loss of optimality.

Reputation, signals, and conjectures

The model of the "chain store" corresponds to a sequential game of entry, where player 1 faces a finite sequence of successive potential entrants, all having the same type. Following an entry, the behavior of player 1 (active or passive) can be observed by all potential entrants, and obviously not in case of no entry. At each stage, the game consists of a tree with perfect information where a player 2 (entrant) plays first and player 1 then plays in case of entry.

Let us denote by a, a', a'' the payoffs of player 1 in case of status quo, entry/passive, entry/active, repectively and similarly b, b', b'' for player 2. The paradox is due to the fact that if the profile "entry/passive" is the only sub-game perfect equilibria (namely $a' > a''$ and $b' > b$) and if the outcome "out" is better for player 1 ($a > a'$), the latter could, by being active during the first rounds, obtain a reputation that will prevent subsequent entries as soon as $b'' < b$.

The analysis in terms of backwards induction shows that the behavior of player 1 vis-à-vis player m does not influence the behavior of player $m + 1$. The only sub-game perfect equilibria of the game with any finite number of entrants corresponds to a sequence of profiles "entry/passive" and reputation does not appear in this model.

As a matter of fact, reputation grows out of uncertainty and this is the basis of the quite similar models of Kreps and Wilson (1982a) and Milgrom and Roberts (1982). As soon as there exists a positive probability p that player 1 is perturbed by having the behavior "always active," a sequential equilibrium appears in the game with n entrants, when n is large enough, with the property that these potential entrants do not enter, except for at most N of

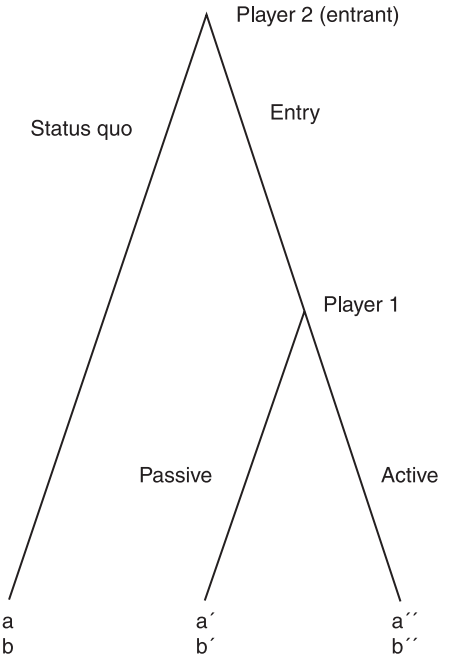

Figure 3.1

them, where this bound depends only on p and not on n. It is enough for player 1 to play at the beginning of the game like a perturbed type to convince his opponents that he is in fact of this type, or at least with a probability high enough. The change of behavior of the following entrants, who, being rational, prefer the status quo, implies that player 1 does not have to offer further proof that he is of this type – hence the equilibrium.

Govindan (1995) recently showed that, by considering stable equilibria (in the sense of Kohlberg and Mertens 1986) one can get rid of the other sequential equilibria, corresponding to "out of equilibrium path" beliefs that are much less plausible (the probability of the perturbed type decreasing after a history "entry/active").

Analogous results relying on similar properties (need for uncertainty, selection through stability criteria) have been obtained by Demange (1992) in the framework of escalation games.

Finally, note that a parallel study of entry problems where a "big" player faces a family of "small" potential entrants gives different results (Fudenberg and Kreps 1987), because the evolution, with the number of agents, of the ratio "immediate loss due to active behavior"/"future gain obtained by reputation" follows another logic.

These kinds of reputation properties have been described by Fudenberg and Levine (1989, 1992) in a more general framework. Consider a repeated strategic form game with complete information where a "long" player,

who stays during all the play, faces a sequence of "short" players, who play successively only once, each one being informed of the past history, which is the sequence of moves used up to now. The behavior of the short player n is a function of his beliefs concerning the behavior of the long player at stage n. If the uncertainty upon the strategy of player 1 is such that, either he has a rational behavior, or he is "programmed" to play in a stationary way a specific mixed strategy of the one-shot game, one obtains a lower bound on the amount player 1 can obtain by building a reputation.

More precisely, let I and J denote the pure strategy sets, and F and G be the one shot payoff function of long and short players (respectively). $\Delta(I)$, resp. $\Delta(J)$, are the mixed strategy sets of the one shot game. B will denote the best reply correspondence of the short player, defined on $\Delta(I)$ with values in $\Delta(J)$. Let σ be an equilibrium strategy of the long player and $\tilde{\sigma}$ the corresponding perturbed strategy. Then, under $\tilde{\sigma}$, the probability that player 1 will play like x, element of $\Delta(I)$, at stage n, if he played like x up to now, goes to one as n goes to infinity: this is the "merging" property which is crucial in this literature. As a consequence, any optimal strategy of a short player n, for n large enough, will be near a best reply to x (using the continuity in the payoffs). Since the long player can always mimick the perturbed type that plays x i.i.d., one achieves as a lower bound for the asymptotic equilibrium payoffs of the long player the quantity:

$$w = \sup_{x \in \Delta(I)} \inf_{y \in B(x)} F(x, y).$$

If the initial game is a game in extensive form and the information after each stage corresponds to the terminal node reached, the result may fail, as in the example shown in Figure 3.2.

By playing a, the long player tries to convince the short player that he is going to keep on playing a, thus that the latter is better off by playing B. However if the initial belief that the long player will play b is greater than 1/2, the first short player will play A; now, by lack of new information, the belief that the short player 2 will have on b will not change, and his behavior will be like that of the first short player and so on . . . The reputation effect does not hold.

This is precisely the phenomenon that occurs in the "chain store" example where the long player will not be checked again, from some stage on, but in this case, it is exactly what he is looking for.

To obtain a lower bound on the payoffs in this new and more general framework one has to introduce a signaling function L defined on $I \times J$, that describes the information of the short player after each stage. One then obtains (Fudenberg and Levine 1992):

$$w^* = \sup_{x \in \Delta(I)} \inf \{F(z, y);\ z \in \Delta(I),\ y \in B(z),\ L(x, y) = L(z, y)\}.$$

In fact, when the long player plays x, everything appears, in terms of informa-

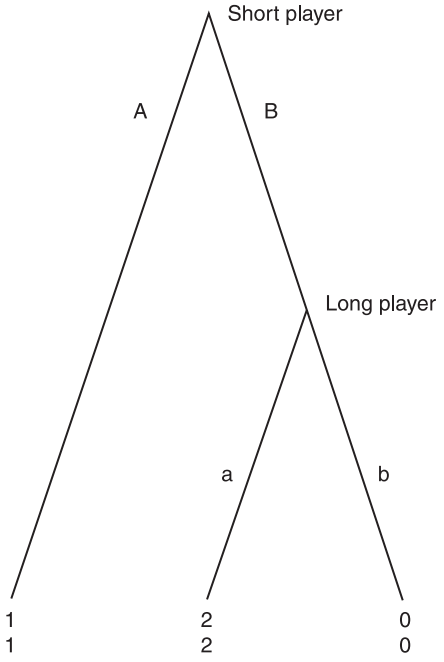

Figure 3.2

tion of player 2, as he was facing a long player playing z – and at the same time, behaving optimally. In the example below, where the left matrix represents the signals of the short player and the matrix on the right the payoffs of both players, the payoff (2, 2), corresponding to the moves (a, A) cannot be obtained through reputation if b is initially more expected than a; player 1 cannot do better than playing c and getting 1.

	A	*B*
a	α	α
b	α	α
c	α'	α''

	A	*B*
a	2, 2	0, 0
b	0, 0	0, 2
c	1, 1	0, 0

One recovers here concepts related to the notion of conjectural equilibrium (Hahn 1978; Battigalli *et al.* 1992). The latter consists, in a strategic form game with signaling function, of a profile of strategies and for each agent, of conjectures on the behavior of others such that:

1 each agent's own strategy is a best reply to his conjectures
2 the signal he is getting from the profile of strategies corresponds to the one he anticipates, given his conjectures.

It is clear that, in this case, a kind of stability is reached, in the sense that, if such a profile is played, there will be no revision of the agents' beliefs, hence the strategies will not change. It is thus conceivable that learning procedures, based on beliefs' evolution in a bayesian optimization process converge to conjectural equilibria, and not only to Nash equilibria (Fudenberg and Levine 1993; Kalai and Lehrer 1993).

Let us consider the following game:

	G	*D*
H	1, 1	1, x
B	x, 1	0, 0

(H, G) which is not an equilibrium for $x > 1$, will be a conjectural equilibrium, if each player thinks that his opponent wants to reach his best equilibrium payoff and only the individual payoffs are privately announced.

Note that this procedure excludes all kinds of experimentation, where at each stage, each player would play a completely mixed perturbation; this would in this case destroy the wrong conjectures.

Let us come back to reputation models but where now two long players compete, the uncertainty being on the behavior of the first. The main difference with the previous study is that the rational behavior of player 2 does not translate in a stage after stage maximization. He has to anticipate the consequences, in terms of player 1's future behavior, of his present choice. In the case of an undiscounted game, with standard signaling (the moves are announced), Cripps and Thomas (1995) have shown that player 1 cannot do better than to force player 2 to have a minimal rational behavior, namely such that his payoff is individually rational. The lower bound on player 1's payoff is in this case:

$$\underline{w} = \sup_{x \in \Delta(I)} \inf \{F(x, y);\, y \in \Delta(J),\, G(x, y) \geq \min_{s \in \Delta(I)} \max_{t \in \Delta(J)} G(s, t)\}.$$

A similar result is obtained at the limit for the discounted case by Cripps, Schmidt, and Thomas (1996), following Schmidt (1993).

It is interesting to notice that the same bound is obtained using a completely different approach, in the framework of repeated games with lack of information on one side (Aumann and Maschler 1995), where each player knows his own payoff. The question is to determine which kind of uncertainty would be most favorable to the informed player, given his type. One can show (see a description and references in Forges 1992: 165–8) that the maximum is reached when the payoff of the perturbation of player 1 is $-G$, hence when his opponent is in a situation of pure conflict; this maximum is then the previous $\underline{w}$. Here again, this quantity is independent of the prob-

ability of the perturbation as soon as this one is positive. More specifically, in the coordination game "battle of the sexes" (Luce and Raiffa 1957: 90) perturbed in an optimal way, one has:

	G	D
H	2	0
B	0	1

Player 1's payoff (rational)

	G	D
H	–1	0
B	0	–2

Player 1's payoff (perturbed)

	G	D
H	1	0
B	0	2

Player 2's payoff

The minimum equilibrium payoff of player 1 in the perturbed game becomes 4/3, while his individual rational level is 2/3.

The intermediary case, where a long player 1 faces a sequence of players 2, each of them playing T stages, or having an expected life length much smaller than the long player, has been studied by Celentani *et al.* (1996). The analysis depends upon the information that the "middle" players have on the strategy of the long player. The most favorable situation for the latter is when his past strategy is revealed, which amounts to considering the T stage game in normal form and assuming standard signaling. Since the long player wants to monitor the behavior of the middle players, his message has to be as explicit as possible, hence he does not want at all to hide his behavior: in opposition with what occurs for the bluff phenomena, here the long player tries to reduce as much as possible the uncertainty of the other players on himself. Indeed, as T goes to infinity, the long player can build a strategy such that a best reply of the middle players will give him the best payoff compatible with their individual rationality constraints.

The same result can be obtained under the following conditions: the middle players are not aware of the strategy of the long player in the game of length T, but learn at each stage revealing information on his moves. In addition, the long player receives at each stage a random signal, function of the move of player 2, but with constant support. It is then clear that all histories of length T will appear with positive probability, hence if player 1 plays, stationary, a strategy in the T stage game, this one will be identified in the long run.

Related questions and extensions

Recall that information problems could concern the strategy, either to predict the next move, or knowing the move, to get more precision on the initial private knowledge. In the second case, taking into consideration the anticipated behavior of his opponent, one player can manipulate his information.

Another approach corresponds to a situation where there is no intial information, but uncertainty concerning the future. The use of a move thus

appears as a signal on the future behavior: in fact it may be compatible only with a subclass of strategies for the subgame to follow, otherwise it would lead to an overall dominated strategy. If the message transmitted this way is explicit enough – in the sense that there is no ambiguity on the compatible strategies – it can be used in the framework of a logical analysis of forward induction to select among equilibria by eliminating unjustifiable paths (see Osborne 1990; Ponssard 1990; van Damme 1989).

In the same vein, a series of articles shows the advantage for a player of reducing his set of potential strategies (automata, sacrifice, penalty), using a similar framework. The analysis depends crucially on the fact that the opponent is aware of this reduction, on the possibility of communication through the moves and on the injective meaning of the signals.

Finally, one should insist on the fact the strategic interaction is fundamentally different in the case of a game than in the case of individual decision making. This is well known, at the behavior level, due to classical strategic game theory, but this also holds at the information level. Indeed, Milgrom and Roberts (1982) already remarked that it is not necessary for a perturbation to be common knowledge for the reputation effect to appear, the fact that there is no common knowledge of the true situation is enough.

This observation leads to a study of questions related to propagation of uncertainty where "domino effects" occur: the lack of public knowledge on the moves leads each player to take into consideration a whole hierachy of situations and decisions including a similar behavior on the part of his opponent. A typical example of the kind of discontinuities that appear may be found in Rubinstein (1989) and an approach to the quantitative evaluation of the propagation of information in Sorin (1998).

Notes

1 "E. Borel was the first author to evolve the concept of a strategy, pure as well as mixed," Von Neumann (1953: 124).
2 See also the comments in Guilbaud (1961), as well as the preface and the translation by Kuhn in Baumol and Goldfeld (1968: 3–9).
3 Borel shows also that it is impossible to mimic randomness, for example to have a behavior, depending on the past, such that your opponent would be facing a sequence of i.i.d. random variables (1938: 119–20).
4 For very general extensions of this procedure of representation of mixed strategies, and sufficient conditions for exact or ε-purification, see Aumann *et al.* (1983) and the references therein.
5 One can find an explicit procedure to construct super-optimal strategies (i.e. strategies that take advantage of the errors of the opponent, cf. *TGEB*: 205–6) in finite games with almost perfect information and incomplete information on both sides, in Ponssard and Sorin (1982).
6 Note also, on a technical level, that Gillies, Mayberry, and Von Neumann (1953: 30) study the convergence of the model with dicrete information to the continuous case, in terms of "distributional strategies."
7 As in the so-called Folk theorem that states that the equilibrium payoffs of an

infinitely repeated (undiscounted) game are the feasible and individually rational payoffs of the one-stage game.

References

Aumann, R. J., Y. Katznelson, R. Radner, R. W. Rosenthal, and B. Weiss (1983), "Approximate purification of mixed strategies," *Mathematics of Operations Research*, 8, 327–41.

Aumann, R. J. and M. Maschler (1995), *Repeated Games with Incomplete Information,* MIT Press.

Aumann, R. J. and S. Sorin (1989), "Cooperation and bounded recall," *Games and Economic Behavior*, 1, 5–39.

Battigalli, P., M. Gilli, and M. C. Molinari (1992), "Learning and convergence to equilibrium in repeated strategic interactions: an introductory survey," *Ricerche Economiche*, 46, 335–77.

Baumol, W. J. and S. M. Goldfeld (1968), *Precursors in Mathematical Economics: An Anthology*, London School of Economics and Political Science.

Bellman, R. and D. Blackwell (1949), "Some two-person games involving bluffing," *Proc. NAS*, 35, 600–5.

Borel, E. (1921), "La théorie du jeu et les équations intégrales à noyau symétrique gauche," *Comptes Rendus de l'Académie des Sciences*, 173, 1304–8; translated in *Econometrica*, 21, 97–100, 1953.

Borel, E. (1924), "Eléments de la théorie des probabilités, Note IV: Sur les jeux où interviennent le hasard et l'habileté des joueurs," Hermann, 204–24; translated in *Econometrica*, 21, 101–15, 1953.

Borel, E. (1938), *Traité du calcul des probabilités et de ses applications, Tome IV, Fascicule II, Applications aux jeux de hasard*, ed. J. Ville, Gauthier-Villars.

Celentani M., D. Fudenberg, D. Levine, and W. Pesendorfer (1996), "Maintaining a reputation against a long lived opponent," *Econometrica*, 64, 691–704.

Cripps, M., K. Schmidt, and J. Thomas (1996), "Reputation in perturbed repeated games," discussion paper, *Journal of Economic Theory*, 69, 387–410.

Cripps, M. and J. Thomas (1995), "Reputation and commitment in two-person repeated games without discounting," *Econometrica*, 63, 1401–19.

Damme, E. van (1989), "Stable equilibria and forward induction," *Journal of Economic Theory*, 48, 476–96.

Demange, G. (1992), "Rational escalation," *Annales d'Economie et de Statistique*, 25–6, 227–49.

Forges F. (1992), "Repeated games of incomplete information: non-zero-sum," in *Handbook of Game Theory*, vol. I, R. J. Aumann and S. Hart, eds, North Holland, 155–77.

Fudenberg, D. (1992), "Explaining cooperation and commitment in repeated games," in *Advances in Economic Theory: Sixth World Congress*, edited by J.-J. Laffont, Cambridge University Press, 89–131.

Fudenberg, D. and D. M. Kreps (1987), "Reputation in the simultaneous play of multiple opponents," *Review of Economic Studies*, 54, 541–68.

Fudenberg, D. and D. Levine (1989), "Reputation and equilibrium selection in games with a patient player," *Econometrica*, 57, 759–78.

Fudenberg, D. and D. Levine (1992), "Maintaining a reputation when strategies are imperfectly observed," *Review of Economic Studies*, 59, 561–79.

Fudenberg, D. and D. Levine (1993), "Self-confirming equilibrium," *Econometrica*, 61, 523–45.
Fudenberg, D. and E. Maskin (1986), "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, 54, 533–54.
Gillies, D. B., J. P. Mayberry and J. Von Neumann (1953), "Two variants of poker," in *Contributions to the Theory of Games, II*, H. W. Kuhn and A. W. Tucker, eds, Princeton University Press, 13–50.
Govindan, S. (1995), "Stability and the chain store paradox," *Journal of Economic Theory*, 66, 536–47.
Guilbaud, G. Th. (1961), "Faut-il jouer au plus fin? (Notes sur l'histoire de la théorie des jeux)," in *La Décision, Colloque international du CNRS, Paris 1960*, Editions du CNRS, 171–82.
Hahn, F. (1978), "On Non-Walrasian equilibria," *Review of Economic Studies*, 45, 1–17.
Harsanyi, J. C. (1967–8), "Games with incomplete information played by 'Bayesian players,' parts I–III," *Management Science*, 14, 159–82, 320–34, 486–502.
Harsanyi, J. C. (1978), "Games with randomly distributed payoffs: a new rationale for mixed strategy equilibrium points," *International Journal of Game Theory*, 2, 1–23.
Kalai, E. and E. Lehrer (1993), "Subjective equilibrium in repeated games," *Econometrica*, 61, 1231–40.
Kohlberg, E. and J.-F. Mertens (1986), "On the strategic stability of equilibria," *Econometrica*, 54, 1003–38.
Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982), "Rational cooperation in the finitely repeated Prisoner's Dilemma," *Journal of Economic Theory*, 27, 245–52.
Kreps, D. and J. Sobel (1994), "Signalling," in *Handbook of Game Theory*, vol. II, R. J. Aumann and S. Hart, eds, North Holland, 849–67.
Kreps, D. and R. Wilson (1982a), "Sequential equilibria," *Econometrica*, 50, 863–94.
Kreps, D. and R. Wilson (1982b), "Reputation and imperfect information," *Journal of Economic Theory*, 27, 253–79.
Luce, R. D. and H. Raiffa (1957), *Games and Decisions*, Wiley.
Milgrom, P. and J. Roberts (1982), "Predation, reputation and entry deterrence," *Journal of Economic Theory*, 27, 280–312.
Montmort, P. de (1713), *Essay d'analyse sur les jeux de hazard*, chez J. Quillau, Paris.
Osborne, M. (1990), "Signalling, forward induction and stability in finitely repeated games," *Journal of Economic Theory*, 50, 22–36.
Ponssard, J.-P. (1975), "Zero-sum games with 'almost' perfect information," *Management Science*, 21, 794–805.
Ponssard, J.-P. (1990), "Self-enforceable paths in extensive form games," *Theory and Decision*, 29, 69–83.
Ponssard, J.-P. and S. Sorin (1982), "Optimal behavioral strategies in 0-sum games with almost perfect information," *Mathematics of Operations Research*, 7, 14–31.
Possel, R. de (1936), "Sur la théorie mathématique des jeux de hasard et de réflexion," reproduced in H. Moulin, ed., *Fondation de la théorie des jeux*, Hermann, 83–120, 1979.
Rubinstein, A. (1989), "The electronic mail game: strategic behavior under 'almost common knowledge,'" *American Economic Review*, 79, 385–91.
Schmidt, K. (1993), "Reputation and equilibrium characterization in repeated games with conflicting interests," *Econometrica*, 61, 325–51.

Selten, R. (1975), "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4, 25–55.
Selten, R. (1978), "The chain store paradox," *Theory and Decision*, 9, 127–59.
Sorin, S. (1992), "Information and rationality: some comments," *Annales d'Economie et de Statistique*, 25–6, 315–25.
Sorin, S. (1998), "On the impact of an event," *International Journal of Game Theory*, 27, 315–30.
Von Neumann, J. (1928a), "Sur la théorie des jeux," *Comptes Rendus de l'Académie des Sciences*, 186, 1689–91.
Von Neumann, J. (1928b), "Zur Theorie der Gesellschaftspiele," *Mathematische Annalen*, 100, 295–320. (Translated in *Contributions to the Theory of Games, IV*, A. W. Tucker and R. D. Luce, eds, Princeton University Press, 13–42, 1959.)
Von Neumann, J. (1953), "Communication on the Borel notes," *Econometrica*, 21, 124–5.
Von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton University Press (quoted from the third edition, 1953).
Watson, J. (1993), "A 'reputation' refinement without equilibrium," *Econometrica*, 61, 199–205.
Watson, J. (1994), "Cooperation in the infinitely repeated Prisoner's Dilemma with perturbations," *Games and Economic Behavior*, 7, 260–85.

4 An appraisal of cooperative game theory

Hervé Moulin

From coalition formation to distributive justice

The model of cooperative games has an interesting history and an ambiguous status within economic theory. For the first two decades after Von Neumann and Morgenstern's book (*TGEB*) (i.e. until the mid-1960s), the analysis of coalition formation and the search for stable agreements was the most active component of the emerging game theory. Subsequently the topic lost its prominence and research on cooperative games mainly followed two methodologically distinct paths: the first one, of a positive nature, explored the deep connections of core stability with the competitive equilibrium of exchange (or production) economies (the seminal work by Debreu and Scarf is dated 1963); the second one, of a normative nature, analyzed axiomatically a number of single-valued solutions such as the Shapley value and Nash's solution to the bargaining problem (although the seminal papers by Nash and Shapley are dated 1950 and 1953 respectively, systematic research on axiomatic solutions did not start until the late 1960s).

The dual normative/positive interpretation of the cooperative game model has often been obscured by formal researchers (some of the most important textbooks surveying game theory, written by some of the most prominent contributors to cooperative game theory, are conspicuously brief on the interpretation of the cooperative model: e.g. Owen 1968, Shubik 1982. This has undoubtedly slowed the progress of the cooperative game ideas in the economics profession at large, and even generated some degree of mistrust for cooperative game analysis. Yet I will argue below that this fuzziness is gradually disappearing as the concepts of cooperative game theory are applied more and more to economics and to other social sciences. I will also argue that the interplay of the two interpretations throws some light on the fundamental debate of normative economics opposing the liberal and welfarist doctrines. In this respect, cooperative game theory adds to our understanding of the universal tension between freedom and justice in at least two ways:

1 the exploration of coalitional stability demonstrates the logical limits of the "efficiency postulate," one of the pillars of the liberal doctrine for

minimal central regulation of social interactions. Specifically the lack of core stability in certain problems of production and allocation of private goods, as well as in the voting context, provides a powerful argument for some form of regulation, some limits imposed by the collectivity on the freedom of association;

2 the axiomatic analysis of "values" and "bargaining solutions" formalizes several fundamental ideas of distributive justice (such as sharing a surplus according to marginal contributions) hence forges important tools for the implementation of welfarist principles of justice at the micro-level.

We introduce the efficiency postulate, on which the cooperative game mode rests. We then define the core, the central stability concept. Next, we summarize the abundant research on the conditions guaranteeing that the core is nonempty. Then we compare several interpretations of these cooperative games where the core is empty and discuss the axiomatic approach to cooperative games and bargaining in general before concluding.

Agreements and the efficiency postulate

Von Neumann and Morgenstern introduced the model of a cooperative game as a poor second best of what strategic analysis ought to be. They make this crucial point after analyzing a normal form version of the divide the dollar by majority game. Each one of the three players independently names one of the other two players: if two players "agree" (in the sense that they name each other) they each receive 50 c and the third player gets nothing; no payments are made if no two players agree. Here is how Von Neumann and Morgenstern analyze the game:

> To begin with, it is clear that there is absolutely nothing for a player to do in this game but to look for a partner, – i.e. for another player who is prepared to form a couple with him. The game is so simple and absolutely devoid of any other strategic possibilities that there just is no occasion for any other reasoned procedure. Since each player makes his personal move in ignorance of those of the others, no collaboration of the players can be established during the course of the play. Two players who wish to collaborate must get together on this subject before the play, – i.e. outside the game. The player who (in making his personal move) lives up to his agreement (by choosing the partner's number) must possess the conviction that the partner too will do likewise. As long as we are concerned only with the rules of the game, as stated above, we are in no position to judge what the basis for such a conviction may be. In other words what, if anything, enforces the "sanctity" of such agreements?
>
> (*TGEB*: 223)

The key observation is that the given game cannot be played independently by three players who do not communicate, who do not agree "outside the game" to coordinate their messages during the game. It is as if the essential decision (which coalition will form) is taken before the formal game of sending independent messages; furthermore this decision is the indivisible privilege of a (of any) coalition of two players: when we agree on a common course of action, there is no way to divide responsibility for the agreement between the players, anymore than we can divide responsibility for the sound of two hands clapping between the two hands. To Von Neumann and Morgenstern, such indivisible representation of preplay agreements is a parasitic assumption, superimposed on the given game out of our inability to analyze in more details the individual interactions leading to such or such agreement: this assumption should ultimately disappear.

> We are trying to establish a theory of the rational conduct of the participants in a given game. In our consideration of the simple majority game we have reached the point beyond which it is difficult to go on formulating such a theory without auxiliary concepts such as "agreements," "understandings," etc. On a later occasion we propose to investigate what theoretical structures are required in order to eliminate these concepts.
>
> (*TGEB*: 224)

The position is methodologically clear: game theory is a model of social interactions built upon the rational behavior of selfish individuals: there is no room in its foundations for the indivisible agreement of several individual players, or for anything akin to a social contract by which individual freedom is alienated to the authority of the collective will expressed in the contract: "Chacun de nous met en commun sa personne et toute sa puissance sous la suprême direction de la volonté générale: et nous recevons en corps chaque membre comme partie indivisible du tout" (Rousseau 1762).

The notion of indivisible agreement is not a primitive concept of the theory; it is an approximation reflecting our imperfect knowledge of the situation at hand. The same basic position will remain dominant through the work of Nash (see the discussion of the Nash program), all the way to recent assessments such as Aumann's: "Formally cooperative games may be considered a special case of non-cooperative games in the sense that one may build the negotiation and enforcement procedures explicitly into the extensive form of the game" (Aumann 1987).

In short, the cooperative game model has been viewed by its inventors and by most of the game theory profession as a "second class" model, one that does not rest on solid methodological foundations. I would argue, on the contrary, that cooperative game analysis relies on a different model of rationality, central to the liberal economic doctrine (e.g. Coase 1960 or Buchanan and Tullock 1967):

> In our analysis we have assumed that individuals are motivated by utility-maximizing considerations and that, when an opportunity for mutual gains exists, "trade" will take place. This assumption is one of the foundations on which economic theory is constructed.

I propose to call this the *efficiency postulate*. This model takes the indivisible agreements within a group of social actors as the basic ingredient of collective action and assumes that, within the set of physical outcomes that the group can jointly achieve, the outcome *a* ultimately selected is an efficient one (i.e. there is no other outcome *b* to which no one strictly prefers *a* and that someone strictly prefers to *a*: the postulate is analogous, at the group level, to the rationality postulate of utility maximization in the non-cooperative model. Both the noncooperative and the cooperative models belong in the realm of methodological individualism (the definition of efficiency – as well as Pareto ordering – rests solely on individual preferences), however in the former the decision power is privately distributed among the participants (each player controlling his/her own set of strategic choices; the determination of an outcome coincides with that of one strategy per player), whereas in the latter it is the indivisible public property of the group as a whole. Neither model can be reduced to the other although the relations between the two models are one of the most interesting topics of the theory (on which more below).

The two seminal questions of the theory are as follows: first, is the efficiency postulate compatible with freedom of association when the various coalitions of players have various opportunities to cooperate? (the question goes back to Von Neumann and Morgenstern). Second, when only the grand coalition has the opportunity to cooperate, what normative principles will help us select a particular compromise on the efficiency frontier? (this question was first raised by Nash in 1951: see below).

The core

The fundamental economic example of a cooperative game is the pure exchange economy à la Arrow-Debreu (e.g. Debreu 1959). Every agent owns certain resources (a certain bundle of private goods) and they freely engage in trade by pairs or in any other coalition (i.e. subgroup) of agents. The efficiency postulate asserts that every opportunity to trade will be exploited. The private ownership structure, however, greatly complicates the analysis because each coalition has its own opportunities to trade and the efficiency postulate must be applied to all coalitions. So an outcome is deemed stable according to the efficiency postulate (it is a core outcome) if (1) it results from a feasible trade benefiting (or at least not hurting) all agents, (2) it is an efficient overall trade, and (3) no coalition of agents (e.g. no pair, no triple and so on) can come up with a better trade on its own (that is, using its own resources).

In general, a cooperative game represents opportunities to trade by a certain amount of surplus (case of transferable utility) or by a set of utility vectors feasible through some undisclosed set of coordinated actions. The full specification of a game includes one such set of cooperative opportunities (one surplus quantity in the transferable utility case) per coalition of players (e.g. 255 such numbers if the game has 8 players), individual players are not deprived of strategic power in such a model, because they can let their potential coalition partners compete for their collaboration so as to gain a larger share of surplus) exactly as the monopolist holding the single copy of a commodity desired by several buyers can bid up the price until extracting most of the surplus from the potential buyers. The set of cooperative opportunities open to a specific coalition of players is derived from the (undisclosed) underlying property right over whatever resources are the object of cooperation (in the case of an exchange economy the private ownership of the initial endowments implies the right to trade those resources at will; in other contexts the property rights may be represented by a technology).

The core is the fundamental stability concept when all coalitions are free to form and automatically reach an efficient agreement (as assumed by the efficiency postulate). An outcome *a* is deemed stable if it is efficient (Pareto optimal) and if no coalition has a cooperative opportunity that makes all members of the coalition better off (or at least one better off and none worse off) than they were under outcome *a*. Thus agreeing on outcome a is compatible with the efficiency postulate of all potential coalitions.

As for any other equilibrium concept, the core concept raises two questions. For what collective decision problems is the core guaranteed to be non empty? If the core is empty, what becomes of the efficiency postulate: is cooperation by direct agreements unworkable? Historically the second question was addressed (by Von Neumann and Morgenstern) almost two decades before the first one, yet we will address them in the opposite (logically more plausible) order.

The divide the dollar game is the simplest example of a cooperative game with an empty core. The majority rule allows any two players out of three to grab the whole surplus. In general the core will be empty if all coalitions of size *s* are capable of generating more per capita surplus than the grand coalition including all *n* players. Sometimes, however, the emptiness of the core comes about in a more subtle fashion. Consider an exchange economy with two indivisible goods (two identical horses) and money, two sellers who each own one horse and two buyers who could each buy up to two horses. The sellers have a reservation price of 1 and 3 respectively and the buyers' willingness to pay are as follows:

buyer 1: 4 for one horse; 5 for two horses

buyer 2: 5 for one horse; 7 for two horses

Efficiency requires that each buyer gets one horse and the core stability property implies that they pay the same price p $3 \leq p \leq 4$, and that each seller receives p.[1] Conversely every such competitive allocation is in the core.

Next imagine that buyer 2's willingness to pay is increasing with the number of horses he acquires (with two horses he can use the carriage sitting in his garage).

buyer 2: 2 for one horse; 7 for two horses

The core of this exchange economy is empty: indeed total surplus in the economy is 3. Seller 1 and buyer 1 extract 3 units of surplus by their own trade therefore the core property implies that seller 2 and buyer 2 get no share of surplus. On the other hand sellers 1 and 2 and buyer 2 can also produce 3 units of surplus, therefore buyer 1 cannot get any surplus in the core. Yet seller 2 and buyer 1 do have the opportunity to trade for one unit of surplus.

Non-empty cores

A large body of research (starting with the pioneering papers of Bondareva (1962) and Scarf (1967, 1971) has explored in great details the conditions under which the core of a cooperative game is non empty. The results include both abstract characterizations and specific applications in economic and political models.

The most important abstract results generalize the idea that the core will be non empty if and only if the per capita surplus of proper coalitions does not exceed the per capita surplus of the grand coalition (this is the balancedness property introduced by Bondareva and generalized by Scarf); other results show that the core is non empty if certain coalitions are systematically prevented from forming, for instance if the players are divided in two classes, say blue and red and all the opportunities to trade happen in the coalitions containing a blue player and a red player (Kaneko and Wooders 1982).

The core of voting games gives important insights into the trade-offs between decisiveness of the voting rule and the stability of agreements among coalitions of voters (a stylized representation of political parties). Clearly under the unanimity rule any efficient agreement improving upon the status quo is stable in the sense of the core (because all coalitions smaller than the grand coalition are deprived of power entirely; at the same time the unanimity rule is quite indecisive: it will stick to the status quo if only one voter opposes the move. At the other extreme, majority voting is quite decisive but leads easily to empty cores (as in the divide the dollar game). Somewhere between these two extremes is the "optimal" configuration of voting rights guaranteeing a non empty core at all preference profiles and as decisive as permitted by the above constraints. This configuration allows any coalition to "veto" a number of alternatives proportional to its size (Moulin 1981; see

also Nakamura 1979, and Moulin and Peleg 1982 for more general results in the same vein).

The core of exchange economies has been studied in great depth. It has been proven non empty when individual preferences are convex (decreasing marginal utilities) but the possibility of an empty core arises as soon as even one agent has non convex preferences (as shown in the second example above). Moreover, when preferences are convex, the core has deep structural connections with the set of competitive equilibrium allocations. In the earlier example, the two sets coincide and this is no coincidence. Generalizing this example leads to the most important result of cooperative game theory. It says that in exchange economies with a large number of agents (each individual agent holding a small fraction of total resources) the set of core allocations and the set of competitive allocations are equal (Debreu and Scarf 1963; see Moulin 1995 for a broad introduction to this result, often called the "Edgeworth proposition"). Therefore, in a broad class of exchange economies, restricted only by the assumption of decreasing marginal rates of substitution, the core captures the fundamental idea of competition among traders and sits at the heart of the "theory of value." In production economies, the same basic link of the core to the competitive idea is preserved as long as the production technologies exhibit decreasing returns to scale (more precisely, as long as the set of feasible aggregate production plans is convex). When the returns to scale are increasing, the notion of competitive price collapses but the core shows more robustness (even though it may also be empty, there are many interesting cases where it is not; see Moulin 1995: Chapter 2, for an introduction).

Empty cores

We turn now to the second question: how do we interpret a cooperative game of which the core is empty? The problem is that our players can only reach a stable agreement if at least one smaller coalition refrains from exploiting its property rights to its unanimous advantage: this seems to contradict the efficiency postulate itself. There are two types of answers.

The first type of answer (of which the first formulation is in *TGEB*) maintains the efficiency postulate for all coalitions and explains the stability of certain agreements by the fact that some member i of a coalition that could profitably "object" to the initial agreement anticipates that, following the move of this coalition, other coalitions will form and that, in turn, he (agent i) will end up worse off than he was under the initial agreement. In the divide the dollar example, the split (1/3, 1/3, 1/3) is deemed stable because an objection like (1/2, 1/2, 0) (players 1 and 2 get together and exclude player 3) will generate a counter objection like (0, 0.6, 0.4) (player 3 bribes player 2 away from her initial deal with player 1, who is left in the cold); anticipating this, player 1 refrains from joining the initial objection and the equal split allocation is stable after all.

The precise definition of a general stability concept along those lines is not a simple matter. Von Neumann's notion of a stable standard of behavior is one (mathematically elegant) such concept: he proposes to think of a set of stable outcomes (instead of a single division of the surplus, a set of possible divisions) and defines a set A of outcomes as stable if (1) any objection to an outcome in A takes us to an outcome outside A, and (2) to any outcome outside A there is an objection taking us back to A. This concept brought some spectacular insights into the process of coalition formation: for instance in the divide the dollar game, one stable set reduces the three outcomes (1/2, 1/2, 0), (0, 1/2, 1/2) and (1/2, 0, 1/2), thus making the plausible prediction that one two-player coalition will form and exploit the third player. Unfortunately, there are many other stable sets: even in the simple divide the dollar game, there are infinitely many of them, most of them with fairly bizarre shapes: Moreover, it turns out that there are cooperative games with no stable set whatsoever (counterexamples were discovered more than twenty years after the publication of *TGEB* and they involve 10 players or more, which gives an idea of the mathematical complexity of Von Neumann's concept!): in the end, the notion of stable sets does not justify the efficiency postulate anymore than the core does.

The bargaining set (of which the first version is due to Aumann and Maschler 1964; it was later "refined" successively by Mas-Colell 1989, Vohra 1991, and Zhou 1994, offers such a justification. It uses a stability property of a single outcome based on the argument given above for equal split in the divide the dollar game, and is never empty. However the bargaining set is in general quite big (it does not narrow down the set of potential agreements very much), and hard to compute. Applications of the concept to economic or political models have been slow to come.

The second type of answer to the puzzle raised by the emptiness of the core is more radical. It amounts to postulating that the instability generated by an empty core is in itself a source of genuine collective costs. The problem is easiest to see in the context of majority voting, where the emptiness of the core is equivalent to the well-known configuration of cyclical majorities (or absence of a Condorcet Winner). In order to explain the "normative assumption – usually implicit but sometimes fairly explicit – that majority cycling is an undesirable political phenomenon, something that we should hope to avoid insofar as possible," Miller gives the following two reasons:

> political choice cannot be stable – for example, no parliamentary government pursuing any set of policies can win a constructive vote of no confidence against every alternative. More specifically, the process of logrolling does not lead to stability, as logrolling coalitions can form and reform in an endless (cyclical) sequence.
>
> Electoral competition between two power-oriented political parties or candidates cannot lead to equilibrium. No matter what platform or set of

> policies one party selects, it can always be defeated, and the outcome of electoral competition – even if modelled under the assumption of complete information – is intrinsically indeterminate and unpredictable, and the resulting electoral victories and attendant outcomes are thus arbitrary.
>
> (Miller 1983)

(Note that Miller gives two additional reasons, one based on non-cooperative manipulations of the agenda, and one based on the violation by the social choices of the collective rationality axiom proposed by Arrow.)

In exchange and production economies, examples of empty cores seem less pervasive than in majority voting and yet they are easily constructed: we gave an exchange example where one buyer has increasing marginal utility; it is perhaps instructive at this point to mention a famous example of Faulhaber (1975) showing how competition as captured by core stability can fail to achieve the efficient outcome. Two identical firms compete to provide a service (say a connection to a phone network) to three identical consumers. Each consumer is willing to pay 7 for the service and the cost function of a firm is as follows:

Units	1	2	3
Cost	6	9	14

Efficiency demands serving all three consumers by a *single firm*, for a total surplus of $7 \times 3 - 14 = 7$. However, any two consumers dealing directly with one firm can generate a surplus of $7 \times 2 - 9 = 5$ or a per capita surplus of 2.5 exceeding 7/3 and consequently the core is empty. The problem here is that if one firm serves all three customers, it must charge a price of at least 4.7, thus making it profitable for a competing firm to lure two customers away (offering them a price of e.g. 4.6), thereby putting the first firm out of business. This is a case of destructive competition calling for a regulation of the "natural monopoly." In the words of Faulhaber: "In order to insure the stability of the most efficient cooperative solution, coercive intervention to restrict market entry becomes necessary" (Faulhaber 1975).

The collectivity of all concerned agents realize that the free, unrestrained formation of alliances will lead to ever recurring destabilizing competition, hence will fail to achieve any efficient outcome at all. Restricting market entry eliminates competition by limiting the freedom of contracting between a firm and a subgroup of the concerned customers.

In the case of cyclical majorities, curtailing the power of coalitions means raising the quotas necessary to reach a decisive consensus. In the case of mass elections, however, it appears that cycles may be an inherent characteristic of democratic systems, one that fosters the pluralism of the political process (in the sense that the losers of today will be the winners of tomorrow), and

restores a kind of higher order stability through the alternation of winning coalitions. In this pluralist view, empty cores are, in fact, more desirable than non-empty ones (Miller 1983).

In both interpretations of the games with an empty core, we and up predicting that the application of the efficiency postulate will be limited to the grand coalition, that intermediate coalitions will not use their cooperative opportunities to the fullest. In the case of the second order stability concepts (e.g. the bargaining sets) coalitions do not form because some members refrain from entering a coalitional deal that is only profitable in the short run; in the case of regulated monopoly, coalitions are forcibly prevented to form (or in the case of increased quotas their formation is made more difficult). And finally, a "dissident" interpretation of core emptiness attaches value to the cyclical instability itself. In all cases the discussion of a cooperative game with a non-empty core and of one without a core are sharply different. Emptiness or non emptiness of the core is an essential qualitative feature of those resource allocation problems where the efficiency postulate is realistic.

The value of a game

The motivation of the axlomatic work discussed in this section is the opposite problem of core emptiness, namely the fact that the core is often too large. Indeed in any cooperative situation where the decision power is fully indivisible (where the strict rule of unanimity prevails) the core is simply the set of efficient outcomes that improve upon (in the Pareto sense) the status quo (initial position); the corresponding set of possible distributions of the cooperative surplus (often called the bargaining range) is normally large, that is to say the efficiency postulate has little if any discriminatory power. Recognizing this as a problem, Nash's seminal paper on axiomatic bargaining starts as follows:

> The economic situation of monopoly versus monopsony, of state trading between two nations, and of negotiation between employer and labor union may be regarded as bargaining problems. It is the purpose of this paper to give a theoretical discussion of this problem and to obtain a definite "solution" – making, of course, certain idealizations in order to do so. A "solution" here means a determination of the amount of satisfaction each individual should expect to get from the situation, or, rather, a determination of how much it should be worth to each of these individuals to have this opportunity to bargain.
>
> (Nash 1951)

Another seminal paper of the axiomatic literature, that by Shapley in 1953, is motivated as follows:

> At the foundation of the theory of games is the assumption that the

> players of a game can evaluate, in their utility scales, every "prospect" that might arise as result of a play. In attempting to apply the theory to any field, one would normally expect to be permitted to include, in the class of "prospects," the prospect of having to play a game. The possibility of evaluating games is therefore of critical importance. So long as the theory is unable to assign values to the games typically found in application, only relatively simple situations – where games do not depend on other games – will be susceptible to analysis and solution.
>
> (Shapley 1953)

Thirty years later, Shubik, who more than any other scholar helped develop the application of these axiomatic "values" to economic and political science, introduces the chapter of his textbook on the Shapley value by the following quote: "The value or worth of a man, is as of all other things, his price; that is to say so much as would be given for the use of his power" (T. Hobbes, *Leviathan*).

All three authors agree that a deterministic solution to each and every bargaining solution is valuable as a matter of principle. Shapley views this as a necessary condition for analysis. Nash and Shubik use the image of the market value determined by the competitive pressure of exchanges to postulate the existence of an equilibrium price for participating in the game.

Note that the goal of picking a unique solution outcome, of computing the value of a game, if it is self-evident to Nash, Shapley, and Shubik, is not a priori compelling. To many die-hard liberals it is an entirely inappropriate objective:

> Under the individualistic postulates, group decisions represent outcomes of certain agreed-upon rules for choice after the separate individual choices are fed into the process. There seems to be no reason why we should expect these final outcomes to exhibit any sense of order which might, under certain definitions of rationality, be said to reflect rational social action.
>
> (Buchanan and Tullock 1962)

> Man has developed rules of conduct not because he knows but because he does not know what all the consequences of a particular action will be. And the most characteristic feature of morals and law as we know them is therefore that they consist of rules to be obeyed irrespective of the known effects of the particular action. How we should wish men to behave who were omniscient and could foresee all the consequences for their actions is without interest to us. Indeed there would be no need for rules if men knew everything – and strict act-utilitarianism of course must lead to the rejection of all rules.
>
> (Hayek 1976)

The antinomy of the two positions would be only superficial if the determination of the unique solution/value resulted from positive equilibrium analysis: the computation of the value would reflect, then, a better understanding of the strategic parameters of the bargaining situation under consideration, and would be of the same nature as the shrinking of the efficient frontier by considerations of core stability. Clearly, Shubik has something like this in mind when he uses the analogy of market value determined by competitive pressures. And the "Nash program" discussed four paragraphs below is an attempt (in my opinion a half successful attempt) to provide equilibrium foundations to the value/solution of a cooperative game.

However the enduring contribution of Nash's and Shapley's seminal work lies in their effective use of a handful of plausible axioms to characterize a unique compromise within the efficiency frontier. These axioms are ultimately justified by normative arguments (a typical example is the axiom of symmetrical treatment of the agents, appearing in several variants under the name of anonymity, equal treatment of equals, one man one vote, or, simply, symmetry): they are the basis of a voluntarist interpretation of collective rationality, whereby some central authority legalized by the collectivity as a whole (the social planner, the benevolent dictator, the state, and so on) selects a specific outcome guided by some publicly known normative principles (those principles engraved in the constitution) and enforces this outcome (individual social actors being coerced into obedience). This approach to social cooperation does run counter to the grain of the liberal tradition poised to minimize the extent to which the collectivity restrains individual freedom of action (see e.g. the minimal state of Nozick 1974).

The need for the principled selection (and collective enforcement) of a compromise on the efficiency frontier can be justified by at least two positive arguments. On the one hand it is a more efficient mode of decision making: reaching unanimous agreement is a slow process with high transaction costs, the stress of haggling and hassling is borne by all social actors. This effectively rules out the unanimity rule for decisions requiring fast processing (e.g. the conduct of war) or involving a huge number of participants (mass elections). On the other hand a just collective diktat is more likely to build consensus than the arbitrary outcome of face to face bargaining, in which luck and bargaining skills select a point within the broad range of core outcomes (of course the unjust diktat of a dictator threatens the consensus even more!).

Yet the heart of the matter is the selection of a particular formula for compromise: say the maximization of the sum or of the minimum or of some other combination of individual utilities; the Shapley value versus the nucleolus; majority voting versus the Borda rule, and so on: see Moulin (1988) for a survey of the axiomatic literature. In the end, we must choose one such formula: their axiomatic comparison is the only way to inform our judgement.

The Nash program (Nash 1953) is an attempt to complement the axiomatic

discussion. The idea is to analyze the bargaining problem by means of a non cooperative "bargaining game" where the strategic moves open to the players mimic the bargaining tactics used in real face-to-face negotiations. Thus the decision power is fully decentralized among the participants in some conventional fashion. If the rules of the bargaining game are cleverly chosen the (non cooperative) equilibrium outcome of the game will be unique, defining a solution, a value for the initial bargaining game; we can then compare this equilibrium solution with the solutions recommended from the normative viewpoint. If the equilibrium solution coincides with a certain (axiomatically derived) solution, we have "implemented" this normative solution by means of a decentralized bargaining procedure.

Nash's initial proposal of a bargaining game to implement Nash's axiomatic solution was rather clumsy (Nash 1953) but subsequent work by Harsanyi 1963 and Rubinstein 1982 uncovered some very plausible bargaining games implementing that very solution. The latter, in particular, imagined that the two bargaining agents take turn making offers and counter offers, with a small probability of the negotiation breaking down after each rejected proposal (see also Young 1993 for a different evolutionary equilibrium story). In a similar vein Gül (1989) and Hart and Mas-Colell (1996) implement the Shapley value by means of a game of offers and counter offers where the author of a rejected offer faces a small chance of being eliminated from the game, and where the players are called at random to formulate an offer (i.e. to propose a way to divide the overall surplus among all remaining participants).

The Nash program draws some fascinating connections between the axiomatic-centralized and the strategic-decentralized approaches to the allocation of cooperative surplus (i.e. the bargaining problem). It gives us some clear answers to the implementation problem. It does not, however, bypass the need for an axiomatic discussion of solutions. To the extent that different solutions can be implemented by different bargaining games (e.g. see Moulin for a game implementing the equal relative benefits rule axiomatized by Kalaï and Smorodinsky 1975, on the Nash program see Binmore 1987) we still have to rely on normative arguments to choose within the limitless diversity of potential solutions (or, for that matter, bargaining games).

Conclusion: whither cooperative games?

After the profound impact of Debreu and Scarf's theorem, cooperative game theory went through a period of relative decline (say from the early 1970s to the mid-1980s); today, cooperative game theory is alive and kicking once again.

If the search for the exclusively positive concepts of stability and coalition formation is moving forward at a slow pace, active research is being conducted to explore the connections between non cooperative bargaining games and axiomatic solutions.

Perhaps more importantly, the axiomatic approach initiated by Nash and Shapley more than 40 years ago has deeply influenced a host of normative economic and political questions. Without any pretense to comprehensiveness, I list a few typical examples of this trend: indexes of political power (Shapley and Shubik 1954), axiomatic cost sharing (Billera and Heath 1982; Tauman 1988; Young 1990), the regulation of natural monopolies (Sharkey 1982) the cost sharing of public goods (Champsaur 1975; Moulin 1987), the fair division of private goods (Thomson 1983; Moulin and Thomson 1988; Moulin 1992) and the cooperative production of private goods (Roemer 1986; Maniquet 1994). Moulin (1995) gives a simple introduction to some of these developments.

Cooperative game theory has finally merged with the mainstream of normative economics, along with social choice theory and implementation theory, where I expect that it will live happily and have many children.

Acknowledgement

Stimulating comments by William Thomson have been very helpful.

Note

1 This well-known argument goes back to Böhm-Bawerk. See e.g. Moulin (1996: Ch. 2).

References

Aumann, R. J. (1987), "Game Theory" in *Game Theory, The New Palgrave*, Vol. 2, J. Eatwell, M. Milgate, and P. Nawman, eds, London: Macmillan.

Aumann, R. J. and Maschler, M. (1964), "The Bargaining Set for Cooperative Games," pp. 443–6 in *Advances in Game Theory* (Annals of Mathematics Studies, 52), M. Dresher, L. S. Shapley, and A. W. Tucker, eds, Princeton, NJ: Princeton University Press.

Billera, L. and Heath, D. (1982), "Allocation of Shared Costs: A Set of Axioms Yielding a Unique Procedure," *Mathematics of Operations Research*, 7, 1, 32–9.

Binmore, K. G. (1987), "Nash Bargaining Theory," pp. 61–76 in *The Economics of Bargaining,* K. G. Binmore and P. Dasgupta, eds, Oxford: Blackwell.

Bondareva, O. N. (1962), "Teorlia iadra v igre n lits, Vestnki Leningrad University," *Mathematics, Mechanics, Astronomy*, 13, 141–2.

Buchanan, J. and Tullock, G. (1962), *The Calculus of Consent*, Ann Arbor: University of Michigan Press.

Champsaur, P. (1975), "How to Share the Cost of a Public Good?" *International Journal of Game Theory*, 4, 113–29.

Coase, R. H. (1960), "The Problem of Social Cost," *Journal of Law and Economics*, 3, 1–44.

Debreu, G. (1959), *Theory of Value*, New York: Wiley.

Debreu, G. and Scarf, H. E. (1963), "A Limit Theorem on the Core of an Economy," *International Economic Review*, 4, 235–46.

Faulhaber, G. (1975), "Cross Subsidization: Pricing in Public Enterprises," *American Economic Review*, 65, 966–77.
Gül, F. (1989), "Bargaining Foundations of the Shapley value," *Econometrica*, 57, 81–4.
Harsanyi, J. (1963) "A Simplified Bargaining Model for the n-Person Cooperative Game," *International Economic Review*, 4, 194–220.
Hart, S. and Mas-Colell, A. (1996), "Bargaining and Value," *Econometrica*, 64, 2, 357–80.
Hayek, F. (1976), "The Mirage of Social Justice," Vol. 2 of *Law Legislation and Liberty*, Chicago: University of Chicago Press.
Kalaï E. and Smorodinsky, M. (1975), "Other Solutions to Nash's Bargaining Problem", *Econometrica*, 43, 3, 513–18.
Kaneko, M. and Wooders, M. (1982), "Cores of Partitioning Games," *Mathematical Social Sciences*, 3, 313–27.
Maniquet, F. (1994), "On Equity and Implementation in Economic Environment," Thesis, Faculté des Sciences Economiques et Sociales, Namur.
Mas-Colell, A. (1989), "An Equivalence Theorem for a Bargaining Set," *Journal of Mathematical Economics*, 18, 129–39.
Miller, N. (1983), "Pluralism and Social Choice," *American Political Science Review*, 77, 734–5.
Moulin, H. (1981), "The Proportional Veto Principle," *Review of Economic Studies*, 48, 407–16.
Moulin, H. (1987), "Egalitarian Equivalent Cost-Sharing of a Public Good," *Econometrica*, 55, 4, 963–77.
Moulin, H. (1988), *Axioms of Cooperative Decision Making*, Monograph of the Econometric Society, Cambridge: Cambridge University Press.
Moulin, H. (1992), "An Application of the Shapley Value to Fair Division with Money," *Econometrica*, 60, 6, 1331–49.
Moulin, H. (1996), *Cooperative Micro-Economics: a game theoretic introduction*, Princeton: Princeton University Press.
Moulin, H. and Peleg, B. (1982), "Stability and Implementation of Peleg (1982), Effectivity Functions," *Journal of Mathematical Economics*, 10, 1, 115–45.
Moulin, H. (1988), "Can Everyone Benefit from Growth? Two difficulties," *Journal of Mathematical Economics*, 17, 339–45.
Nakamura, K. (1979), "The Vetoers in a Simple Game with Ordinal Preferences," *International Journal of Game Theory*, 8, 55–61.
Nash, J. (1951), "The Bargaining Problem," *Econometrica*, 18, 155–62.
Nash, J. (1953), "Two-person Cooperative Games," *Econometrica*, 21, 128–40.
Nozick, R. (1974), "Anarchy, State, and Utopia," New York: Basic Books.
Owen, G. (1968), *Game Theory*, New York: Academic Press.
Roemer, J. (1986), "Equality of Resources Implies Equality of Welfare," *Quarterly Journal of Economics*, 101, 751–84.
Rousseau, J. (1762), *Du Contrat Social*, Paris.
Rubinstein, A. (1982), "Perfect Equilibrium in a Bargaining Mode," *Econometrica*, 50, 97–109.
Scarf, H. (1967), "The Core of an N Person Game," *Econometrica*, 35, 50–69.
Scarf, H. (1971), "On the Existence of a Cooperative Solution for a General Class of N-Person Games," *Journal of Economic Theory*, 3, 169–81.
Shapley, L. (1953), "A Value for N-Person Games," pp. 307–17, in *Contributions to the*

Theory of Games II (Annals of Mathematics Studies, 28), H. W. Kuhn and A. W. Tucker, eds, Princeton, NJ: Princeton University Press.
Shapley, L. and Shubik, M. (1954), "A Method for Evaluating the Distribution of Power in a Committee System," *American Political Science Review*, 48, 787, 792.
Sharkey, W. (1982), *The Theory of Natural Monopoly*, Cambridge: Cambridge University Press.
Shubik M. (1982), *Game Theory in the Social Sciences*, 1, Cambridge, MA: MIT Press.
Tauman, Y. (1988), "The Aumann–Shapley Prices: A Survey," in *The Shapley Value: Essays in Honor of Lloyd Shapley*, Al Roth ed., Cambridge: Cambridge University Press.
Thomson W. (1983), "Problems of Fair Division and the Egalitarian Solution," *Journal of Economic Theory*, 31, 211.
Vohra, R. (1991), "An Existence Theorem for a Bargaining Set," *Journal of Mathematical Economics*, 20, 19–34.
Von Neumann, J. and Morgenstern, O. (1944), *Theory of Games and Economic Behavior*, New York: Willey.
Zhou, L. (1994), "A New Bargaining Set of an N-Person Game and Endogenous Coalition Formation," *Games and Economic Behavior*, 6, 512–26.
Young, H. P. (1990), "Producer Incentives in Cost Allocation," *Econometrica*, 53, 4, 757–65.
Young, H. P. (1993), "An Evolutionary Model of Bargaining," *Journal of Economic Theory*, 59, 1, 145–67.

5 The coalition concept in game theory

Sébastien Cochinard

Introduction

One of the most difficult and major problems of game theory consists in understanding how players may choose to organize themselves in a game in order to get joint-maximizing profits. The problem raises in fact two different questions: which organizations are likely to emerge from a game (first problem) and how are members of these organizations going to share its gains (second problem)? Hence, if these organizations are coalitions of players, a solution to the game shall specify those coalitions that are likely to form (ex post stable coalitions and/or coalitions that could form during the process at an intermediary step) and the way players, in each coalition, shall share common utility they get through their coalition. In *Theory of Games and Economic Behavior* (*TGEB*) Von Neumann and Morgenstern proposed developing a theory of n-person games in the framework of the characteristic function model in order to solve the second problem. Later, Shapley (1953), Aumann and Maschler (1964), and many other authors have responded to the question by proposing various solution concepts: respectively, the Shapley value, the bargaining set, the core, the kernel, the nucleolus, etc. (Although the core is a major solution concept in cooperative games, we will not tackle it here, except in comparing it to bargaining sets, because we want to focus our attention on coalition structure. We will only give a brief definition and speak of the "core of a coalition structure." We refer to Hervé Moulin's chapter in this book for a presentation of the core.)

In cooperative game theory it seems hard to disentangle the two problems: the way players share the worth of the coalition depends on which coalitions are likely to emerge from the game. It is this link between the two problems that research on coalition structure games tries to elucidate, seeking a better understanding of how a given coalition structure affects each player's individual payoff. Once this question is answered (and there are many ways to answer it), one can face the first problem (how players choose to organize themselves) from different viewpoints. Either one links explicitly the two problems considering that the "sharing" game within each coalition and the game between coalitions that determines stable coalitions should have

common features,[1] or one separates them. A way of avoiding arbitrary or ad hoc aspects of a specific choice of a solution concept may consist in introducing first a non-cooperative step into the game, which describes the whole set of possible links among players. Starting from a graph-theoretic extension of the Shapley value, Aumann and Myerson (1988) propose for instance justifying the existence of a cooperation structure in a cooperative game by a non-cooperative "linking" game, the issue of which is a subgame perfect Nash equilibrium.

Both cooperative and non-cooperative games are needed in order to analyze coalitions. Following Nash's (1951) ideas, one should consider these approaches as complementary, rather than antagonist. As a matter of fact, bargaining theory and literature about non-cooperative implementation of cooperative solution concepts shows under which conditions one can design extensive-form games in order to get "efficient" or equitable payoffs (in the sense of the core, of Shapley value or of the Nash bargaining solution) as subgame-perfect Nash equilibria of these non-cooperative games (see for instance Moldovanu 1992, Moldovanu and Winter 1992, or Perry and Reny 1994). In a non-cooperative environment, i.e. without binding agreements between players, but where free communication of strategies is possible, Bernheim, Peleg, and Whinston (1987) propose a refinement of Nash equilibrium (the "coalition-proof Nash equilibrium" or CPNE for short) that takes into account some coalitional deviations. Many other coalitional refinements of non-cooperative equilibria have been proposed in various contexts, for instance "coalition-proof communication equilibria" or "strong communication equilibria" (Einy and Peleg 1992) which both refine communication equilibria in games with incomplete information.

One must recognize the revival of interest in cooperative solution concepts and coalition formation theories, and it is not merely due to the development of implementation literature. Theories or models have been developed that aim at highlighting the close relation between cooperative and non-cooperative representations, as "TOSS" in Greenberg 1990. They allow for comparisons of payoffs evolving from different bargaining situations (see Xue 1993; or Arce 1995), which is of great help to economists. In the wake of Von Neumann and Morgenstern's ideas,[2] this research aids an understanding of how various frameworks (institutional settings, negotiation framework, social norms, behavioral assumptions, etc.) may stand behind well-known solution concepts. In another direction, cooperative game theory deals also with games with graph-restricted[3] or hierarchical communication structures (see for instance the work of van den Brink and Gilles or of Vasquez-Brage *et al.* in the February 1996 issue of *Games and Economic Behavior*; see also Amer and Carreras 1995; Derks and Gilles 1995 in the *International Journal of Game Theory*).

Models of coalitional-form games

The Von Neumann and Morgenstern model of coalition

Any theory of cooperative games with N (N>2) players must include an analysis of coalition[4] because one cannot disregard the fact that in such games, subsets of players (different from N) have typically the possibility of cooperating and choosing joint strategies in order to share a common payoff. Indeed, if in a two-player game, everything adds up so that each player has to decide if she is willing to cooperate or not,[5] the problem seems more intricate with more than two players, since one should wonder with whom one wants to enter into partnership and what non-partners are going to do. Aiming at simplifying an exhaustive treatment of all these options, one can decide to "sacrifice the strategic structure of the game" (Weber 1994: 1286) and merely consider its coalitional aspects.

The model used by Von Neumann and Morgenstern, based on the characteristic function, lies on two additional behavioral assumptions concerning players (see Roth 1988): (1) opportunities made to a coalition of players are established without any reference to players outside the coalition, and (2) members of a coalition may conclude costless binding agreements to share the coalition's worth in any way, so that it is not necessary to model explicitly the actions players intend to adopt in order to reach these agreements.

A game (N, v) in characteristic form or in coalitional form (also called coalitional game, for short) is defined by a set of players N and a function v, the characteristic function of the game, which assigns a real number called the "worth" of the coalition to each subset of N (each coalition). The word "coalition" is used here in a neuter sense without a priori institutional or structural references. The function v, the number and identity of players, together with the communication rules among players completely define the game. Once this coalitional representation has been specified, players form the "grand coalition" N and bargain to share the total value v(N). In the most general model, the results of the allocation process depend more on the distribution of power among the players, as described by values of v, than on the process itself (the undefined players' strategies). Thus, the characteristic function is really a description of the structure of power in the game.

The "Von Neumann and Morgenstern solution" has been historically the first solution concept studied in this class of games. In order to find it, one first defines an imputation of the game. An imputation $z = (z_i)$, i = 1 to N, is a vector of payoffs (one payoff z_i for each player in N) that are both individually rational (for each player i, $z_i \geq v(\{i\})$, where $v(\{i\})$ stands for individual worth of player i as a singleton-coalition) and efficient (or collectively rational, that is: $\sum_{i=1}^{N} z_i = v(N)$). Moreover, one defines a comparison criterion between payoff vectors: domination. A payoff vector x dominates another pay-

off vector y through coalition S if, for each player $i \in S$, $x_i > y_i$ and $\sum_{i \in s} x_i \leq v(S)$. More generally, x dominates y if there exists at least one coalition S such that x dominates y through S. Then the Von Neumann and Morgenstern solution (hereafter vN&M solution) of a game (N,v) is the set I of imputations not dominated by elements in I. One can interpret I as the set of stable organizations (one also speaks of "stable set" for vN&M solution), within which the same players can however be treated differently by receiving different payoffs, depending on the element of I considered (see Aumann 1989).

The coalition structure is already implicit in *TGEB*: internal stability (a solution does not dominate another solution) and external stability (given an imputation lying outside I, there exists a solution which dominates it in I) of the set of solutions in fact isolates stable structures (Aumann and Myerson 1988: 176; Kurz 1988: 155).

However, assumption (1) of Von Neumann and Morgenstern's model seems rather unrealistic in many real-life situations. We should be able at least to specify its validity range: Harsanyi (1977) shows that this assumption is indeed sustainable if we admit that there is a symmetric repartition of available information among coalitions. Coalitions in Von Neumann and Morgenstern are rather static objects and taking into account more complex negotiations among coalitions (like "cutting across" coalitions, as suggested in Aumann and Drèze 1974: 231) would turn out to be a difficult task in their framework. Furthermore, one can exhibit games whose stable sets are empty (for instance the famous, but far from trivial, ten-player game of Lucas 1968).

One can extend Von Neumann and Morgenstern's model in three main directions: (1) obtaining a solution concept which, for any coalitional-form game, would give a unique vector as "the" solution of the game: the Shapley value answers this issue in an axiomatic way; (2) allowing bargaining at the level of "middle" size coalitions (say a bargaining between S and the complementary coalition N\S for instance): that is what can be found in Harsanyi (1977) with the "generalized Shapley value"; (3) taking into account more complex situations where players are constrained to bargain only with some permissible coalitions within a "coalition structure" and, possibly, with different levels of coalition structures (a player i may belong to coalition S and to coalition T with $S \not\subset T$ and $T \not\subset S$): games with coalition structures (or "cooperation structures" in the sense of Myerson, 1977), where players' communication is limited through a partition of the set of players, through a "nested structure of unions" (see Owen 1977) or through a graph (see Amer and Carreras 1995) represent today the most powerful answers to these problems.

Coalition and the Shapley value

Shapley (1953) solves axiomatically the problem of finding a unique solution for any game in coalitional form (N, v). Three (or four, depending on the presentation) axioms are needed to find this solution, called the Shapley value. The first axiom, or symmetry axiom, states that any permutation of players in N leaves the Shapley value unchanged (as long as values taken by function v are changed properly, the position of player i in N does not matter). Then one introduces the concept of a "carrier" of a game in coalitional form. A coalition R is said to be a carrier of the game if and only if the worth of any coalition S remains unchanged when one restricts S to its intersection with R ($\forall S \subseteq N$, $v(S \cap R) = v(S)$). A "null player" or "dummy player" is a player who stays out of all carriers of the game. The Shapley value of a null player is zero. The second axiom, or axiom of the carrier, states that players in a carrier must share the total value of the carrier v(R) among them without allocating anything to null players (one has $v(R) = v(N)$). (One sometimes divides this axiom into a null player axiom and an efficiency axiom, which amounts globally to four axioms.) The third axiom, or additivity axiom, states that the Shapley value of game (N, v + w) equals the Shapley value of game (N, v) plus the Shapley value of game (N,w).

The remarkable result of Shapley lies in the proof that there exists only one vector function (one component per player), called the Shapley value, defined on the set of games in coalitional form whose set of players is N, that satisfies the axioms presented above. One should however note that the strongest requirement comes from the second axiom, since players in a carrier R ($R \subseteq N$, the complement, if it exists, consisting in null players) must share the total available worth for N, v(N). The Shapley value for player i, noted $Sh_i(v)$, is given by the following formula:

$$Sh_i(v) = \sum_{S \subseteq N \setminus i} \frac{|S|!(|N|-|S|-1)!}{|N|!} \left(v(S \cup \{i\}) - v(S)\right)$$

where |X| denotes the cardinal of X.

In order to compute the Shapley value for player i, it is sufficient to consider coalitions S to which i does not belong, counting for each of these coalitions the number of its members and knowing its worth v(S). (There are numerous tricks to short circuit this fastidious calculus and compute more quickly the Shapley value when faced with games presenting some kind of symmetry in coalitions' worth or in players' situations in N; see Aumann 1989, for instance.) One can also say that the above formula expresses the fact that the Shapley value can be seen as the weighted sum of marginal contributions (the $v(S \cup \{i\}) - v(S)$ terms) of player i to each coalition in N.

From the point of view of the player computing its Shapley value, there is a kind of "statistical" analysis of the coalition, each player evaluating the numerous potential games (Shapley 1953: 307 speaks of "prospects") and

realizing a "reasonable compromise" (Aumann 1989) among different options (of participation to such and such coalition) through an indicator which is the Shapley value. (The Shapley value can more formally be interpreted as the expected marginal contribution of a player, the expectation being calculated with respect to an adequate stochastic process.) The Shapley value offers a measure of the power of a player, in situations without a priori institutional restrictions (for instance on the bargaining abilities of players) other than those reflected by the characteristic function. Its domain of application ranges over a wide number of topics, especially in political science but also in economics. Much theoretical literature has been devoted to enlarging the set of games from which one can compute the Shapley value: multi-choice cooperative games (Hsiao and Raghavan 1986), cooperative games with coalition structures (Driessen and Tijs 1986).

The coalition model we have presented until now hides a major assumption concerning the crucial question of with whom players are able to negotiate, i.e.: when is negotiation an "effective negotiation" (Myerson 1986)? Players who negotiate "effectively" do so with the idea that the negotiation process will tend to fulfill their preferences in equilibrium (in the wide meaning of equilibrium, this discussion being relevant as well to non-cooperative game theory). They will never negotiate over an outcome which will turn out to be Pareto-dominated in *their* set of attainable outcomes. If there exists an imputation which improves each coalition member's payoff and if this coalition is negotiating effectively, then members should agree on this imputation unless this agreement contradicts a previous agreement some members would have with outside players (in the framework of another effective negotiation within another coalition). For example, this effectiveness assumption applies only to the grand coalition N in Nash bargaining model, it applies to all negotiations between $\{i\}$ and subcoalitions $S \subseteq N \setminus \{i\}$ in the computation of the Shapley value.

Taking into consideration other effective coalitions will lead to model situations where such symmetric treatments would appear unrealistic: thus, one can define which coalitions are able to negotiate effectively thanks to ψ-stable structures, to coalition structures or to an "affinities" graph. Moreover, one can imagine more subtle and realistic situations, where the effective negotiating ability of one coalition is balanced by the effective negotiating ability of other coalitions: what is proposed in literature is then the "objections"/"counter-objections" framework of bargaining sets. Now we present Harsanyi's (1977) extension of the Shapley value which relies on an effective negotiation between coalitions S and complementary coalitions $N \setminus S$.

Harsanyi's coalition model

Harsanyi (1977) departs from Von Neumann and Morgenstern's coalition model by enlarging the following assumption: opportunities made to a

coalition of players are established without any reference to players outside the coalition. In a first attempt to weaken this assumption, one may reduce the influence of players who stay outside the coalition S to the threat exerted by the whole set of these players, considered as a homogeneous group, i.e. one can view them as a coalition itself, this coalition being defined as the complementary coalition N\S. Therefore, we are led to analyze potential conflicts between S and N\S: we consider all coalitions S and their opponents N\S, then isolating for each player i the coalitions S to which he may belong. Taking into account this kind of conflict, we define each player i's utility as a function $U_i(\sigma_S,\sigma_{N\setminus S})$ of a threat strategy σ_S of S (it is a common strategy for all players in S) and $\sigma_{N\setminus S}$ of N\S: the worth v(S) of coalition S is nothing other than the sum of these utilities U_i for all players i belonging to S.

Three alternative assumptions have been suggested concerning coalitions' abilities to commit to offensive or defensive threats against other coalitions. These theoretical assumptions evolve from research aiming at deriving the characteristic function v from the strategic form of the game (i.e. seeing v as a function of the strategies of players): that is exactly what we have done in the previous paragraph, computing v(S) from σ_S and $\sigma_{N\setminus S}$. A first assumption consists in allowing players in S to guarantee themselves the maximal sum of individual payoffs against the best offensive threat of N\S: one speaks therefore of minimax representation (of the strategic-form game). In a second assumption, coalitions S and N\S play defensive equilibrium strategies against each other. The equilibrium strategy of S maximizes the sum of individual payoffs in S, and reciprocally for N\S: one speaks then of defensive equilibrium representation. A third assumption, developed by Harsanyi (1963), generalizes Nash's rational threat criterion and leads to the following equilibrium strategies for coalitions S and N\S: the equilibrium strategy of S (respectively, N\S) maximizes the sum of individual payoffs in S (N\S) minus the sum of individual payoffs in N\S (S): one speaks of rational threats representation. Using this criterion, Harsanyi obtains a "generalized Shapley value" which depends not only on marginal contributions of player i to coalitions S (the terms $v(S\cup\{i\})-v(S)$), but also on all conflicts between complementary coalitions (as summarized in $v(S)-v(N\setminus S)$).

To each of these assumptions on coalitions' strategic abilities, one can associate a different scenario for the coalitional-form game. These assumptions are never equivalent except in the case of games with "orthogonal coalitions" (Myerson 1991: 418–26). Such games appear for instance in the significant framework of pure exchange economies where a coalition's most offensive strategy against another consists in declining to bargain with it. In other contexts, a coalition S would indeed gain more by inducing a coalition T to bargain, rather than just refusing to play; in which case, it seems to be difficult to agree to assumption (1) of Von Neumann and Morgenstern's model. Indeed coalition S's offensive ability is established *with explicit reference* to another coalition T, i.e. to players outside original coalition S (this

offensive "ability" is nothing other than a strategy, this strategy of S could well vary for different coalitions T).

We proposed two answers to the first two extensions of coalition model suggested above; what about the third one, which no longer consisted in evaluating the different possibilities of belonging to a coalition as in the Shapley value, but rather examined the issue of shifting one's coalition? Let us take the example of a three-player coalitional-form game with players 1, 2, and 3, the problem being to consider player 1 leaving {1,2} to {1,3}.

Harsanyi (1977) suggests then distinguishing between discriminant and non-discriminant solutions to such a game. A discriminant solution is one where members of a coalition (say 1 in {1,2}) share its value (1 obtaining its generalized Shapley value with respect to ({1,2}, v) game) before bargaining with the complementary coalition (coalition {3} in this case). As for Harsanyi (1977), the discriminant solution is the consequence of a communication failure, i.e. of an asymmetry of information between coalitions (see on this point Pérez-Castrillo 1994, who proposes the following non-cooperative scenario: in case of information trouble, exogenous agents or institutions, alien to original players, compete in order to bring players on the coalitions they form.

Harsanyi's framework is one of perfect communication, whose solution is the non-discriminant one and where each player obtains its generalized Shapley value only after all bargainings between complementary coalitions are finished. In our example, player 1 effectively obtains its generalized Shapley value only after {1} has bargained with {2,3}, respectively {2} with {1,3}, and {3} with {1,2}. We will refer later to the distinction discriminant solution vs. non-discriminant solution in the coalition structure games setting in terms of CS value vs. AD value (extensions of Shapley value). The discussion will oppose two definitions of coalition (with respect to the goal of a coalition and the reasons for its formation) and distinct scenarios for the game in question. Here again, the timing of the steps "bargaining" and "sharing" (obtaining one's value) will be crucial in order to determine which is the "good" concept to apply. However, the distinction between CS and AD values will rely more accurately on the "essential" (super-additive) or "inessential" (merely additive) character of the game between coalitions. The equivalence between communication failure across coalitions and additivity of the game between coalitions is not obvious.[6]

After a brief presentation of the problem of sharing the coalition's worth among its members, we can observe that suggested answers cannot deal with situations where asymmetric treatment of coalitions, "effective" negotiation abilities of coalitions, complex "balanced" bargaining abilities are needed. We present now a first approach of the game-theoretic analysis of such situations thanks to ψ-stability and bargaining sets theories.

Taking into account coalition structure

Stable structure

As Von Neumann and Morgenstern (1944) recognize the potential existence of coalitions, they construct an N-person game theory different from their 2-person game theory. Nevertheless, Luce and Raiffa (1957) declare that they do not directly explain the coalition formation process. Indeed, in their view, the major obstacle to the development of a coalition formation theory lies in the absence of explicit assumptions concerning communication and collusion between players in the characteristic function model of Von Neumann and Morgenstern.

One should notice here the difference between the coalitional phenomenon and the collusive phenomenon. A model assuming the existence of coalitions without coalition structure does not explain collusion, since considering that any coalition may form as in classical coalitional-form games cannot lead to any particular prediction about how players are likely to collude (for an application of these ideas with core as solution concept and labor unions as coalition structure, see for instance Gabszewicz and Hansen 1972).

In their ψ-stability theory, Luce and Raiffa (1957: 220) developed the idea that by endowing a game with an a priori "coalition structure," one could explain, together with the usual structure of power as described by the coalitional form, limited collusion, i.e. existence of intermediate size coalitions, not reduced to mere singletons $\{i\}$ and different from N, "internally" cooperative and "externally" non-cooperative, as economic and social real-life situations may display.

As a matter of fact, one can observe that some coalitions are more easily obtained than others, and this phenomenon cannot be explained by the mere difference in coalitional payoffs. The origins of such an asymmetry are found in exogenous factors: whether they are historical, geographical, sociological, linguistic, political, or legal, they underlie the formation of "effective" coalitions such as labor unions. Otherwise, some endogenous factors operate too: the game structure itself is then involved in the determination of coalitions. Let us go back to our three-player game example and suppose moreover that the game is superadditive and majoritary (each two-player coalition earns the grand coalition's value). In this setting, two-player coalitions have more incentives to form and refuse independent (each player separately) negotiation with a third player, than in a unanimity game where only grand coalition has a non-null value (cf. Myerson 1991: 445).

Coalition structure is nothing other than a partition of the set of players, by which players are constrained in a first step to belong to such or such coalition. Starting from this hypothetical original situation (if player i belongs to coalition S, if S is given through a particular coalition structure), one can evaluate qualitatively the influence of an original coalition on future

coalitions. Function ψ represents the rule by which admissible coalitions change; it simply associates the set of all possible future coalition structures to each coalition structure. The final coalition structure is defined as the stable structure of the game.

The above discussion about the origins of a coalition structure is largely related to Harsanyi's theory of discriminant/asymmetric solutions. Only the conclusions differ. Considering a coalition structure is equivalent to making an assumption about communication and collusion between players in order to study the effects of such an asymmetry on the evolution of the composition of coalitions, i.e. on their stability. As for Luce and Raiffa, function ψ can only be computed on the basis of empirical research.[7] From a theoretical point of view, if one wants to determine the quantitative effects of a coalition structure, one is led to define "coalition structure games," and the associated extension of solution concepts (vN&M solution, Shapley value, etc.) from coalitional-form games to such games.

Bargaining sets and coalition structure

We have presented in the introduction the bargaining set (Aumann and Maschler 1964) as an approach to the problem of sharing the worth of a coalition among its members. In addition, the bargaining set takes into account our first problem of coalition formation as it is closely related to the concept of coalition structure. At last, it allows us to analyze situations of reciprocal coalitional threats, through a characterization in terms of objections/counter-objections, and to define stable situations thanks to the concept of justified objection.

The bargaining set is defined by a coalition structure B (where $B = (B_k)$, for $k = 1, \ldots K$) together with a set of payoff vectors satisfying conditions close to those satisfied by imputations (individual rationality and efficiency) in vN&M solution. In fact, we will simply extend the initial definition of the imputations set I for games (N, v) to games (N, v) including a coalition structure B: we will name this set X. It is defined as the set of payoff vectors x such that $x_i \geq v(\{i\})$, for i = 1 to N (x is individually rational, definition and interpretation are therefore identical for X and I) and we have $\forall k \in \{1, \ldots K\}, x(B_k) = \sum_{m \in B_k} x_m = v(B_k)$, i.e. the K players in each coalition B_k of B share completely the worth of their coalition among themselves (x is efficient in the narrow sense that it is efficient only within coalition B_k for the set X; in comparison z was efficient in grand coalition N for the set I).

Given a coalitional-form game (N,v), a coalition structure B, a payoff vector x in X, a coalition B_k in B and two players i and j in B_k, an *objection* of player i against player j is a pair consisting in a set S containing i and not j, and in a payoff vector in S, $y = (y_s)$ with $s = 1, \ldots S$, such that (1) $y_i > x_i$,

(2) $y_s \geq x_s$ for $s \in S \backslash \{i\}$ and (3) $y(S) = \sum_{s \in S} y_s \leq v(S)$. In words, saying that player i has an objection against player j is simply saying that i is able to enter a group S of players (here one speaks of a "group" rather than of a coalition, the latter being reserved for initial elements of B), to which j does not belong, in order to get a strictly better payoff than in its original coalition B_k, whereas other players obtain payoffs in S at least as good as in their respective original coalitions. Condition $y(S) \leq v(S)$ requires simply that members of a new group S cannot gain more than coalition S itself by adding up their individual payoffs.

A *counter-objection* is defined in the same manner by giving a set T containing j and not i, and a payoff vector in T, $z = (z_t)$, for $t = 1, \ldots T$, such that (1) $z_t \geq x_t$ for $t = 1, \ldots j, \ldots T$ (player j and other members of T obtain at least the same payoffs as initially), (2) $z_t \geq y_t$ for $t \in S \cap T$ (in order to draw members of S into T and thus "counter-object") and (3) $z(T) \leq v(T)$. Thus, we can model the negotiation process here as series of objections and replying counter-objections which correspond to potentially numerous threats between players, who are initially distributed according to a given coalition structure, and join little by little[8] distinct groups in order to form a new coalition structure. These groups gather players coming from coalitions that are not compelled to be complementary in N, as was the case in previously reviewed coalition models (see above). An objection against which no counter-objection exists is said to be "justified": it describes a situation where it is in player i's interest to quit her/his coalition to form a definitively stable new group, since she/he has a positive advantage to do so and there exists no credible threat against her/him. The bargaining set is then the set of payoff vectors against which no justified objection exists. It perfectly depicts a situation where the coalition structure is stable, since it is basically not worth relinquishing her/his coalition for any player, the consequences of such a decision being *neither optimal nor stable*. Here we emphasize that we need the non-existence of both justified objections in the bargaining set (optimality) and of any counter-objection in justified objections (this means we are then only concerned with stability against *credible* deviations).[9] Of course, the definition of the bargaining set is bound up with the initial coalition structure B, and one could suspect that the bargaining set is empty for some coalition structures. Nevertheless, it has been proved (Davis and Maschler 1963; other proofs have been given later) that whatever game (N,v) and coalition structure B, if X is non-empty, then the bargaining set is non-empty.

In the same way the Shapley value enables the measuring of players' power in a coalitional-form game; we would like to obtain a measure of objecting/counter-objecting players' ability, without having to compute explicitly the bargaining set. This approach leads to introducing the concepts of *excess* of a coalition, with respect to a vector x (simply defined as $v(S) - x(S)$, where

$x(S) = \sum_{s \in S} x_s$) and of *maximum excess* s_{ij} of a player i against a player j. Given v, B, B_k in B, i and j in B_k, S containing i and not j, we define the maximum excess as $s_{ij}(x) = \max \{v(S)–x(S), i \in S, j \notin S\}$: when player i considers the set of all coalitions to which he may belong and player j cannot, s_{ij} is the maximum gain i can get from belonging to coalition S, which represents also player i's "relative strength" vis-à-vis player j (the s_{ij} have to be compared with marginal contributions in the Shapley value). The kernel (Davis and Maschler 1967), which is a subset of the bargaining set, is defined, given v and B, as the set of payoff vectors x of X such that, for any B_k in B and for all i and j in B_k, either $s_{ij} \leq s_{ji}$ or $x_j = v(\{j\})$. In words, a vector x is in the kernel if, for all players i and j, the relative strength of player i vis-à-vis player j does not exceed the relative strength of player j vis-à-vis player i and, in the case where player i is "relatively stronger" than player j, i cannot threaten j to reduce her/his payoff to x_j, since this payoff is precisely lowered to its individually rational minimum level $x_j = v(\{j\})$ (see the definition of X above). If one considers the whole set of players j, coalitions S related to payoffs lying in the kernel are thus "balanced"[10] with respect to the relative strengths of player j. Other major developments have risen, such as the nucleolus or Maschler's α-power model (this model in particular and bargaining sets have led to experimental approaches of coalition's analysis; see Rapoport 1990), they also take into account considerations close to concepts of excess and maximum excess.

The core (formally, the set of imputations $z = (z_i)$, $i = 1, \ldots N$, such that $\forall S \subseteq N, \sum_{i \in S} z_i \geq v(S)$) depends on a selection criterion (namely, "blocking") close to objecting for bargaining sets: the core can be defined as the set of imputations against which there is no objection (a coalition that could object to an imputation would be said to "block" this imputation). However, no particular coalition structure is implicit in a core allocation, in opposition to bargaining sets (Kurz 1988). Where the core takes only into account original deviations, bargaining sets consider that potential threats (objections) launched by some players may be counter-balanced by other players (counter-objections). For a survey of recent developments in bargaining sets theory and of results linking core to bargaining sets, we refer to Einy and Wettstein (1996).

One may also wish to take into account games with a large number of players. Games with a continuum of players (Aumann and Shapley 1974) have been constructed in order to represent coalitions with a continuum of members (as represented by a real interval, say [0; 1]). For an economic example, see Gabszewicz and Hansen (1972), where one looks for properties of the core of a coalitional-form game with "unions," in the sense of a coalition structure made of two types of factor owners, themselves subdivided into members and non-members of unions (and not in the sense of

Owen 1977). A "mixed economy" is modeled with both continuum of agents and unions (as atoms of the characteristic function v; for a brief presentation of the measure-theoretic tools needed in this field, we refer to Weber 1994: section 14).

Aumann and Maschler (1964) and Davis and Maschler (1967), notably by explicitly defining coalition structure B and applying solution concepts to B rather than to N, have initiated an ambitious research program for game theory. First, one defines the coalition structure extension (N, v, B) of a coalitional-form game (N, v). Second, one generalizes coalitional-form games solution concepts to games (N, v, B) with coalition structure. Last, the third step of this research program aims at studying the stability of the coalition structure, and, in order to achieve this goal, isolates "stable structures" defined as solutions of a new game Gv, the "coalition formation game." Research on games with coalition structure follows the two directions of the core and of the Shapley value; we will treat, but not exhaustively, the second one.

However, solution concepts presented until this point, such as the bargaining set, consider the coalition structure as exogenously given and do not perfectly respond to our first problem, which consists in exhibiting coalitions supposed to emerge from the game. Research on endogenous fomation of coalitions (see Zhou 1994) aims at filling in this gap.

Coalition structure games and endogenous formation of coalitions

Aumann and Drèze research program

In coalition structure games, views on the coalition problem are changed. One looks at determining stable coalition structures rather than describing individually rational and efficient payoffs. The fundamental paper on coalition structures is by Aumann and Drèze (1974) as it defines the extension (N, v, B) of a coalitional-form game and generalizes to (N, v, B) games the following six solution concepts (relative to (N,v) games): the vN&M solution, the core, the bargaining set, the nucleolus, the kernel, and the Shapley value. Following this article, two solution concepts have essentially been used in order to analyze coalition formation: core and Shapley value.

The "core of a coalition structure" is defined (both definitions have been demonstrated to be equivalent, see Kurz 1988) either as the core of game (N, v, B), or as the set of non-dominated coalition structures, the domination rule being defined on coalition structures (and not merely on payoffs; see Shenoy 1979). Thus, the core and another solution concept named "dynamic solution" are used as stability criterions of coalition structures. It is worthwhile noticing here that solution concepts used for the domination rule in game (N, v, B) (where one compares different payoff vectors for one given coalition structure B) may in general differ from solution concepts (core, dynamic

solution) used in the coalition formation game where naturally B varies. On the contrary, it seems legitimate for Kurz (1988) to demand some consistency in the sense that both solution concepts should be identical (for instance, one will use Shapley value for games within each B_k of B, and also for the game between the B_k, identified as the coalition formation game). But before trying to explain how players organize themselves into coalitions, one should study how a given coalition structure modifies each player's payoff. In the next paragraph, we take the example of the Shapley value.

Coalition structure values

Aumann–Drèze value (hereafter AD value) is presented axiomatically, as the Shapley value, but axioms are stated relative to B_k rather than to N. Therefore one is again confronted with Harsanyi's discriminant/non-discriminant debate. With AD value, each coalition forms in order to obtain its value (i.e. coalitional payoff) and nothing more, the bargaining between coalitions being a distinct problem. The definition of the AD value includes (in the "relative efficiency" first axiom, also present in Owen 1977) the first assumption of von Neumann and Morgenstern's model: a player's payoff does not depend on her/his eventual contribution to a coalition lying outside its actual coalition (through an extra bargaining for instance).

One may share another view (see Kurz 1988), closer to the non-discriminant solution, that describes coalitions as forming only in order to get a posteriori a better bargaining power in the final bargaining over grand coalition's value and not merely to obtain their value at once. One defines in this framework a different value named CS value (for coalition structure). The discriminant solution is then just restricted to the particular case of inessential games where there is nothing to be shared among coalitions (we have $v(\underset{k\in K}{\cup} B_k) = \sum_{k\in K} v(B_k)$, so that no bargaining outside coalitions occurs. These distinct views about coalition's goal involve distinct scenarios in the coalition structure game. What game does really matter for final results: game inside each coalition B_k, "game" between coalitions B_k or game within "grand coalition" N?

In the AD value case, coalition B_k is the real entity which forms in the coalition structure to obtain its coalitional value $v(B_k)$. Members of coalition B_k bargain only among themselves in order to share $v(B_k)$. In the CS value case, coalition appears when all its members commit themselves to bargain (this commitment is precisely *credible* when coalition B_k is *stable*) and the real entity that forms at the end of the game is the grand coalition N. In this case, coalition structure B can be considered as a mere intermediate bargaining tool used to raise individual payoffs. This scenario involves some subtle negotiations between individual players within each coalition B_k and among all

coalitions B_1, B_2, . . . B_k, . . . B_K (considered as "players" negotiating in the name of the entire coalition).

One proceeds to the analysis of coalition structure stability in two different ways, depending on assumptions on other members' reactions to the departure of a player from her/his coalition. In a first model, coalitions that see one of their members going out collapse into singletons of players, whereas in a second model, players stay together in a reduced new coalition. There exists no general result on coalition structure stability for all N-players game; results may even differ depending on the stability model considered. However, studying stability in coalition structure games may help understand and predict the final social organization in these games. Games without stable coalition structures may thus model intrinsically unstable situations.

The distinction between AD value and CS value stands on the essential (super-additive) or inessential (additive) nature of the game between coalitions. Formally, either $v(\cup_{k\in K} B_k) > \sum_{k\in K} v(B_k)$ and in this case, CS value analysis shows that the "good" (effective) coalition to look at is N, or $v(\cup_{k\in K} B_k) = \sum_{k\in K} v(B_k)$ and in that case, one has to look at B_k and AD value is concerned. "Non-super-additivity of game (N, v) seems to be the most compelling explanation for the formation of a coalition structure" (Aumann and Drèze 1974: 232–4). Due to non-superadditivity, coalition B_k has no incentives to bargain with coalition B_l ($k \neq 1$), and hence players found themselves naturally bargaining within a coalition structure. Non-super-additivity is itself justified by the existence of asymmetries of information, communication problems or barriers of various natures (Greenberg 1994: 1308–10, establishes more precise causes of non-super-additivity in the case of local public goods economies).

If one does not want to consider any hypothetical game between coalitions, one is led with Zhou (1994: 515–16) to take the following position in the debate about the formation of subcoalitions S. As for Zhou, either one studies coalitions' formation, and non-super-additivity being one of its fundamental causes, the model is not super-additive and there is no other coalition to consider than "grand coalition" N; or the model is super-additive and then it seems hard to explain coalition formation in such a framework. Models[11] allowing players to belong to more than one coalition, but constraining the set of admissible coalitions, organize a real game between coalitions and show endogenous formation of coalitions.

Endogenous formation of coalitions

Aumann and Myerson (1988) take a different view on coalition formation, using an extension of the Shapley value (Myerson 1977). Instead of

considering disjoint coalitions (the coalition structure is indeed a partition), they look at more inclusive structures, namely "cooperation structures," which consist in cooperation graphs whose vertices figure players and links, bilateral cooperations between players forming pairs. Thus a non-cooperative linking game is built up, where players may choose if and with whom they establish such links. Moreover, one does not preclude the fact that a player might well have a "non-myopic view," enabling her to "predict" the future consequences of linking with a second player, himself potentially able to link with a third player, and so on. As the link formation scenario is thoroughly non-cooperative and public, one can assume a common knowledge "rule of order" (Aumann and Myerson 1988: 180) among the players. If one assumes in addition (1) that a link, once formed, can never be broken, (2) that this game is finite (the linking game comes to an end and the (N, v) game can finally take place) and (3) that after the last link has been formed, each pair of unlinked players is offered the opportunity to form a link, one can establish that the linking game always results in a well-defined cooperation graph.

Each player's payoff is evaluated with the Shapley value, extended to these cooperation structures, themselves resulting from the linking game described above. In this context, the Shapley value of a "classical" coalitional-form game can be interpreted as an evaluation of the "prospects" of players in the particular case where there exists free and total communication among players, i.e. where the cooperation graph is completely connected, each pair of players being linked. In the case generalized by Myerson (1977), players' prospects are totally changed if the cooperation graph is not completely connected. However, the linking game being one of perfect information, it possesses some nice features, one of them being to have subgame-perfect Nash equilibria in pure strategies (Selten 1975). A cooperation structure is then said to be "natural" if it results from a subgame-perfect Nash equilibrium of the linking game: to each such equilibrium is associated a unique (natural) cooperation graph which is a final graph of the linking game.

Here cooperation or coalition structures arise endogenously in opposition with bargaining sets and coalitional value literatures, and thus seem less arbitrarily added to the original game. Moreover, elements of negotiation and of communication are clearly treated apart from the bargaining itself: during the negotiation step (modeled as the linking game) players are not allowed to sign binding contracts. Finally, stability analysis is simplified; if no subgame-perfect equilibrium leads to an additional link to be formed, the graph of the game is said to be stable. However, Aumann and Myerson's result does not take into consideration coalitions that are not "internally connected" (Myerson 1991: 446), i.e. those coalitions that are not connected components of the cooperation graph and consist in players coming from distinct connected components. Connected components of the cooperation graph are thus an interesting representation of Myerson's concept of "effective coalitions" in

the case of coalitional games with cooperation structure (see Myerson 1991: 447).

Let us put aside for a while endogenous formation of coalitions and look at the related problem of internally connected coalitions. Owen (1977) proposes a model in which players may belong to different active coalitions, with the restriction that these coalitions should be "nested" (i.e. ranked by inclusion). The nested set of such active coalitions, called "unions," is the "nested coalition structure": it consists in a family of finer and finer partitions of the set of players. Each union acts in the name of its members as one unique agent carrying on the bargaining among unions, one restrictive underlying assumption to this model being that "agents for different unions have equal bargaining ability, independent of the size of the union that they represent" (Myerson 1991: 451). The solution to this model, the Owen value, is defined axiomatically as an extension of the Shapley value to TU[12] coalitional games with systems of unions: an induced "classical" coalitional game (the "quotient" game) is played by unions, utility within each union being shared among its members with respect to their abilities of entering other unions. In a paper (Vasquez-Brage *et al.* 1996), both directions of graph restrictions (Myerson) and of unions (Owen) have been merged and give rise to a new "allocation rule[13] for graph-restricted games with systems of unions."

A slightly different literature analyzes the constraints on coalition formation resulting from a "hierarchical organization structure" among the players, as encountered in the theory of the firm. One also speaks of permission structures because each player has to get permission for her/his actions either from all her/his superiors (conjunctive approach), or at least from one of them (disjunctive approach). Superiors can then be said to have a veto power over the cooperation of their subordinates. With the help of techniques related to lattice theory (rather than graph theory), one studies coalitional possibilities in (N, v) games with such constraints through core-like concepts (for a recent reference and a short survey, see Derks and Gilles 1995, where some nice geometric characterizations are proved in a general case and especially in a convex case). These models show under which conditions (on v, on formable coalitions) one is led to infinite exploitation or limited exploitation of subordinates.

Finally, we can go back to our previous problem and try to define, in a rather approximate axiomatic way, what should be the conditions a solution concept should verify in order to model the endogenous formation of coalitions. Here is the answer given by Zhou (1994: 513), an adequate concept requiring the following three properties: (1) not being a priori defined for any payoff vector of a given coalition structure, (2) always selecting a non-empty set of payoff vectors for some coalition structures, and (3) not systematically containing payoff vectors for all coalition structures.[14] On the basis of these three requirements, Zhou proposes a new bargaining set whose distinctive features are to be constructed apart from any particular given coalition structure, to define objections and counter-objections through coalitions instead

of players, and especially to call for the non-empty intersection of original objecting coalition S and counter-objecting coalition T. These new features enable the use of Zhou's bargaining set as a tool for modeling the endogenous formation of coalitions, in addition to its payoff distribution properties. Other major developments towards an endogenous treatment of coalition formation include Hart and Kurz (1983), Bennett (1983), and Kahan and Rapoport (1984).

Non-cooperative treatments of coalition

Recognizing the fact that coalitions can form and effectively operate in non-cooperative environments, i.e. without allowing players to sign binding agreements, has led Aumann, as early as 1959, to define the strong Nash equilibrium concept (Aumann 1959). A Nash equilibrium is strong if and only if no coalition can correlate (on the coalition's strategy) to deviate from it so that all its members are in a better position after deviation, the strategy of the complementary coalition being given. In this model, the ability of correlation of players is at the origin of formation of coalitions. But, as Bernheim, Peleg, and Whinston pointed out, this ability turns out to be "in fact a complete freedom" (Bernheim *et al.* 1987: 3), since no further restriction is imposed on the deviating coalition.

In a non-cooperative environment where players are free to communicate their strategies to each other, the "coalition-proof Nash equilibrium" (Bernheim *et al.* 1987) tries to take into account the effects of deviating coalitions on Nash equilibrium through a recurring definition (the recurrence is stated in terms of the size of coalitions). The central point is to impose constraints on deviating coalitions, similar to those imposed on players in Nash equilibrium: valid deviations must be Pareto optimal among "self-enforcing" agreements (in the restricted sense[15] that no proper subcoalition can agree to deviate to a mutually better position). Coalition-proofness ideas have been applied in the context of games with cooperation structure (Ferreira 1996). Thus both directions of research are merged in order to build a new concept, deviations being valid if they are Pareto-optimal among self-enforcing agreements (as in coalition-proof Nash equilibria), but for coalitions of connected players only (forming a cooperation structure "à la Myerson").

However, as Bernheim, Peleg, and Whinston noticed (Bernheim *et al.* 1987: 3, n. 2), coalition-proof Nash equilibrium cannot take into account situations where a member of a deviating coalition could agree on a further deviation with a player alien to this coalition: when the original coalition deviates, only subcoalitions of that coalition may further deviate. Recent models (see Greenberg 1994: 1330) that allow for a subcoalition, say T, of a deviating coalition S, to collude with a subcoalition, say Q, of the complementary coalition N\S, yield results that strongly depend on assumptions concerning the information known by Q about T (which may or may not be common knowledge, for instance).

We have supposed that any coalition forms a priori on the initiative of its own members, so that the share of the coalition's value each player gets from its membership corresponds to a solution concept in cooperative game theory. If one gives up this assumption (Pérez-Castrillo 1994), one may wish to analyze non-cooperative situations, where one distinguishes players who have the ability to form a coalition, called principals, and players who have not, called agents (this by analogy with principal–agent literature, although a major distinction here is that we stay in a complete information setting whereas principal–agent models deal mostly with games with incomplete information). Here, coalition structure is viewed as the endogenous result of competition between principals. Pérez-Castrillo (1994) shows the equivalence between the set of subgame-perfect Nash equilibria of a non-cooperative game (provided that principals are numerous enough to ensure competition) and the stable solution[16] of a cooperative game,[17] in which coalitions are formed on the initiative of their own members.

Literature concerning non-cooperative implementation of cooperative solution concepts, essentially the Shapley value and the core (for instance, we were concerned in the previous paragraph with implementation of the stable solution through subgame perfect Nash equilibria) aims to show under which conditions one can consider cooperative payoffs, which generally include in their axiomatic definitions some fairness or equity principles, as the results of a decentralized non-cooperative (in extensive or strategic form) process. Some recent results in this area include Moldovanu (1992), Moldovanu and Winter (1992), and Perry and Reny (1994).

Conclusion

In conclusion, we have seen that there is no one satisficing answer to the two issues of sharing the coalition's worth and finding stable coalitions. Forming a coalition has not the same meaning, depending on the literature and the model we look at, results of the coalition formation game being often bound to the communication structure considered in the game. However, four major ideas emerged from cooperative game theory in order to understand the role of coalition: domination (in the vN&M solution), blocking (in the core), objection and coalition structure (in bargaining sets) and excess of a coalition (in the kernel and the nucleolus). The idea of "justified objection" is particularly interesting because it allows for modeling a notion of credible threat at coalitional level, close to the one currently used at individual level in non-cooperative game theory. The range of game-theoretic models aiming at answering coalitional issues seems apparently wide, and suggested solutions, heterogeneous: for instance, cooperative games in coalitional form (with or without coalition structure) and bargaining sets or coalitional values, non-cooperative games in strategic or extensive form[18] and coalition-proof Nash equilibria.

However, recent research enables us to bring back some unity to results on

coalition with the help of abstract games[19] theory, originally stated in Von Neumann and Morgenstern (1944) and further generalized by Greenberg (1990). These new developments highlight the extremely general role and power of domination and stable set: cooperative solution concepts (vN&M solution, bargaining sets, core) are viewed as stable sets of peculiar abstract games, the features of the negotiation processes relative to those concepts are then precisely stated and easily compared within the same format. As for non-cooperative games, semi-stable theory (see Roth 1976; Asilis and Kahn 1991) allows for a new characterization of non-cooperative equilibrium concepts (coalition-proof Nash equilibria, renegotiation-proof equilibria), especially in infinite games. These results are more than a mere "stable set implementation theory": they build a new and interesting bridge between apparently disconnected fields and concepts.

Notes

1 For instance, Kurz 1988, views as a natural requisite that solution concepts used in the two games should be identical, and in that case, to the Shapley value.
2 Vannucci, 1996, presents even results concerning "implementation of vN&M stable sets through recurrent sets," that are at the heart of evolutionary game theory solution concepts.
3 Graph theory and game theory intersect in many other ways; the most classical example is found in games in extensive form where game-theoretic representation makes use of trees since Kuhn (1953). Since Berge (1962), it has also been recognized that "kernel" or "semi-kernel" concepts of oriented graph theory are in strong connection with stable sets of game theory.
4 Coalition is precisely defined as a non-empty subset S of the set of players N.
5 Von Neumann and Morgenstern (1944) consider a strategic-form game where each player chooses the coalition he wants to belong to.
6 One can at least say that lack of super-additivity sometimes explains defaults of communication between coalitions, since nothing is distributed across coalitions (additive case, AD value, discriminant solution), nothing is to be bargained on. The reverse proposition seems suspect, there could well be other ways to model informational asymmetries within the framework of superadditive games. See Aumann and Drèze 1974, for an exhaustive discussion on this point.
7 For a finer interpretation of the role played by ψ-stability theory in terms of "individually stable" coalition structures, see Greenberg (1994: 1313).
8 There are different definitions of bargaining sets leading sometimes to slightly different results, each has its particular name, and one should rather speak of a *family* of solution concepts when dealing with bargaining sets. The one we chose is the most common in the literature and is called $M_{i,j}$. In their original paper, Aumann and Maschler (1964) did consider alternative definitions, and, for instance, objections made by several players and not just one single player at a time. See for example Zhou 1994, for an alternative definition.
9 We will again find later the idea of prescribing constraints on deviating coalitions in "coalition-proof Nash equilibrium" in the context of non-cooperative games.
10 Strictly speaking, one should not use here the term "balanced" in order to avoid confusion with the major concept of "balancedness," prooved by Bondareva, Shapley, and Scarf to be decisive in order to show non-emptiness of the core.
11 Greenberg (1994: 1327) gives a list of such works.

12 TU is for "transferable utility." A major dichotomy in coalitional-form games is made between games where players may offer side-payments as a part of their payoffs to other players, and hence called TU-games and games where players are not able to make such offers, i.e. non-transferable utility games. Other tools and techniques are then needed: v is no longer a function but a correspondence, and v(S) is no longer a real but a closed convex subset of R^S. See Weber (1994) for a first approach and a short survey.
13 Shapley value, Myerson value, Owen value and generally cooperative solution concepts are "allocation rules" since they answer the problem of sharing the worth of N.
14 Although properties 2 and 3, as Zhou notices, seem antagonistic, too strong conditions for a solution concept may lead to its non-existence in some cases, as for the core; and too weak conditions, constant results independent of the coalition structure, as in Aumann–Maschler bargaining sets, which does not seem desirable for a solution concept.
15 One must not confuse this definition with the *false* (see Aumann 1990) idea according to which Nash equilibria should always be self-enforcing.
16 The "stable solution," in this context, is a generalization of the core to games that are not necessarily superadditive.
17 In fact, two cooperative games are considered in the paper, and the equivalence of the non-cooperative game to each of them is demonstrated.
18 Bernheim *et al*. (1987: 8–11) then define the concept of "perfectly coalition-proof Nash equilbrium." See Einy and Peleg (1992) for an alternative solution concept, namely "the communication proof equilibrium," in the framework of extensive-form games where restrictions are put on information available within coalitions.
19 An abstract game (A, R) is a pair consisting in a set A of objects (for instance, imputations in coalitional-form games; strategy profiles in strategic-form games or triples consisting in a strategy, a coalition and a history in repeated games with coalition structures, see Asilis and Kahn 1991) together with a relation R on this set, say a dominance relation. The vN&M solution is nothing else than the stable set of the abstract game consisting in the set A=I of imputations together with the dominance relation defined above.

References

Amer, R. and Carreras, F. (1995) "Games and cooperation indices," *International Journal of Game Theory* 3, 239–58.

Arce, M. (1995) "Social norms and core outcomes in a sharecropping economy," *Journal of Economics* 2, 61, 175–83.

Asilis, C. and Kahn, C. M. (1991) "Semi-stability in game theory: a survey of ugliness," in *Game Theory and Economic Applications*, Proceedings, New Delhi, India, December 1990, *Lecture Notes in Economics and Mathematical Systems*, B. Dutta, D. Mookherjee, T. Parthasarathy, T. S. E. Raghavan, D. Ray, and S. Tijs, eds, New York: Springler-Verlag.

Aumann, R. J. (1959) "Acceptable points in general cooperative n-person games," in *Contributions to the Theory of Games IV*, Princeton: Princeton University Press.

Aumann, R. J. (1989) *Lectures on Game Theory*, Underground Classics in Economics, Boulder–San Francisco–London: Westview Press.

Aumann, R. J. (1990) "Nash equilibria are not self-enforcing," in *Economic Decision Making: Games, Econometrics, and Optimisation: Essays in Honor of Jacques Dréze*, J. J. Gabszewicz, J.-F. Richard, and L. Wolsey, eds, pp. 201–6, Amsterdam: Elsevier Science Publishers.

Aumann, R. J. and Drèze, J. H. (1974) "Cooperative games with coalition structures," *International Journal of Game Theory* 3, 217–37.

Aumann, R. J. and Maschler, M. (1964) "The bargaining set for cooperative games," in *Advances in Game Theory*, M. Dresher, L. S. Shapley, and A. W. Tucker, eds, Annals of Mathematical Studies, 52; pp. 443–76, Princeton: Princeton University Press.

Aumann, R. J. and Myerson, R. B. (1988) "An application of the Shapley value," in *The Shapley Value*, A. E. Roth, ed., pp. 175–91, Cambridge. Cambridge University Press.

Aumann, R. J. and Shapley, L. S. (1974) *Values of Non-Atomic Games*, Princeton: Princeton University Press.

Bennett, E. (1983) "The aspiration approach to predicting coalition formation and payoff distribution in sidepayment games," *International Journal of Game Theory* 12, 1–28.

Berge, C. (1962) *Graphes et Hypergraphes*, Paris: Dunod.

Bernheim, B. D., Peleg, B., and Whinston, M. D. (1987) "Coalition-proof Nash equilibria I and II," *Journal of Economic Theory* 42, 1–29.

van den Brink, R. and Gilles, R. P. (1996) "Axiomatizations of the conjunctive permission value for games with permission structures," *Games and Economic Behavior* 12, 113–26.

Davis, M. and Maschler, M. (1963) "Existence of stable payoff configurations for cooperative games" (abstract), *Bulletin of the American Mathematical Society* 69, 106–9.

Davis, M. and Maschler, M. (1965) "The kernel of a cooperative game," *Naval Research Logistics Quarterly*, 12, 229–59.

Davis, M. and Maschler, M. (1967) "Existence of stable payoff configurations for cooperative games," in *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, M. Shubik, ed., Princeton: Princeton University Press.

Derks, J. J. M. and Gilles, R. P. (1995) "Hierarchical organization structures and constraints on coalition formation," *International Journal of Game Theory* 24, 147–63.

Driessen, T. S. H. and Tijs, S. H. (1986) "The core and the τ-value for cooperative games with coalition structures," in *Game Theory and Economic Applications*, Proceedings, New Delhi, India, December 1990, *Lecture Notes in Economics and Mathematical Systems*, B. Dutta, D. Mookherjee, T. Parthasarathy, T. S. E. Raghavan, D. Ray, and S. Tijs, eds, New York: Springer-Verlag.

Einy, E. and Peleg, B. (1992) "Communication-proof equilibria," Discussion Paper 9, July 1992, Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem.

Einy, E. and Wettstein, D. (1996) "Equivalence between bargaining sets and the core in simple games," *International Journal of Game Theory* 25, 65–71.

Ferreira, J. L. (1996) "Endogenous formation of coalitions in non-cooperative games (abstract)," Tenth Italian Congress on Game Theory and Applications, University of Bergamo, Italy, March 4–5.

Gabszewicz, J. J. and Hansen, T. (1972) "Collusion of factor owners and distribution of social output," *Journal of Economic Theory* 4, 1–18.

Greenberg, J. (1990) *The Theory of Social Situations. An Alternative Game-theoretic Approach*, Cambridge: Cambridge University Press.

Greenberg, J. (1994) "Coalition structures," in *Handbook of Game Theory* vol. II, R. J. Aumann and S. Hart, eds, chapter 37, Amsterdam: Elsevier Science B. V.

Harsanyi, J. C. (1963) "A simplified bargaining model for the n-person cooperative game," *International Economic Review* 4, 194–200.
Harsanyi, J. C. (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Institutions*, Cambridge: Cambridge University Press.
Hart, S. and Kurz, M. (1983) "Endogenous formation of coalitions," *Econometrica* 54, 50–69.
Hsiao and Raghavan (1986) "Multi-choice cooperative games," in *Game Theory and Economic Applications*, Proceedings, New Delhi, India, December 1990, *Lecture Notes in Economics and Mathematical Systems*, B. Dutta, D. Mookherjee, T. Parthasarathy, T. S. E. Raghavan, D. Ray, and S. Tijs, eds, New York: Springer-Verlag.
Kahan, J. P. and Rapoport, A. (1984) *Theories of Coalition Formation*, London: Lawrence Erlbaum Associates.
Kuhn, H. (1953) "Extensive games and the problem of information," in *Contributions to the Theory of Games*, H. Kuhn and A. Tucker, eds., Princeton: Princeton University Press, 193–216.
Kurz, M. (1988) "Coalitional value," in *The Shapley Value*, A. E. Roth, ed., chapter 11, pp. 155–73, Cambridge: Cambridge University Press.
Lucas, W. F. (1968) "A game with no solution," *Bulletin of the American Mathematical Society*, 74, 237–9.
Luce, D. and Raiffa, R. (1957) *Games and Decision*, New York: Wiley.
Maschler, M. (1992) "The bargaining set, kernel and nucleolus: a survey," in *Handbook of Game Theory* vol. I, R. J. Aumann and S. Hart, eds, Amsterdam: Elsevier Science B. V.
Moldovanu, B. (1992) "Coalition-proof Nash equilibria and the core in three-player games," *Games and Economic Behavior* 4, 565–81.
Moldovanu, B. and Winter, E. (1992) "Core implementation and increasing returns to scale for cooperation," Discussion Paper 23, Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem.
Myerson, R. B. (1977) "Graphs and cooperation in games," *Mathematics of Operations Research* 2, 225–9.
Myerson, R. B. (1986) "Negotiation in games: a theoretical overview," in *Uncertainty, Information and Communication: Essays in Honor of K. J. Arrow*, vol. III, Helleer, Starr, and Starrett, eds, chapter 1, Cambridge: Cambridge University Press.
Myerson, R. B. (1991) *Game Theory: Analysis of Conflict*, Cambridge, MA: Harvard University Press.
Nash, J. F. (1951) "Non-cooperative games," *Annals of Mathematics* 54.
Owen, G. (1977) "Values of games with a priori unions," in *Essays in Mathematical Economics and Game Theory*, R. Henn and O. Moschlin, eds, pp. 76–88, New York: Springer-Verlag.
Pérez-Castrillo, J. D. (1994) "Cooperative outcomes through noncooperative games," *Games and Economic Behavior* 7, 428–40.
Perry, M. and Reny, P. J. (1994) "A non-cooperative view of coalition formation and the core," *Econometrica*, 62, 4, 795–817.
Rapoport, A. (1990) *Experimental Studies of Interactive Decisions*, Theory and Decision Library Series C: Game Theory, Mathematical Programming and Operations Research, Kluwer Academic Publishers.
Roth, A. E. (1976) "Subsolutions and the supercore of cooperative games," *Mathematics of Operations Research* 1, 43–9.

Roth, A. E. (1988) "Introduction to the Shapley value," in *The Shapley Value*, A. E. Roth, ed., Cambridge: Cambridge University Press.
Selten, R. C. (1975) "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory* 4, 22–55.
Shapley, L. S. (1953) "A value for n-person games," in *Contributions to the Theory of Games* vol. II, H. Kuhn and A. W. Tucker, eds (Annals of Mathematical Studies, 28), Princeton: Princeton University Press.
Shenoy, P. P. (1979) "On coalition formation: a game-theoretic approach," *International Journal of Game Theory* 8, 133–64.
Vannucci, S. (1996) "Von Neumann stable sets and recurring sets," Tenth Italian Congress on Game Theory and Applications, University of Bergamo, Italy, March 4–5.
Vasquez-Brage, M., Garcia-Jurado, I., and Carrerras, F. (1996) "The Owen value applied to games with graph-restricted communication," *Games and Economic Behavior* 12, 42–53.
Von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.
Weber, R. J. (1994) "Games in coalitional form," in *Handbook of Game Theory* vol. II, R. J. Aumann and S. Hart, eds, chapter 36, Amsterdam: Elsevier Science B. V.
Xue, L. (1993) "Farsighted optimistic and conservative coalitional stability," mimeo, McGill University.
Zhou, L. (1994) "A new bargaining set for a N-person game and endogenous coalition formation," *Games and Economic Behavior* 6, 512–26.

6 Do Von Neumann and Morgenstern have heterodox followers?

Christian Schmidt

Introduction

The publication in 1944 of *Theory of Games and Economic Behavior* (*TGEB*) by Von Neumann and Morgenstern is generally regarded as the starting point of game theory. But what theory does this refer to? According to a first precise, but restrictive meaning, it is a mathematical theory whose base is made up of games known as zero-sum games. The theory can also be understood in a wider, albeit rather vague, sense. Its most extreme expression is provided by Aumann: "Game theory is a sort of umbrella or 'unified field' theory for the rational side of social science, where 'social' is interpreted broadly, to include human as well as non-human players (computers, animals, plants)" (Aumann 1987b: 2).

The first meaning meets the definition of a theoretical object, while the second aims at identifying its application field. But the identification of the application field of game theory (second meaning) includes reference to elements drawn from the definition of its theoretical object. Indeed, how can the "rational side" of social science be stressed without criteria borrowed from mathematical theory (first meaning)? Besides, there is no guarantee that there exists a perfect connection between the object of game theory in mathematical theory's narrow meaning and the "unified field" theory. The theory of zero-sum games obviously cannot on its own and without auxiliary constructions serve as an "umbrella" to a study whose ambition is, from the very beginning, to deal with all social phenomena resulting from an interaction between rational decision-makers (Von Neumann 1928, 1959: 13). Since *TGEB*, theoreticians have been led to suggest solutions to fill this gap.

As we have already shown in another paper, the theory of non-cooperative games sketched out by Nash in his four seminal papers (1950a, 1950b, 1951, 1953) marked a break with the research program assigned to cooperative game theory by Von Neumann and Morgenstern in *TGEB* (see Chapter 2, this volume). Our purpose here is to study another aspect of this issue.

For lack of a *stricto sensu* unified theory, a consensus amongst theoreticians has developed over the years around a hard core which, beyond the differences between *TGEB* and Nash's contributions, corresponds with what may be called traditional game theory. However, such apparent consensus

should not be misleading. Theoreticians still disagree on the foundations of game theory (Binmore 1990, 1992, 1993). Various propositions have been made since 1944 to build up the theory on other bases.[1] These heterodox theories have not been accepted by the majority of game theoreticians; but they include a number of concepts that are being further looked into. The purpose of the present work is to reconsider some of the above-mentioned alternatives in the light of a player's rational behavior via-à-vis his information about other players. We will show how they refer to suggestions already made by Von Neumann and Morgenstern in *TGEB*, which passed unnoticed or were purposely ignored.

Rationality and knowledge of others in *TGEB*

The meaning of the hypothesis of players' rationality in game theory is an issue which Von Neumann implicitly raised in his first demonstration of the minimax theorem (Von Neumann 1928). It re-emerged within a similar problematic in *TGEB* (1944), where Von Neumann and Morgenstern sought to draw out its implications. These two historical contributions encompass the main elements required for an understanding of a controversy which has run through the history of game theory since *TGEB*.

The nature of the problem

The 1928 paper sets forth two propositions about the players' information which at first seem somewhat incompatible. On the one hand, Von Neumann infers from his definition of the strategy concept that each player must choose his strategy while totally ignoring the strategies chosen by the other players (Von Neumann 1959 (1928): 19). He assumes that all the available information about the players' strategies is indeed already included in this strategy's definition. On the other hand, as seen from the players' viewpoint, the mathematical identity between maximum and minimum values requires the introduction of each player's knowledge about the other's strategies (Von Neumann, 1959 (1928): 23). As he formulates this latter constraint, Von Neumann realizes that such knowledge does not provide the player with the same type of information as that provided by the rules of the game. However, he does not suggest any means for reconciling these two hypotheses on the players' information. According to him, this situation is related to the way in which the game, as defined by the theory, incorporates uncertainty (Von Neumann 1959 (1928): 26). In a more contemporary wording, one would say that players' expectations are endogenous in game theory. For Von Neumann, if there is a problem, the solution should be sought by going deeper into the game's rules.

In 1944 *TGEB*'s authors are confronted with the same problem, and they connect it, this time explicitly, with the question of rationality. An immediate interpretation of the game's maximum and minimum values can be provided

in the case of a two-person zero-sum game where these values express maximization (vs. minimization) of each player. But this does not permit an understanding of how the game works. The analogy with the 1928 paper is given by the following quotation:

> Observe that from the point of view of the player I who chooses a variable way τ_1, the other variable can certainly not be considered as a chance event. The other variable, in this case τ_2, is dependent upon the will of the other player, which must be regarded in the same light of "rationality" as one's own.
>
> (*TGEB*: 99)

Each player decides upon his strategy while ignoring the other's, which now means that the players' rationality, expressed by the choice of their maximizing strategies, should not depend on the knowledge of the (also rational) strategy of the other player. At the same time, however, the game's solution, as viewed by Von Neumann and Morgenstern, is accessible to the players only if each of them has some information about the strategy that the other has chosen, i.e. about its rationality, since the strategy is the result of a rational choice. The shift of the players' information onto the players' rationality raises a number of issues; these account for the introduction of additional hypotheses in the reasoning process in *TGEB*.

> Indeed, the rules of the game prescribe that each player must make his choice (his personal move) in ignorance of the outcome of the choice of his adversary. It is nevertheless conceivable that one of the players, say 2, "finds out" his adversary; i.e. that he has somehow acquired the knowledge as to what his adversary's strategy is. The basis for this knowledge does not concern us; it may (but need not) be experience from previous plays. At any rate, we assume that the player 2 possesses this knowledge. It is possible, of course, that in this situation 1 will change his strategy; but again let us assume that, for any reason whatever, he does not do it. Under these assumptions we may then say that player 2 has "found out" his adversary.
>
> (*TGEB*: 105)

The definition of the players' rationality exclusively falls within the game's rules. It does not depend on the knowledge that a player may have about the other's strategy, because the zero-sum two-person game, as analyzed in *TGEB* is a simultaneous one-shot game. For the same reason, the conditions under which such knowledge is acquired are viewed as being outside of the game. However, the fact of possessing such information does have some impact on the rational choice of the player's strategy. Thus the way in which the game will develop will differ according to whether, for instance, player 2 knows player 1's strategy, while player 1 is ignorant of player 2's strategy, and

vice versa (see minorant and majorant games, *TGEB*: 100–1). In another respect, such extra knowledge which player 2 has in the first case will be an advantage for him only if player 1 does not change his strategy. Now, whether player 1 maintains or changes his strategy depends on his rational choice. It is difficult in those conditions for player 1's rationality to be independent from his possibly knowing that player 2 has "found him out."

Toward solving the problem

The path suggested in *TGEB* to answer the question within the limits of a two-person zero-sum game goes through three stages: construction of the majorant and the minorant games; use of the symmetry hypothesis; normative interpretation of the theory as a knowledge shared between players.

*Auxiliary minorant and majorant games (*TGEB*: 100, 106–7, 149).*

Two auxiliary games Γ_1 and Γ_2 are added to the initial game Γ. Γ_1 and Γ_2 are identical to Γ except on one point. In Γ_1, player 2 knows player 1's strategy when he chooses his own. In Γ_2, it is player 1 who knows player 2's strategy when he chooses his. In Γ, as long as the game is in its normal form, neither player knows the other's strategy when he makes his choice. Γ_1 and Γ_2 can be interpreted by taking either player's viewpoint.

The construction of games Γ_1 and Γ_2 enables us to pinpoint what makes the difference between the player's rationality when it is inferred from the abstract formulation of the game (Γ) and when it applies to the player as he makes his choice (Γ_1 and Γ_2). Let us consider V, V_1, and V_2 as the values of games Γ, Γ_1, and Γ_2 respectively, $V_1 \leq V \leq V_2$. The issue raised by the interpretation of the players' rationality is thus led back to the mathematical determination of the game, in accordance with the method favored in *TGEB*.

The construction of auxiliary games gave rise to various uses. The theory of metagames developed by Howard (1970) originated from it.

Aumann also made use of a variation of this procedure as the basis for his handling of irrationality in game theory (Aumann 1992).

*The symmetry hypothesis (*TGEB*: 109–10, 165)*

The construction of auxiliary games Γ_1 and Γ_2 is based on the introduction of asymmetric information between players 1 and 2. The results obtained should not differ from those of the players who benefit from such asymmetry. The symmetry hypothesis therefore means that only players' roles differ in each auxiliary game, but not the player who plays the role.

The symmetry hypothesis transforms the two formulations of rationality into an information issue, in accordance with the problematic raised in Von Neumann's 1928 paper. It is obvious in the *TGEB* perspective when applied

to a two-person zero-sum game. It looks more stringent when the hypothesis is applied to less restrictive games.

Two distinctive symmetry hypotheses are set out in *TGEB*. The first one, which will be qualified as "weak," is derived from the initial definition of the zero-sum game. It establishes the possibility of interchanging players within the game without modifying the payoff functions. The second one, which will be qualified as "strong," consists in making the interchange of players' roles feasible and it does affect the payoff functions (*TGEB*: 109, n. 1). Only weak symmetry is used in the construction of Γ_1 and Γ_2. The strong symmetry hypothesis appears only in Γ.

Nash used both symmetry hypotheses. He extends the first hypothesis to non-cooperative, *n*-person games (Nash 1951). He utilizes the second hypothesis in his model of negotiation (Nash 1950a, 1953). He hesitates as to what interpretation should be given the weak hypothesis with regard to each player's behavior, before finally adhering to Von Neumann and Morgenstern's in *TGEB*. As for the strong symmetry hypothesis, included in the solution of the negotiation game, it means that the only difference between the players lies in the information contained in the mathematical description of the game (Nash 1953).

Neither Von Neumann and Morgenstern nor Nash made any further comment about the implications of the strong symmetry hypothesis on the players' behavior. It was Harsanyi who connected it to the definition of the player's rationality in the formulation of negotiation games which he developed on the basis of the similarity he worked out between Nash's negotiation model and that of Zeuthen (Harsanyi 1956). The strong symmetry hypothesis implies that each player in the two-person negotiation game behaves in exactly the same manner as the other if roles were interchanged. Abandoning the symmetry hypothesis would correspond to something irrational in their behavior.

The understanding of the strong symmetry hypothesis has been suggested by Schelling, who based his criticisms on the example of the dollar-sharing game (Schelling 1959, quoted by Schelling 1960, Appendix b). Schelling's criticisms led Harsanyi to stop defending the symmetry hypothesis on the basis of the players' rationality (Harsanyi 1961). Harsanyi later suggested a somewhat different interpretation: the basis for the symmetry hypothesis might not be directly derived from the rational behavior of maximizing players: it should be sought in the capacity of the theory, which incorporates symmetry, to determine the solution of the negotiation game (Harsanyi 1977, n. 12).

*Players' knowledge of game theory (*TGEB*: 148)*

If interpreted in a normative way, the theory of two-person zero-sum games provides each player with a behavior rule. Such prescriptions exactly express the rational behavior of the players in the theory. Besides, each player knows

that his adversary also knows the theory. He is therefore able to "find out" the other's rational strategy while ignoring the strategy which the other has effectively decided upon (and vice versa as per the symmetry concept). If the "found out" player were to change his strategy, he would behave in a way that would then contradict the theory's prescriptions. Assuming that the theory is complete, the knowledge of it shared by both players thus allows for a reconciliation between the contradictory indications according to which each player should know about the other's strategy in order to behave rationally in the course of the game. One player's complete ignorance of the strategy chosen by his adversary becomes compatible with his abstract knowledge of the strategy recommended to the other player by the theory. This has served as a basis and has been developed, especially by Selten, to show that in non-cooperative games the player only has to know that the knowledge of the game theory is shared with the other players, to be able to predict reliably the strategies chosen by the other players (Selten 1978; Selten and Leopold 1982).

However, while this interpretation is in *TGEB*[1] the hypothesis of a knowledge of the theory shared between players is not used to this end by Von Neumann and Morgenstern. It is mentioned in *TGEB* as a support to the theory. What the authors discuss is whether it gives an epistemological basis which suffices to found the validity of game theory whose mathematical structure they establish, i.e. the zero-sum game theory. In the authors' view, this heuristic procedure only provides an "indirect" argument that allows for proving the logical existence of the theory, but is not sufficient to guarantee that the thus obtained theory is the one which was looked for (*TGEB*: 48, n. 5). According to Von Neumann and Morgenstern, game theory should be established on a mathematical basis that should be independent from the heuristic procedures which have served to elaborate it (minorant and majorant games, knowledge of the theory by the players, etc.).

Suspicion about the "indirect argument" provided by the player's knowledge of the theory can be related to the self-realization device that threatens every theory which describes phenomena generated by expectations based on a shared belief in this theory. Morgenstern had already pointed out such a risk before *TGEB* (Morgenstern 1934). A "fanciful theory" whose rationale would be shared by both players could be fallaciously legitimated. However, this heuristic argument is where the Nash equilibrium is derived from. Without knowing the other players' strategies, each player knows, through the theory and through the fact that the others know it too, that his best strategic response corresponds to equilibrium. Von Neumann and Morgenstern's "indirect argument" criticism therefore anticipates a weakness in the research program which has been developed on the basis of the Nash theory of equilibrium.[2]

Such criticism can be understood in different ways. The argument inferred from the players' shared knowledge of the theory is only applicable to a *complete* theory. This means that the theory should contain a unique prescription for the players, whatever the stage of the game they are in. This is

not the case in many non-cooperative games, since more than one equilibria exist and the theory provides no information whatever that would allow the players to choose between them. Such insufficiency in Nash's program can be interpreted in Von Neumann and Morgenstern's terms by observing that the procedure followed by Nash has failed to determine the "satisfactory" theory and has only allowed for the identification of a logically plausible group of theories. This is how most game theoreticians have understood it when they sought to give a more refined definition of the Nash equilibrium in order to fill in this gap, by taking different paths (Selten's "perfect" equilibrium (1975), Kreps and Wilson's "sequential" equilibrium (1982) . . . for extensive form games; Meyerson's "correct" equilibrium (1978); Kalai and Samet's "persistent" equilibrium (1984); and Kohlberg and Mertens' "strategically stable" equilibrium (1986) . . . for normal form games). They lead to a plurality of non-cooperative game theories.

Von Neumann and Morgenstern's criticisms can also be interpreted in a more radical way. If the knowledge shared between players of a supposedly complete theory is not sufficient to demonstrate its legitimacy, there may be more serious reasons behind it. Let us remember that such a procedure tends to reduce all the information which each player might (or might not) have down to the sole prescriptions formulated by the theory. Then, the player's lack of information about the strategy effectively chosen by his adversary is offset by the knowledge of the other's rationality via the hypothesis of the theory as a common knowledge between players; this boils down to admitting that the players' rationality can be entirely reduced to the rationality of the theory's modeler. Such conjecture was first questioned by Schelling on the basis of counter-examples of "pure coordination" games (Schelling 1960: Appendix c). This leads Schelling to suggest a different interpretation of the solution of a game in terms of the equilibrium drawn from focal points.

Evaluation and criticism

These three stages of reasoning for clarifying the problem raised by the players' information about the others are used in *TGEB* to accompany the new demonstration of the minimax theorem. Since for Von Neumann and Morgenstern this "fundamental theorem" is at the core of the theory, their respective importance is related to its demonstration.

While the construction of both minorant and majorant games means nothing more than a heuristic procedure that is logically independent from the theory itself, the symmetry hypothesis is a crucial element. Von Neumann and Morgenstern even suggest that it could serve as a basic axiom to the theory (*TGEB*: 168).

As for the knowledge of the theory shared by the players, its status is less obvious. On one hand, Von Neumann and Morgenstern like to stress that their demonstration of the minimax theorem is free of any kind of such hypothesis.

> There is nothing heuristic or uncertain about the entire argumentation. We have made no extra hypotheses about the intelligence of the players, about "who has found out whose strategy" etc. Nor are our results for one player based upon any belief in the rational conduct of the other – a point the importance of which we have repeatedly stressed.
>
> (*TGEB*: 160)

On the other hand, they acknowledge that for a player there is no strategy that would be optimal in any case (*TGEB*). If, for instance, one of the players chooses a strategy that moves away from the theory's prescriptions, the strategy prescribed by the theory to the other player is not always the best one for him. Von Neumann and Morgenstern illustrate this via a number of games ("stone, paper, scissors," matching pennies, etc.). The problem disappears when such a situation is viewed as resulting from a mistake by the first player due to his lack of knowledge of the theory. This is obviously the case if it is assumed from the start – i.e. before the game begins – that both players have a perfect knowledge of the theory and that each knows that the other knows it. Indeed, the extra-mathematical hypothesis of the knowledge of the theory shared between players does not intervene in the demonstration of the minimax theorem, but it can be useful to delineate its field of interpretation.

Dissident interpretations and heterodox research programs

The issue raised in *TGEB* is worsened in the case of non-cooperative games with a Nash equilibrium solution. According to this theory, each player must choose his best strategy in response to the others without knowing what the latter have actually chosen. Such a prescription is not applicable in every case. Let us consider for example two situations: (1) when the game has more than one equilibrium, none of which corresponds to a set of dominant strategies; (2) when, for some reason, one of the players moves away from the rational behavior prescribed by the theory. In both cases, players cannot identify their best strategic responses without complementary information on the strategies chosen by the others.

For most theoreticians until recently, this deficiency was the consequence of the provisional incompleteness of the Nash equilibrium theory. They have attempted to remedy it by narrowing the link between the interpretation of the theory and the Bayesian hypothesis of individual choices under uncertainty. Harsanyi and Selten derived from a Bayesian treatment of each player's anticipations of the others' behaviors, the "risk dominant" criterion, which should enable the players to select one equilibrium in (1) (Harsanyi 1975; Harsanyi and Selten 1988). Aumann, by working on his "correlate equilibrium" concept (Aumann 1987a) showed that it is possible to introduce one player's irrational behavior into a possible state of the game, provided it is supposed that all players share a Bayesian understanding of the prior

distribution of probabilities which each player assumes before the game starts (Aumann 1988).[3]

Another diagnosis has been established by a few researchers. According to them, such a gap in the theory cannot be filled by adding Bayesian hypotheses. They recommend a deeper change in what the majority's tradition considers to be game theory. Their viewpoint may be qualified as heterodox. We shall now examine two illustrations: Howard's metagames theory and Schelling's focal points theory. The former has a direct connection with *TGEB* as it develops the auxiliary games concept imagined by Von Neumann and Morgenstern in *TGEB*. The latter's filiation with *TGEB* is less evident. On one hand, Schelling is skeptical about a general game theory project, reducible to those fundamental properties derived from individual rationality, in the way in which Von Neumann and Morgenstern conceived their zero-sum two-person game theory. On the other hand, several objections against the inadequacies of the orthodox theory of non-cooperative games meet, through a different path, those already formulated in *TGEB*. More generally, Von Neumann and Morgenstern's doubts and warnings concerning the difficult relationship between the solution of a game and the rational standards of players' behaviors should be recalled:

> the rules of rational behavior must provide definitely for the possibility of irrational conduct on the part of others . . . In whatever way we formulate the guiding principles and the objective justification of "rational behavior," provisos will have to be made for every possible conduct of "the others." Only in this way can a satisfactory and exhaustive theory be developed. But if the superiority of "rational behavior" over any other kind is to be established, then its description must include rules of conduct for all conceivable situations – including those where "the others" behaved irrationally, in the sense of the standards which the theory will set for them.
>
> (*TGEB*: 32)

Howard's metagames

Every game can be interpreted in two ways. One of them gives a detailed description of the sequences in the players' decisions starting from the initial up to the terminal stage of the game ("extensive" form). The other one only defines the functional relation between players' choices and the pay-off values ("normal" form). The extensive form provides information on the game which the normal form does not indicate. The manner in which the information on the game is understood by the players affects it. Assuming that players know the game and that they share the knowledge of its source, the situation is not the same according to whether it is its extensive form or its normal form which is known by the players. When it is the extensive form that the players know, each of them can in principle know the others' previous

choices at the time when he must make his own ("perfect recall"). In the case of a perfect information game, the last player to make his move knows all the strategies which have already been chosen by the others. When players know only the normal form of the game, each must choose his strategy while ignoring the others'. The situation is for them as if they played simultaneously. It suggests some connection between, on the one hand, the contradictory hypotheses about the knowledge of the others' strategies in the definition of each player's rational behavior and, on the other hand, the representation of the game, the knowledge of which is shared by the players. Analyzing this connection is the starting point of Howard's propositions to solve the problem (Howard 1971: 11–23).

Both the extensive and the normal forms are extreme representations of the game and one can pass from one to the other. While a game in its extensive form possesses one and only one normal form, a normal form may belong to several different games. But they do not exhaust the possible representations of the game, and theoreticians have suggested various intermediate representations between these two extreme ones.[4] Howard's metagames are representations of this kind.

Between normal and extensive form

Starting from a game in its normal form, Howard selects, amongst all games of extensive form with the same normal form, one specific game for each player. This game has a strategic interest for the players because it conveys the knowledge of all the strategies chosen by the other players. Howard's metagames must be seen as a generalization of the construction of the majorant game in *TGEB*. The general approach is as follows: to each *n*-person game there correspond *n* metagames which are constructed on the basis of *TGEB*'s majorant games. Those *n* metagames can be represented by a unique *n k*-person game, where each *k* player has the knowledge of the choices made by each of the other players. Such a procedure, which refers to a level-1 knowledge of the strategic choices of the other players, can be worked further by constructing the level-2 metagames (the knowledge of the level-1 knowledge), and so on. In other words, the construction procedure of auxiliary games as implemented in *TGEB* can be worked down ad infinitum (Howard 1970, 1971: 55). It enables the players to have access to a level of information possessed by the observer.

Let J be a 1, 2, . . . *k*, . . ., *m*-person game. All players know its normal form N(J). In the set of games in extensive form with the same normal form, there is a game for each player which is particularly useful to him, i.e. the game in which he is the last to make his move while fully knowing about the others' strategies. The kJ metagames can thus be derived from the initial game J. Each kJ metagame corresponds to a game different from J which can be represented in a normal form. To do this, it suffices to replace the space of 1, 2, . . ., *k* players in the initial game by an F set which contains all those *f*

functions that link the S$m-k$ other players' strategies to Sk. k, a player in the J game, is both a player in k J and an observer of J in k J. The duality of player k's functions in k J enables one to overstep the contradicting hypotheses concerning player k's knowledge about the other player's strategies, as it appears in the orthodox game theory.

The construction of metagames is a technical process which was not first implemented by Howard (see majorant game in *TGEB*). It has since been often used by orthodox theoreticians. It can be noticed that Aumann implicitly refers to it when he introduces irrationality in game theory (Aumann 1988). Aumann models the information system associated with the players under the form of a metagame. In his two-person game example, each auxiliary game reflects two separate states of information for each player over the other (Aumann 1988: 217–20). The novelty in Howard's contribution does not lie in his technics of metagames, even generalized, but in a more ambitious project whose aim is to change game theory by replacing it by a general theory of metagames (Howard 1987).

For Howard, contradictory hypotheses as formulated by game theory in relation to a rational player's knowledge about the others' strategies, are but one of the signs of the problems that have to be faced while defining the player's rational behavior in traditional game theory. According to him, such problems result from an inadequacy in the very definition of a game in the theory. Not only can the game be represented in many more ways than its normal and extensive forms, but above all the player can construct his own representation(s) on the basis of some objective knowledge of the game's data and rules. Obstacles concerning the player's rationality arise from the fact that in game theory the representations of the game are assumed to be common to the theoretician and to the players. By rejecting such simplifications, Howard's purpose is to construct another theory. In Howard's theory, each player has a specific set of metagames at his disposal, which can be derived from the information contained in the initial game's normal form. Players decide upon their strategies on the basis of their sets of metagames.

Theory and interpretation

Howard's metagame theory has two parts: a mathematical part, which develops the formal procedure of metagames' construction on the basis of an n-person initial game; and a prescriptive part which guides each player both in selecting one metagame amongst all his possible metagames and in choosing his rational strategy in this selected metagame.

The structure of the mathematical theory of metagames is very simple. An intuitive representation can be as follows. Let J be a game; it has been shown above how to construct those k J metagames in association with k players. The procedure is also applicable to k J metagames by constructing those $i\,k$ J metagames of k J, where i stands for the nth player in k J, and so on ad infinitum. All such transformations are arranged by levels from the initial

game J. The hierarchically organized infinite set of all metagames make up the mathematical object of metagames theory. Howard sums it up by the formula of "infinite metagame tree" (Howard 1971: 55).

Howard only sketched out the second part of the theory. As in traditional game theory, players are omniscient. But omniscience in the metagame theory only lies in each player's knowledge of the subjective infinite set of all his metagames. On the other hand, players do not necessarily possess a complete knowledge of the others' sets of metagames. Therefore they can be misled in their expectations of the others' strategic choices, even when they know that the other players are rational.

Up to what level can players further work into the construction procedure of metagames to rationally decide upon their strategies? Howard showed that in most cases, players can choose their strategies on the basis of information contained in their level-1 metagames. However, in some cases, the information conveyed by a higher-level metagame can be instructive for the players. Thus, in the prisoner's dilemma game, one must wait to reach level-2 metagames before rational outcomes appear whose payoff values correspond to a cooperative equilibrium solution (Howard 1971: 58–60). Such property is not surprising as metagames of a higher than 1 level can be regarded as mental representations of the repeated initial game.

For Howard, the mathematical construction of metagames leads to a subjective theory to be applied to a qualitative game. This has the merit of taking into account the difference between the player's knowledge and the observer's (or modeler's) knowledge. It opens a path to a subjective treatment of the player's rationality which is not based on a Bayesian-like hypothesis. By combining the metagames procedure and the choice of a qualitative method, Howard extends his theory to incomplete information metagames that result from each player's partial or complete ignorance of the other players' preferences. More accurately, each player has a belief about others' preferences, but this belief is not necessarily true. The player thereby makes a rational choice while ignoring the true game in which he is playing. Such an approach to the players' incomplete information gave rise to a number of developments, especially through the somewhat similar concept of hyper-games (Schmidt 1994).

The main merit of metagame theory is to reverse the traditional relation between the modeler's and the player's viewpoint on the game. In metagame theory, the player is not supposed to possess the knowledge of the game which the modeler has; instead, the knowledge which the players may have about the game is given to the modeler. However, the development of metagame theory raises a number of difficulties. Each player defines his rational strategy within the subjective set of his metagames and possesses only one belief, itself subjective, on the other players' preferences. There is no guarantee that the sets of those metagames always possess a non-empty intersection. On the other hand, one of the advantages of metagame theory lies in the way in which it handles incomplete information. Incomplete information means here that players may be misinformed about the initial game, i.e. that they

may not know the "actual" game in which they are playing. Unfortunately, metagame theory does not provide for the analysis of the coherence of all such choices. Anyway, Howard's approach to metagames has opened up a new way of linking game theory to knowledge investigation.

Strategies' coordination and focal points in Schelling

Schelling also refers to Von Neumann and Morgenstern's majorant and minorant games, but the observation he derives from them is different. In a zero-sum game, the information about other players possessed by the player who moves last always provides him with an advantage which enables him to make a maximum gain (majorant game). This is no longer necessarily the case in non-zero-sum games. This is shown in non-cooperative games, where the player who moves first while ignoring the others' choices makes a higher gain. Two-person non-cooperative games with two Nash equilibria when the players' interests are diverging provide traditional illustrations ("chicken game," the "battle of the sexes," etc.). Schelling remarks that in such situations it is often more advantageous for a player, in choosing his strategy, to place himself in his minorant game (Schelling 1960: 161).

Strategic information

Schelling's shortcut is an invitation to work further on his analysis. The construction of the majorant metagame in a game of which only the normal form is known by the players, does not inform the player about the strategy which has actually been decided upon by the other players. The complementary information he thus acquires on the game takes the form of an extension of the possible states of the game. The utility for the player to whom these complementary possibilities are to be revealed varies according to the game's structure, i.e. to players' specific configurations of interests (or preferences). As he moves first, the player gives information to the others on the strategy he has actually chosen. This information reduces the others' uncertainty. Its final impact is to be measured not only from the point of view of the players who receive this information but also from that of the player who gives it. Indeed, nothing stops the player who moves first from anticipating the consequences of such information on the others' strategies. The player who decides to move first also uses the metagame angle. His advantage lies in that he is the first to know the strategy actually chosen by the player who moves first, i.e. himself. According to the game's structure, this information may or may not benefit the player who controls it.

Between the potential information provided by metagames and the actual information provided by the player who moves first, there is some space available in the players' strategies for intermediate types of information, given by a player to another player. Instead of moving first, the player can make a conditional bid concerning his strategic choice. Let us take a two-person

non-cooperative game. Player 2 will, for example, inform player 1 that "if player 1 did not play his strategy A, then he, player 2, would play b." This type of offer corresponds to what, since Lewis (1978) and other logicians have called "counterfactuals."[5] The player who makes this bid, as well as the one who moves first, provides information on his strategic choice to the other player. But the nature of this information is different. It results from a mental projection where the one who makes this announcement puts himself in the position of the player who would move last, as in metagames. When this information is conveyed to the other player, it is potential and can become effective only if/when the strategy actually chosen by the other player is actually displayed.

Figure 6.1 illustrates the connections between three types of information on the strategies of others and their impact on the player' rational choices. By moving first, player 1 provides no useful information to player 2, whose strategy is dominant. Player 2 therefore rationally chooses his strategy while ignoring player 1's selected strategy. The same goes for player 1. The prescription of

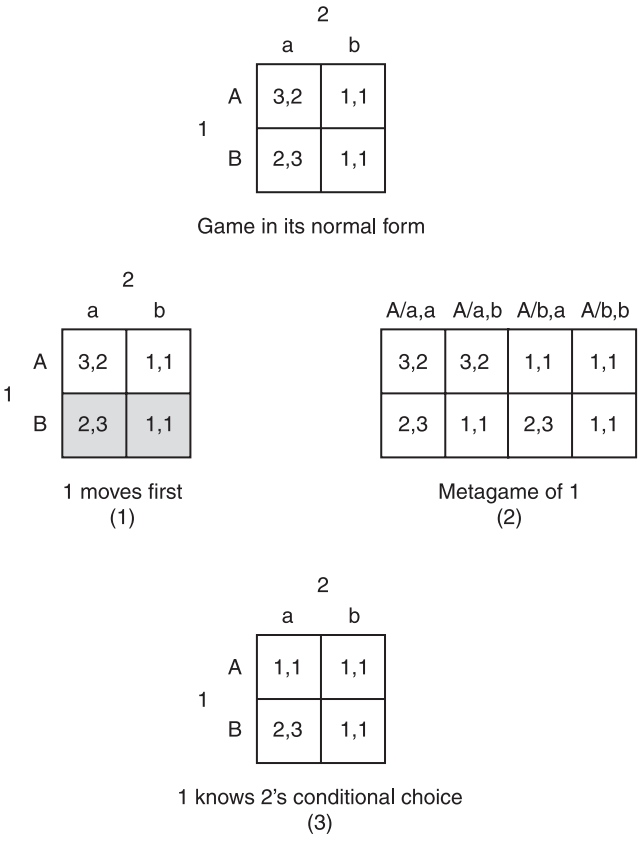

Figure 6.1

the orthodox theory, whereby each player must choose his rational strategy without knowing that which has been chosen by the other, is easily fulfilled.

Transformation (2) reveals to player 1 that there exists a rational outcome which corresponds to a Nash equilibrium in the metagame whose related payoff values benefit player 2 (B; A/b,a). Such complementary information on the game, which does not directly appear in its normal form, will hamper 1 in choosing his rational strategy. The metagame possesses two Nash equilibria and player 1 no longer has a dominant strategy. The Nash equilibria in metagame (2) are regarded as "irrational outcomes" for player 1 due to the "minimax-regret."

Transformation (3) modifies the game to player 2's advantage. The transformation results from an effective – or tacit only – complementary information. This information is represented by player 2's conditional bid to player 1 on the choice of his response strategy. If player 2 gives player 1 no information of this kind, player 1 may be led to anticipate on the basis of his metagame. He can logically fear that player 2 won't choose the strategic response linked with the equilibrium most advantageous to player 1. In this case, the dissuasive bid is replaced by what Schelling calls a "tacit deterrence," which results from player 1's expectation.

This example illustrates a number of Schelling's favorite ideas. The information that a player can acquire on the other's strategy and that which he can give to the other about his own strategy both modify or do not modify the data on the game according to the specific configuration of the players' interests, or preferences, in the game. The impact of this information on the definition of the players' rational strategies also appears to be quite varying. In some cases, the information acquired on the other player favors the player who has it; in other cases, it hampers or even penalizes him. As for the information voluntarily (or involuntarily) conveyed to the other, it may, according to the situation, benefit or penalize the player who gives it. Its final impact will depend on the one hand on the nature of such information (actual, potential, or conditional), and, on the other hand, on the respective positions of the recipient and the information-provider according to the configuration of their interests in the game. In any case, for Schelling, this information cannot be regarded as neutral since it is likely to change the definition of the whole of the players' strategies and therefore to be used as a strategy by the players. For example, the decision to move first or to make a bid to deter the other are strategic decisions as such for Schelling. Taking them into account requires an extension of the game as traditionally defined in the theory.

From coordinating expectations to focal points

These observations led Schelling to reconsider the endogenization of the players' anticipations, an issue that Von Neumann raised as early as 1928. Von Neumann's assertion that the players' uncertainty in those games which

he calls "strategic" applies only to the strategies chosen by the other players, is absolutely correct. But the handling of such uncertainty cannot be reduced to a traditional problem of individual choice under uncertainty. Indeed, each player in game theory must choose his strategy while ignoring the others' strategic choices, but the separate treatment of each player's uncertainty boils down to ignoring a core characteristic of the game, which lies in the interdependence of the players' expectations. In extending this observation Schelling explains the issue of the relationships between the player's rationality and his information about others' choices, as a consequence of the problems faced by the players in coordinating their expectations.

For Schelling, the root of such problems is to be found in the limits of the rationality assumption. He first shows, as opposed to Harsanyi's repeated assertions (1957, 1961, 1962), that the symmetry hypothesis, in the strong acceptation, cannot be inferred from the maximizing rationality ascribed to the players by game theory (Schelling 1960: Appendix B). As opposed to Harsanyi's and Selten's views (Harsanyi and Selten 1988: 342–3), he demonstrates that the symmetry hypothesis is neither a necessary nor a sufficient element to endogenize players' anticipations. Schelling carries on and gives a number of examples of pure coordination games, where players' personal rationality, even if rationality is common knowledge, does not enable them to coordinate their expectations (see the famous "meeting points game" example). Whatever the methods suggested by later theoreticians to reinforce the definition of players' rationality, especially in their successive attempts towards further refining the Nash equilibrium, none of them provides for the coordination of the players' strategic choices within the Schelling examples (Schelling 1960: Appendix C).

It could be argued that, as shown by Schelling, the coordination of players' strategic choices is not necessarily done by coordinating their expectations. In a two-person coordination game with two equilibria, where players' interests are absolutely identical, it only takes one of the players "randomly" choosing his strategy first to have the second one – who is informed about the selected strategy – choosing rationally and thereby ensuring the game's coordination on either equilibrium. However, according to game theory, such a prescription is not contained in the description of the game. Unless the game's definition is changed, one must adhere to Schelling and acknowledge that the strategies' coordination is indissolubly linked with the coordination of strategies' expectations. Such a coordination cannot be reduced to the mere juxtaposition of rational expectations between independent players.

On the basis of his examples of coordination games, Schelling grew convinced that the solution to the problem of the coordination of players' expectations was to be sought in a way that would not consist in further studying individual rationality as traditionally dealt with in game theory. Such coordination can be done on the basis of focal points, common to the players and which serve as a support for their anticipations. Instead of trying directly to expect the strategy rationally chosen by the other, the players will

seek to bring out a basis common to their mutual expectations. This common reference belongs to the contextual environment in which players are acting (i.e. configuration of interests and preferences), without being part of the game as defined by the theory. Focal points result from research by the players along a procedure which Schelling sketches by way of the following analogy:

> The type of "rationality" or intellectual skill required in these games is something like that required in solving riddles . . . A riddle is essentially a two-person problem; the methodology of solution depends on the fact that another person has planted a message that in his judgment is hard to find but not too hard. In principle one can neither make up or solve riddles without empirical experience; one cannot deduce a priori whether a rational partner can take a hint. Hint theory is an inherently empirical part of game theory.
>
> (Schelling 1960: 295, n. 3)

Schelling's approach to rationality marks his break from the dominant approach developed in the theory of non-cooperative games since Nash. It alters the exceptions/rule relationship. For Schelling, pure coordination games are not pathological cases of the non-cooperative game theory; rather, they provide pedagogic supports to point out and magnify the specificities of a coordination problem that concerns all the classes of non-cooperative games. The knowledge of the game which the theory conveys to players is, allowing for exceptions, inadequate to enable them to coordinate their anticipations for reaching the solution. To do so, it is not the hypothesis of individual rationality that should be common knowledge between players but some information outside the game which the players can make use of when needed. The object of such common knowledge makes it quite different from that in the orthodox game theory. It is not a priori provided to the players through the assumption of rationality. They must find it out. Procedures to select and to treat the information are required, which, while being outside the game, can nonetheless be drawn from the empirical context to which it is related.

For Schelling, the player's discovery and use of focal points occur by way of a threefold process: (1) Look for peculiarities contained in the concrete situation of the game (asymmetries, irregularities of all sorts, etc.); (2) Amongst the peculiarities, select those which are "significant," due to the fact that they may be interpreted in a common way between players; (3) Choose a strategy by adjusting behavior to those prescriptions derived from such common interpretation. The three operations are implemented intuitively in the examples of coordination games studied by Schelling. The lack of formalization must be linked to the limitations of the examples studied. In one-shot games, operations (1), (2), and (3) are obviously independent from the logical structure of the game. This is no longer the case with multiple-sequence

games and a fortiori with repeated games. In both cases, players, while looking for focal points, may use some information drawn from the game's progress (or from their own knowledge of previous plays), which opens the way for systematizing operations 1, 2, and 3.

The orthodox temptation

Some authors try to link the search for focal points with the application of "forward induction." In game theory the signal provided to the other players by the choice of the player who operates at a previous stage of the game is seen as a "clue" which contributes the solution of the coordination enigma, according to Schelling's analogy (van Damme 1989). One can also revisit Schelling's examples, as for the meeting point game, in increasing the number of its possibilities and handling it in a repeated game. On the basis of the information collected by players during each play, several authors seek to work up a calculation algorithm which provides players rules for the quickest way for players to coordinate (Crawford and Haller 1990; Ponssard 1993, 1994). Other researchers work toward inductively reconstituting which processes guide towards the focal points within experimental games (Roth, 1985).

These recent works prove that there exist links between the heterodox theory of coordination by focal points and the orthodox game theory. Such links must not conceal the differences. The focal points theory studies the specificities of each game according to its own narrative support (context), while the search for a formal treatment of coordination procedure aims at a general endogenous protocol. Actually, the objective is to discover a general treatment rule for handling the peculiarities attached to the course of each game. Unfortunately, it can be observed that the number of such peculiarities rapidly increases as the number of players and of related pure strategies increases too. In these circumstances, the quest for a general coordination game theory could not make sense. Schelling's followers are faced with a type of problem met at another level by Von Neumann and Morgenstern in *TGEB*, when the increased number of players in cooperative games rendered the study of their coalitions more and more complicated, making a unified theory of cooperative games questionable.

The extension of Schelling's concepts to repeated games brings out the various ways to study the connections between simple games and repeated games (or, more precisely, "super-games"). A super-game can be devised as the representation of a simple game identically repeated from one phase to the next. But any game can also be regarded as the result of a simplification carried out on a super-game. The formalization of focal points would have to go through the reconstruction of the super-games, on the basis of which the coordination games presented by Schelling have been defined. This would avoid such treatments being criticized as artificial.

Notes

1 See, for example, H. Greenberg's theory of social situations (Greenberg 1990) which is analyzed elsewhere (Schmidt 2001).
2 Therefore the assumption that the common knowledge of the game leads to a Nash equilibrium was progressively relaxed (Aumann and Brandenburger 1995).
3 Aumann later relaxed this hypothesis (Aumann 1999).
4 As, for example, Harsanyi's standard forms of a game (Harsanyi and Selten 1988).
5 Roughly speaking, in logic the counterfactuals are a special kind of the broad class of conditional propositions.

References

Aumann, R. J. (1987a), "Correlated equilibrium as an expression of Bayesian Rationality," *Econometrica*, 55.

Aumann, R. J. (1987b), "Game theory," in *Game Theory, the New Palgrave: A Dictionary of Economics*, Eatwell, J., Milgate, M. and Newman, P., eds, London, Macmillan.

Aumann, R. J. (1992), "Irrationality in game theory," in *Economic Theories of Policies*, Arrow, K., ed., Cambridge, Cambridge University Press.

Aumann, R. J. (1994), "Notes on interactive epistemology," mimeo.

Aumann, R. J. (1995), "Backward induction and common knowledge of rationality," *Games and Economic Behavior*, 8.

Aumann, R. J. (1999), "Interactive epistemology II: probability," *International Journal of Game Theory*, 28, 302–30.

Aumann, R. J. and Brandenburger, A. (1995), "Epistemic conditions for Nash equilibrium," *Econometrica*, 63, 1161–80.

Binmore, K. (1987), "Modeling rational players I," *Economics and Philosophy*, 3.

Binmore, K. (1990), "Foundations of game theory," in *Advances in Economic Theory, Proceeding of the 6th World Congress of the Econometric Society*, Vol. 1, Cambridge, Cambridge University Press.

Binmore, K. (1992), *Essays on the Foundations of Game Theory*, Oxford, Basil Blackwell.

Binmore, K. (1993), "De-Bayesing game theory," in *Frontiers of Game Theory*, Binmore, K., Kirman, A., and Tani, P., eds, Cambridge, MA, MIT Press.

Binmore, K. and Samuelson, L. (1996), "Rationalizing backward induction?" in *Rationality and Economic Behaviour*, Arrow, K., Colombatto, E., Perlman, M., and Schmidt, C., eds, London, Macmillan.

Crawford, V. P. (1991), "Thomas Schelling and the analysis of strategic behavior," in *Strategy and Choice*, Zeckauser, R. J., ed., Cambridge, MA, MIT Press.

Crawford, V. P. and Haller, H. (1990), "Learning how to cooperate: optimal play in repeated coordination games," *Econometrica*, 3.

Fudenberg, D. and Tirole, J. (1991), *Game Theory*, Cambridge, MA, MIT Press.

Hammond, P. J. (1995), "Consequentialism, structural rationality and game theory," in *Rationality and Economic Behaviour*, Arrow, K., Colombatto, E., Perlman, M., and Schmidt, C., eds, London, Macmillan.

Harsanyi, J. C. (1956), "Approaches to the bargaining before and after the theory of games: A critical discussion of Zeuthen's, Hick's and Nash's theories," *Econometrica*, 24.

Harsanyi, J. C. (1957), "Rejoinder on the bargaining problem," *Southern Economic Journal*, 24.
Harsanyi, J. C. (1961), "On the rationality postulates underlying the theory of cooperative games," *Journal of Conflict Resolution*, 5.
Harsanyi, J. C. (1962), "Bargaining in ignorance of the opponent's utility function," *Journal of Conflict Resolution*, 6.
Harsanyi, J. C. (1975), "The tracing procedure," *International Journal of Game Theory*, 4.
Harsanyi, J. C. (1977), *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge, Cambridge University Press.
Harsanyi, J. C. and Selten, R. (1988), *A General Theory of Equilibrium Selection Games*, Cambridge, MA, MIT Press.
Howard, N. (1970), "Some developments in the theory and applications of metagames," *General System*, Yearbook of the Society for General System Research, 15.
Howard, N. (1971), *Paradoxes of Rationality: Theory of Metagames and Political Behavior*, Cambridge, MA, MIT Press.
Howard, N. (1987), "The present and the future of metagame analysis," *European Journal of Operational Research*, 32, 1–25.
Kalai, E. and Samet, D. (1984), "Persistent equilibria," *International Journal of Game Theory*, 13.
Kohlberg, E. and Mertens, J. F. (1986), "On the strategic stability of equilibria," *Econometrica*, 5.
Kreps, D. and Wilson, R. (1982), "Sequential equilibria," *Econometrica*, 4.
Lewis, D. (1978), *Counterfactuals*, Cambridge, MA, Harvard University Press.
Mertens, J. P. (1987), "Supergames and repeated games," in *Game Theory, the New Palgrave, a Dictionary of Economics*, Eatwell, J., Milgate, M., and Newman, P., eds, London, Macmillan.
Morgenstern, O. (1934), "Volkommene vor Aussicht und Wirschaftliches Gleichgewich," in *Zeitschrift für National Alökonomie*, 6, English trans.: *Perfect Foresight and Equilibrium* (1937).
Myerson, R. (1978), "Refinement of the Nash equilibrium concept," *International Journal of Game Theory*, 7.
Nash, J. F. (1950a), "The bargaining problem," *Econometrica*, 18, 155–62.
Nash, J. F. (1950b), "Equilibrium points in n-person games," *Proceedings of the National Academy of Science*, 36.
Nash, J. F. (1951), "Non-cooperative games," *Annals of Mathematics*, 54.
Nash, J. F. (1953), "Two-person cooperative games," *Econometrica*, 21.
Osborne, M. J. and Rubinstein, A. (1994), *A Course in Game Theory*, Cambridge, MA, MIT Press.
Ponssard, J. P. (1990), "Self-enforceable paths in games in extensive form: a behavior approach based on interactivity," *Theory and Decision*, 19.
Ponssard, J. P. (1993), "Formalisation des connaissances, apprentissage organisationnel et rationalité interactive," in *Analyse Economique des Conventions*, Orléan, A., ed., Paris, PUF.
Roth, A. E. (1985), "Toward a focal point theory of bargaining," in *Game theoretic Models of Bargaining*, Roth, A. E., ed., Cambridge, Cambridge University Press.
Schelling, T. C. (1959), "For the abandonment of symmetry in game theory," *The Review of Economics and Statistics*, 3.

Schelling, T. C. (1960), *The Strategy of Conflict*, Cambridge, MA, Harvard University Press.
Schmidt, C. (1990), "Dissuasion, rationalité et magasins à succursales multiples," *Revue d'Economie Politique*, 5.
Schmidt, C. (1994), "Preferences, beliefs, knowledge and crisis in the international decision-making process: a theoretical approach through qualitative games," in *Game Theory and International Relation*, Allan, P. and Schmidt, C., eds, Aldershot, Edward Elgar.
Schmidt, C. (2001), "From the 'standards of behavior' to the 'theory of social situations': a contribution of game theory to the understanding of institutions," in *Knowledge, Social Institutions and the Division of Labour*, Porta, P. L., Scazzieri, R., and Skinner, A., eds, Cheltenham, Edward Elgar, 153–68.
Selten, R. (1975), "Reexamination of the perfectness concept for equilibrium points in extensives games," *International Journal of Game Theory*, 4.
Selten, R. (1978), "The chain store paradox," *Theory and Decision*, 9.
Selten, R. and Leopold, V. (1982), "Subjunctive conditionals in decision and game theory" in *Philosophy of Economics*, Stegmüller, W., Balzer, W., and Spohn, W., eds, Berlin, Springer-Verlag.
Shubik, M. (1992), "Game theory at Princeton, 1949–1955: A personal reminiscence," in *Toward a History of Game Theory*, Weintraub, E. R., Durham, Duke University Press.
Van Damme, E. (1989), "Stable equilibria and forward induction," *Journal of Economic Theory*, 48.
Von Neumann, J. (1959, 1928), trans. from the German as *On the Theory of Games Strategy*, in *Contributions to the Theory of Games*, Vol. IV, Tucker, A. W. and Luce, R. D., eds, Princeton, Princeton University Press, 13–42.
Von Neumann, J. and Morgenstern, O. (1944), *Theory of Games and Economic Behavior*, Princeton, Princeton University Press.
Weintraub, E. R., ed. (1992), *Toward a History of Game Theory*, Durham, Duke University Press.

7 From specularity to temporality in game theory

Jean-Louis Rullière and Bernard Walliser

Hell, it's the others.
(Jean-Paul Sartre)

Introduction

From its very beginning, non-cooperative game theory was primarily interested in the coordination of individual agents and formally defined various equilibrium notions whose main properties were precisely scrutinized. These notions were first introduced from the modeler's point of view, expressing necessary conditions ensuring a compatibility between expectations and plans for each player and between plans for all players. Such an approach is not in prior accordance with methodological individualism, which requires the emergence of an equilibrium to result from the conjunction of individual behaviors without the introduction of an outside entity. Hence these notions have to be justified from the players' point of view: the modeler exhibits concrete coordination processes by which they succeed in adjusting their respective actions in order to reach an equilibrium state.

All through its history, game theory has shown moreover a shift in the examined processes, from psychical, implemented by introvert agents, to physical, implemented by extrovert agents. Initially, equilibrium was achieved through introspective reasoning of players able to simulate the others' intended strategies by leaning on a common knowledge of game characteristics. Recently, equilibrium was attained by interactive learning processes of players able to mutually adapt their implemented behaviors by crystallizing into a common experience the past play of the game. Such a substitution between cognitive specularity and effective temporality induces a substantial change in the methodological status of game theory, but methodological individualism stays problematic in both cases.

In its first section, the chapter is concerned with players' mental computation mechanisms about classes of games which imply ever-more sophisticated simulated temporality and where specularity progressively finds its limits. The games considered are first static when stated in normal form, then become dynamic when expressed in extensive form, and finally introduce various

kinds of uncertainty reduced with time. In its second section, the chapter is concerned with players' dynamical adaptation rules embedded in types of processes which assume ever less demanding specularity and where temporality progressively finds its limits. The implemented processes imply first belief revision of others' characteristics then forecast from the past of others' strategies, and finally reinforcement according to past utilities of one's own strategies.

Reasoning in static games

In game theory, each player is an entity quasi-isolable from his social environment whose deliberation process is rational in that he achieves compatibility between three basic choice characteristics (Walliser 1989). Cognitive rationality ensures adequation between *beliefs* about the outside world and available pieces of information, a strong version stating that the player forms expectations by handling knowledge along the usual Bayesian rules. Instrumental rationality ensures adequation between given *opportunities* and well anchored *preferences*, a strong version asserting that the player chooses a plan by maximizing a utility function under constraints and according to prior expectations. By combination, a player is Bayesian rational if he links information and plan by maximizing expected utility, against exogenous uncertainty in a passive environment or endogenous uncertainty in a strategic environment.

In a static game, two players are linked exclusively through their preferences, since the utility of one player not only depends on his own action, but on the other's. In order to achieve his deliberation process, each player is then required to anticipate the other's behavior, hence to simulate mentally the other's deliberation process by putting himself in the other's shoes. He assumes that the other acts along the same choice principles as he does himself, i.e. he postulates that the other is rational instrumentally as well as cognitively. He has moreover to predict the other's choice characteristics that generally differ from his own, and these characterictics – summarized in the player's type – are assumed to be commonly known (until later).

However, when simulating the other's behavior according to his own beliefs, each player has to deduce how the other predicts his own behavior, hence how the other simulates his own deliberation process at a second level. Each player gets engaged, by a play of mirrors, into a sequence of crossed expectations at successive levels on the other's strategy (I expect that you expect that I expect . . . that you do action a). This sequence is grounded, by the corresponding play of mirrors, on a sequence of crossed knowledges about his choice characteristics (I know that you know that I know . . . that your type is t). By the fact that the preceding specularity is potentially unended, it remains to be shown that the coupled mental processes may converge toward some kind of belief equilibrium without appeal to an external entity.

The cognitive foundations of equilibria rest on a model where Bayesian rational players face an uncertainty space containing their actions and where each of them chooses a strategy conditionally on the uncertain state. The first assumption concerns common knowledge of players' rationality, i.e. each player knows that the other is Bayesian rational, knows that the other knows he is rational, and so on till infinity. The second assumption concerns common knowledge of game structure, i.e. each player knows the other's type as well as the game rules, knows that the other knows them and so on. The third assumption concerns common knowledge of players' independence of play, i.e. it is commonly known that the players decide their actions without assuming that these actions are correlated in any way.

The three preceding assumptions lead to a rationalizable equilibrium, in which each strategy of a player is a best response to the other's expected strategy, itself considered as a best response to the other's expectation and so on till closing the loop at some finite level (Bernheim 1984; Pearce 1984). But it is harder to single out a Nash equilibrium, where each strategy of a player is a best response to the other's equilibrium strategy, the loop being closed simultaneously for both players at the second level. It would be necessary to introduce the more drastic assumption that the player's conjuncture on the other's strategy is common knowledge (Tan and Werlang 1988) or at least mutual knowledge (Aumann and Brandenburger 1995). Hence, methodological individualism gets into trouble, since it is necessary to explain how these expectations can be guessed, otherwise than by assuming that the players consider the Nash equilibrium to be the adequate equilibrium notion.

Even if players' mental reasoning justifies some equilibrium notion, it says little about how the players select a specific equilibrium state, in as much as the definition and selection problems can be considered as sequentially treated by the players. For differentiated equilibria, one can be eliminated or privileged by general comparison criteria (Pareto optimality, risk dominance, symmetry, etc.), which have to be common knowledge among the players. For identical equilibria, one can be singled out as a focal point (Schelling 1960) by using local criteria based on background knowledge (habits, saliences, etc.), which have to be common knowledge too. Hence, methodological individualism is again in question, since players refer to conventions which are shared by all of them, but which are neither included in the game description nor constructed by the players.

Reasoning in dynamic games

When introducing time into the game structure represented now by a game tree, several adaptations have to be brought to the static framework of rational deliberation. First, the player's preferences, opportunities and prior beliefs are not only considered as exogenous but also stable, excluding phenomena such as addiction induced by past actions. Second, cognitive rationality is grounded on bayesianism which states that players' current beliefs

evolve according to revision rules, when observing others' actions which constitute their main means of communication. Third, instrumental rationality is grounded on consequentialism, which states that the action chosen at each step by expected utility maximization depends exclusively on its future consequences.

In a dynamic game, the opportunities of two players are linked together, since the available actions of a player at same node of the game tree depend on the past actions of both. Each player is confronted with a second reason to anticipate his opponent's actions, hence to simulate his opponent's future behavior on all further periods traced on a subjective time scale simulating the objective one. As he knows that the other does the same, he has to anticipate again how the other predicts his own future actions, leading to crossed expectations on sequences of actions based on personal beliefs of game structure. However, he is confronted at each step with the two arrows of time: cognitive rationality acting forward from observation of past actions, instrumental rationality acting backward from expectation of future actions.

The extensive form of a game may be reduced to normal by introducing the notion of a strategy, which indicates what a player's action would be in all situations where he is liable to play. A strategy is then decided by each player at the very beginning of the game and is assumed to be faithfully implemented, the combination of strategies defining a unique path in the game tree. As far as instrumental rationality is concerned, the two forms are equivalent since, when arrived at a given node, a player implements the action indicated by his strategy, and determined by its only future consequences. As far as cognitive rationality is concerned, the two forms may differ since, when finding himself at a given node, the player is assumed to follow his strategy rigidly without assessing again the observed past which may differ from the expected one.

Coming back to the cognitive foundations of equilibria, the three assumptions of common knowledge of players' rationality, of game structure and of players' independence seem to justify a subgame perfect equilibrium. Such an equilibrium states that the players' strategies form a Nash equilibrium in each subgame, starting at some node, and is obtained, in finite games, by the backward induction procedure. It expresses that the future behavior is time-decomposable in the sense that each player at an end node chooses his best action, then each player at a preceding node chooses his best action knowing the action his follower will take and so on, till reaching the root node. The subgame perfect equilibrium is then justified by the fact that the last player is rational, that the last but one player is rational and knows that his follower is rational and so on.

However, the procedure leads to the backward induction paradox (Binmore 1987), showing an incompleteness of the player's reasoning when he considers himself mentally in a node unreached by the equilibrium path. He cannot be content then with looking only at the future of the game and keeping his strategy, but he has to understand why his opponent has

deviated and he may be induced or not to deviate himself according to the interpretation of the other's deviation. Aumann (1995) solves the paradox by saying that such a deviation can simply not happen and should not be considered by the player, since common knowledge of rationality ensures the equilibrium path to be implemented. Binmore (1995) answers that a deviation is always possible because the player's rationality is not a physical law and may be violated, hence taking this fact into account may ruin the backward induction principle.

When considering himself mentally at a node off equilibrium, the player appeals in fact to counterfactual reasoning, and has to revise in some way his beliefs about his opponent. He can call into question any of the three basic assumptions he made as well as some auxiliary assumptions, and according to the assumption refuted and modified, the subgame perfect equilibrium appears to be maintained or not. The revision rules of each player consist more precisely in ranking all assumptions according to an epistemic entrenchment index in order to abandon the less entrenched until restoring belief consistency. Methodological individualism is again challenged since, even if the revision rules can be included in player's type and considered as common knowledge as usual, they act really as collective conventions able to interpret univocally possible deviations.

Reasoning in games with uncertainty

The player's uncertainty is structural when it deals with the game structure in static or dynamic games, especially with the player's characteristics which stay well known by the player himself but no longer by his opponent. Such incomplete information is reduced, by virtue of the Harsanyi doctrine (Harsanyi 1967–8) to a standard form, i.e. probabilized uncertainty attributed to the play of a passive new player assimilated to nature. In a first step, all the choice characteristics of each player are summarized in a set of possible types which he can adopt, the range of all players' types being considered as common knowledge. In a second step, a prior probability distribution is defined globally on the type space and affected to nature which plays first but secretly, such a common prior being itself common knowledge (Morris 1995).

The treatment of structural uncertainty brings to the fore a compromise, with regard to the preciseness of each player's prior information, between the two first levels of his beliefs about his opponent. At level one, the player's knowledge is moderately limited, since he is endowed with a probability distribution on the other's type, and he knows perfectly the support of that probability distribution. At level two, the player's knowledge gets strongly demanding, since that probability distribution, obtained by conditioning the common prior on his own type, is assumed exempt of ambiguity. Such a drastic treatment recovers, as soon as level two is reached, a certainty lost at level one, hence this rules out that a player has only a bounded confidence degree in his assessment of the other's type.

The treatment of structural uncertainty also emphasizes a symmetrization, with regard to the players' respective knowledge on their types, when climbing in the first two levels (Garrouste and Rullière 1994). At level one, the players' knowledge looks asymmetric at first glance, since each ignores the other's type while knowing his own, even if they are already placed in similar positions. At level two, the players' asymmetry becomes perfectly known by both, and both analyze it by conditioning on the same prior distribution, so that their computations become perfectly parallel and moreover precoordinated. This severe treatment restores, as soon as level two is reached, a symmetry lost at level one, hence this rules out that a player has a subjective assessment on the other's characteristics grounded on personal experience.

The player's uncertainty is factual when it deals with game past play in a true sequential game, especially with the player's past moves, which are well known by the player himself but not by his opponent. Such imperfect information is again reduced to a standard form, that is probabilized uncertainty, even if it looks this time specific to each of the concerned players rather than shared. In a first step, for the player who happens to move, the nodes of the game tree he is unable to discriminate are gathered in an information set integrated in the game structure and considered again as common knowledge. In a second step, the nodes of that information set, where each corresponds to a precise sequence of past moves of all players, are affected with a probability distribution reflecting the player's private beliefs.

From the modeler's viewpoint, a perfect bayesian equilibrium is defined by two conditions expressing through a loop the necessary consistency between the player's actions and beliefs at each information set (Kreps and Wilson 1982). On the one hand, the best action intended by the player is computed by the backward induction procedure, in accordance with a belief weighting the nodes included in the information set. On the other hand, the belief adopted by each player is adapted through the Bayesian rule, knowing that the past actions of the opponent are considered as optimizing ones. If the equilibrium states appear theoretically as fixed points of that strategy-belief loop, the modeler is aware of no general algorithm for calculating them, and is content with validating equilibria obtained by intuition and trial.

From the player's viewpoint, a perfect Bayesian equilibrium can hardly result from an autonomous mental process, in which he revises his beliefs and adapts his strategies along the simulated progress of the game. Even if all players agree to consider the perfect Bayesian equilibrium as the relevant equilibrium notion, they are still confronted with both a computation and a selection problem. They try in fact to simulate the modeler's role, that is, the role of an outside Nash regulator, who suggests to the players some equilibrium state which both have an incentive to accept. However, such a fictitious entity is quite evidentially incompatible with methodological individualism, because he acts as a coordination institution which does not result from players' behaviors alone.

In summary, the ambition to achieve common knowledge of an equilibrium state by the sole reasoning of players looks quite out of reach, since it obliges each player to take the place of the modeler himself. Prior information needed at the beginning of the game about game structure and players' rationality gets more and more drastic when introducing uncertainty, the features of which must be common knowledge. New knowledge built from prior information by each player assumes the last to have a very strong cognitive rationality since he simulates altogether the opponents' behavior and his own. Belief coordination of all players on the same equilibrium state if many are available seems impossible without introducing conventions which are very context-dependent and must be common knowledge.

These drawbacks lead to the reliance on an equilibrium state on the common experience of players engaged in learning when involved in a real-time game, temporally revealing at first stance as a complement, then more drastically as a substitute to specularity. Prior information is reduced and stays specific to each player, while new information is supplied by observation of implemented actions all along the play of the game, in a public and cumulative way. New knowledge is obtained by revision of prior knowledge at each new message in accordance with plain updating rules, and becomes even implicit when action is directly related to information through reinforcement rules. Action coordination of all players toward an equilibrium state results univocally from convergence of the dynamic process, at least under some preconditions on players' prior beliefs or players' matching rules.

Learning on game structure

Players are at first uncertain about some elements of the game structure, such as players' types (beliefs, preferences) or nature's laws, but they become able to observe concretely the successive actions and states in a repeated game played now in real time. They remain endowed with perfect instrumental and cognitive rationality, the last allowing them to revise their prior beliefs on the game structure each time they receive a new message, according to various revision rules. Players try more precisely to reveal hidden information about their opponents at each period by postulating their rationality, such an inductive process being very similar to the process of accusing on the basis of supposed intentions. Time plays the role of a constant supplier of original messages, which drives the player to update his beliefs on exogenous features between two periods, but keeps the choice of a strategy at each period to be of a specular nature.

In a game with conflicting interests, the player who reveals true information gets generally better off while the other gets worse off, hence the last has an interest in manipulating the information revealed by deviating his action from the optimum. The player does it either stochastically in order to blur information and render it useless for the other or systematically in order to induce false information and acquire a reputation in the other's eyes. In a

game with converging interests, both players may be better off by information revelation, and in games with separate interests, each player is indifferent to the other's information treatment. Two well-known paradoxical games are of the last type, where players' actions consist in spelling out the acquired information and where players' utilities are uniquely concerned with information truth.

The hats' paradox considers three players having on their head either a black or a white hat, and observing initially the others' hat color, but not their own hat's color. At the beginning of the game, an observer knowing for instance that all hats are black announces publicly that "*there is one black hat*," a message which was already privately known by all players, but which becomes accordingly common knowledge since each learns that the others know. At each period, the players go out if they think that they know the color of their own hat (and are rewarded if they are right), and this move constitutes a message which can be interpreted by the other players. Between two periods, each player revises his beliefs about the eight possible color combinations, and more precisely eliminates those appearing inconsistent with the others' actions when simulating their reasoning.

The generals' paradox considers two generals allied to fight a common enemy, but placed in different valleys and having then to coordinate the right moment to attack in order to win. Since this moment depends on the occurrence of a favorable event only observed by the first general, he sends a message to the other when it occurs, but this message has a small probability of being lost. Hence, when receiving the message, the second general sends a counter-message confirming that he got it, but this message has the same probability of being lost, the exchange of messages going on until one does get lost. When receiving one more message, a general gets one level higher in the crossed knowledge about the occurrence of the favorable event, but when receiving no further message, he does not know whether his last one got lost or whether the response got lost.

In typical games, the dynamic process leads to perfect knowledge of the game structure for each player, at least if the messages are sufficiently rich (convergence being due to the law of large numbers in a probabilistic framework). Since all players receive the same public messages which become common knowledge and since all reveal missing information by the simulation of others' behavior, the process may even converge towards common knowledge of the game structure. In games with separate interests, where players act in isolation except for revealing others' information, the eventual equilibrium coincides with the actions taken by all players under common knowledge of the game structure. Concerning the two paradoxes, where actions consist moreover in announcing the computed posterior knowledge, the virtual equilibrium exactly consists in the achievement of common knowledge of the exogenous nature's state.

In the hats' paradox, the hats' colors become common knowledge in three steps, since the possible worlds (color combinations) being finite, mutual

knowledge at level three becomes common knowledge. In the generals' paradox, the event occurrence never becomes common knowledge and the generals never attack, since the possible worlds (number of messages) being infinite, mutual knowledge never becomes common knowledge at some hierarchical level. However, even in the last game, coordination can be approximatively achieved (the generals will attack with high probability) by weakening common knowledge to almost common knowledge of the game payoffs (Monderer and Samet 1989). Cognitive rationality then gets bounded in the sense that players believe at each hierarchical level with a non-degenerate probability of being right, but crossed knowledge still continues to infinity.

Learning on game play

Players are now more directly uncertain about the game future play, that is, about their opponents' future strategies and nature's future states, but they observe again the past actions and states in a really repeated game. If they stay endowed with perfect instrumental rationality, their cognitive rationality becomes bounded and allows them only to expect others' future strategies from past actions, according to various forecasting rules. Players no longer try to reveal the others' types in order to forecast their behaviors, but are content with linking their expected strategies to their past actions (past strategies are not observable) by leaning on an inertia principle. Time plays the role of a schedule for the successive actions and reactions of the optimizing players, which allow them to mutually adapt their endogenously determined strategies, without further need for crossed expectations or knowledge.

In fact, players achieve passive experimentation when information gathered about others' actions is simply a by-product of the normal play of the game, but is nevertheless diversified since the learning process is non-stationary. But they may also try active experimentation when they deviate voluntarily from their optimal strategy in order to test the others' reactions, even if such a behavior is hard to interpret for the opponents. They have in fact to deal with a trade-off between exploration in search of original information about their strategic environment and exploitation of relevant information for ameliorating the ongoing strategy. The trade-off already looks very complicated when playing against a passive stochastic environment as in the two-armed bandit problem (Aghion *et al.* 1991), but is even harder in a strategic environment where the opponents shift in their answers.

The epistemic learning process is rational when a player has a prior probability distribution over the other's dynamic strategies, and revises it in a Bayesian way each time he receives a new message about the implemented actions. A very simple rule assumes that he eliminates progressively the strategies incompatible with the observed actions, an incremental rule which works as long as the obtained messages do not contradict the current

knowledge. In such a process, each player is perfectly aware that his opponent is learning too, since he considers that the other uses not only stationary strategies, but strategies conditioned in a more subtle way by the past history of the game. However, the game structure is assumed to be sufficiently stable along the process since new opportunities make it necessary to enlarge the set of possible strategies while evolving preferences necessitate contemplating even more sophisticated strategies.

The epistemic learning process is adaptive when the player directly expects a probability distribution on the other's future actions in an extrapolative way with regard to the past observed actions and to the initial assessment. The most classical rule is fictitious play, which states that the expected mixed action is equal to the frequency of past ones, a pure mean extrapolation of the past which can be applied whatever the observed actions. In such a process, each player reasons implicitly as if the other's behavior were stationary and tries to find out his general trend, even if this trend is duly re-evaluated at each period for the next one. By contrast, the game structure is allowed to vary within some limits throughout the process, since the trend of observed actions will reflect that evolution with some delay, even if new actions become available and are progressively tested.

An adaptative learning process may follow a cycle (Shapley 1964), but is liable to converge towards a Nash equilibrium if the game is endowed with a simple enough structure. A rational learning process eventually converges towards a subjective Nash equilibrium (or self-confirming equilibrium), such that the messages the player gets thereafter induce him to change his beliefs and thus his strategies (Kalai-Lehrer 1993). Under some conditions (Battigalli *et al.* 1993), it is a weaker notion than a Nash equilibrium since the corresponding beliefs are only locally true (at the equilibrium state) and not globally (at non-equilibrium states). As a restriction to methodological individualism, a necessary precoordination condition for converging is that the beliefs contain a grain of truth, i.e. give a positive probability to the other's actually followed strategy.

The selection of an equilibrium state is naturally achieved, besides those of the forecasting rules, by the initial beliefs considered as exogenous, even if they may result from preceding experience extracted from similar games. In a game with conflicting interests, asymmetries in the game structure – including those in the prior beliefs – may favor some equilibrium states with large bases of attraction. In a game with converging interests, and especially in a coordination game, asymmetries exist only in the game history, and players may be guided towards a specific equilibrium state thanks to rules exploiting them (Crawford and Haller 1990). Limiting again methodological individualism, these last rules are assumed to be common knowledge among the players, in order to help them polarize toward a specific target among equivalent ones.

Learning on individual payoff

Players are ultimately uncertain too about their own payoff function, relating their utility on the others' actions and on their own, and all they observe in a repeated game is the payoff they got in the past with each possible action. They are endowed with null cognitive rationality since they no more hold beliefs about their environment, and with weak instrumental rationality since they no more optimize their behavior but fix it according to various reinforcement rules. Players no longer anticipate their future payoffs by forecasting the others' future actions, but are content with playing more and more often the actions with the best past results that are also assumed to give the best future results. Time plays the role of support for an optimization algorithm achieved by the player, which improves his payoffs step by step according to past experience, in a purely forward procedure which need not involve any form of expectation.

Players do active experimentation quite mechanically when, at each period, there is a small probability of switching from the ongoing action to any other, independently of the actions' past utilities. They do active experimentation more deliberately when they adopt a stochastic behavior rule conditioned on the past utilities of each action, but which is more diversified than maximization in that it never abandons an action. If available actions depend on some parameter, they may even look for an original action in the neighborhood of an action with good results, or try a new action lying between two actions with good results. In some circumstances, players are, moreover, able to observe the actions implemented by other players in a similar position as well as their corresponding results, hence they imitate players that have succeeded.

The learning process is behavioral when each isolated player is submitted to a trial-and-error process which makes him reinforce the good strategies and inhibit the bad ones in view of their past results. One rule considers that he chooses at each period an action with a probability which is proportional to the cumulative utility it obtained in the past (Arthur 1993). This rule associates to each action a utility index which aggregates its past utilities by computing their sum rather than their average, thus creating a positive feedback in favor of the best actions more often used. Moreover, the rule assumes that an action is choosen stochastically in proportion to the past utility index rather than deterministically by keeping the best action, thus allowing all along the play some form of active experimentation.

The learning process is evolutionary when each player is represented by a subpopulation of agents and when each agent is allowed a fixed strategy but reproduces according to the utility he gets. The classical replicator rule considers that one agent is selected randomly in each subpopulation, that these agents meet together, and that each agent reproduces proportionally to the utility he obtained (Weibull 1995). The rule considers that the reproduction rate depends linearly on obtained utility assimilated to fitness, and it is less harsh than the all-or-nothing rule which assumes that the winner divides in

two while the loser dies. Moreover, the rule is of a deterministic nature as concerns the selection process (but not the encounter process), hence it is frequently supplemented by a regular stochastic supply of mutant strategies to maintain variability.

A behavioral process converges toward an equilibrium notion where a player is aware of no more opportunity to improve his observed utility, a notion badly characterized up to now but which often reduces to pure Nash equilibrium. An evolutionary process converges toward an equilibrium notion better characterized, especially for symmetric games where it is stronger than Nash equilibrium, but depends on the polymorphic or monomorphic character of the population. This equilibrium notion is, however, weaker than an evolutionary stable equilibrium, at which a strategy used by a player cannot be invaded by a small amount of any single mutant strategy (Maynard Smith 1982). Methodological individualism is called into question by the fact that the players' interaction and reproduction mechanisms are not precisely described with regard to players' individual behaviors.

The selection of an equilibrium state is naturally achieved, besides the reinforcement rules, by the various influences acting on the system and inducing a unique path conditioned by context as well as by history. In a deterministic system, the selected equilibrium depends heavily on the prior index distribution on strategies or on the initial population distribution in an evolutionary process. In a stochastic system, the selected equilibrium depends only probabilistically on the initial state, since the system is driven in some direction by the concrete occurrences of players' encounters, players' mutations, and players' behaviors. Methodological individualism is even better satisfied when these stochastic factors are defined by physical laws (neighborhood encounters) rather than by more or less reducible social laws (imitation dynamics).

Specularity and temporality may be used sequentially (but simultaneously by all players) as play goes on, since they have complementary properties to coordinate agents in a given context. Specularity allows a fast adaptation to an original situation by leaning on some shared conventions, but it is unable to take into account all unexpected contingencies and may be marred by equilibrium selection problems. Temporality fits all hazards as soon as they occur and progressively builds coordination mechanisms, but acts very slowly and it is unable to prevent the players from being locked in at second-best situations. Hence, specularity precedes temporality when an initial adaptation to context is followed by a more subtle one and temporality precedes specularity when players become aware of the non-optimality of an emerging equilibrium state and correct it.

Specularity and temporality may be retained respectively by heterogeneous players in the whole progress of a given game, since the players adopt contrasted fashions and possess differentiated capacities (Haltiwanger and Waldman 1985). Specularity is favored by players who prefer future-oriented

deliberation because they dislike facing unexpected events, and who possess the abilities of abstraction needed to forecast globally the strategic future. Temporality is favored by players who prefer past-oriented deliberation because they are convinced only by recognized facts, and who possess the capacities of learning needed to draw the lessons of past experience. Hence, speculative players have to simulate the learning process of the less sophisticated lived-style players while the lived-style players have to test the practical achievements of the less concrete speculative agents.

Specularity and temporality may finally be selected by all players involved in a game all through the play, in accordance with the complexity as well as the duration of the game. Specularity is better fitted to games where the structure is sufficiently simple and already well known by the players and which is played only once or in an identical way a small number of times. Temporality is better fitted to games where the structure is rather complicated and partially hidden from the players and which are played with some variation a great number or an indefinite number of times. However, for most games played recurrently but nevertheless transparently enough, specularity and temporality work together, for instance when collective coordination devices are unconsciously constructed by players' actions, but act in return consciously upon players' behaviors.

References

Aghion, P., Bolton, P., Harris, C., and Jullien, B. (1991) "Optimal Learning by Experimentation," *Review Economic Studies*, 58, 621–54.

Arthur, W. B. (1993) "On Designing Economic Agents that Behave like Human Agents," *Journal of Evolutionary Economics*, 3, 1–22.

Aumann, R. J. (1995) "Backward Induction and Common Knowledge of Rationality," *Games and Economic Behavior*, 8, 6–19.

Aumann, R. J. and Brandenburger, A. (1995) "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63(5), 1161–80.

Battigalli, P., Gilli, M., and Molinari, M. C. (1993) "Learning and Convergence to Equilibrium in Repeated Strategic Interactions: an Introductory Survey," mimeo, Instituto di Economia Politica, Università L. Bocconi; King's College, Cambridge.

Bernheim, B. (1984) "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–28.

Binmore, K. G. (1987) "Modeling Rational Players: I," *Economics and Philosophy*, 3, 179–214.

Binmore, K. G. (1995) "Backward Induction and Rationality," Discussion Paper, London School of Economics.

Crawford, V. P. and Haller, H. (1990) "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica*, 58, 571–96.

Garrouste, P. and Rullière, J. L. (1994) "Survival or Abandonment of the Axion of Symmetry in Game Theory: From Mises-Hayek to Harsanyi-Schelling," mimeo, Université Lumière, Lyon 2.

Haltiwanger, J. and Waldman, M. (1985) "Rational Expectation and the Limits of Rationality: an Analysis of Heterogeneity," *American Economic Review*, 75, 326–40.

Harsanyi, J. C. (1967–8) "Games with Incomplete Information Played by Bayesian Players," *Management Science*, 14, November 1967, 459–62; 15 January 1968, 320–34; March 1968, 486–502.
Kalai, E. and Lehrer, E. (1993) "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–45.
Kreps, D. and Wilson, R. (1982) "Sequential Equilibria," *Econometrica*, 50, 863–94.
Maynard Smith J. (1982) *Evolution and the Theory of Games*, Cambridge University Press.
Monderer, D. and Samet, D. (1989) "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1, 170–90.
Morris, S. (1995) "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11(2), 227–53.
Pearce, D. (1984) "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–50.
Rubinstein, A. (1989) "The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge,'" *American Economic Review*, 79, 385–91.
Schelling, T. C. (1960) *The Strategy of Conflict*, Harvard University Press.
Shapley, L. (1964) "Some Topics in Two-Person Games," in M. Dresher, L. Shapley, and A. Tucker (eds) *Advances in Game Theory*, Princeton University Press.
Tan, C. and Werland, S. (1988) "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 370–91.
Walliser, B. (1989) "Instrumental Rationality and Cognitive Rationality," *Theory and Decision*, 27, 7–36.
Weibull, J. (1995) *Evolutionary Game Theory*, MIT Press.

Part III

Applications

8 Collective choice mechanisms and individual incentives

Claude d'Aspremont and Louis-André Gérard-Varet *

Contemporary economic analysis has recognized the importance of "asymmetric information" not only for markets, but also for many different organizations. The resulting "theory of incentives" aims to define rules, or institutional mechanisms, likely to lead individual agents to make collectively optimal choices and to reveal all private information necessary for an efficient collective choice. A new chapter in game theory has thus been opened: the strategic analysis of economic institutions, their stability, or the conditions under which they may be changed.

The driving forces at the origin of the theory of incentives go well back in the history of economic thought. They can be traced in the first observations regarding possible market failures (externalities, public goods). They are also found in the well-known "controversy over socialism" of the 1930s, which highlighted not only the question of whether collective ownership was compatible with decentralization, but also, with von Hayek (1945), the informational dimension of the workings of markets. Whenever information is dispersed, "equilibrium" requires a process of communication. It may then be in the agents' interest to manipulate such a process and not to reveal fully the information they have at hand.

Wicksell (1896) followed by Samuelson (1954) both held the similar view that, in order to achieve an optimal level of collective goods, it was necessary to centralize some information but that: "it is in the selfish interest of each person to give false signals, to pretend to have less interest in a given collective-consumption activity than he really has" (Samuelson 1954: 388–9). However, Samuelson believed that in a competitive economy, with only privately owned goods, the problem of encouraging agents to reveal their preferences correctly could not arise. It was Hurwicz (1960, 1972) who showed that this problem, as well as the strategic phenomena associated with it, are not due to the introduction of public goods, at least when the number of agents is finite. The theoretical analysis of planning such as by Drèze and de la Vallée Poussin (1971) and Malinvaud (1971), raises similar questions irrespective of whether goods are owned privately or publicly. And it was indeed Vickrey (1961) who provided an essential impetus in our understanding of incentives in his study of

* Louis-André Gérard-Varet died 2001.

the strategic aspects of auctions or tenders, viewed as exchange mechanisms in which the value of the good exchanged is "uncertain."

Informational problems play a significant role in any market involving bilateral contracts. This is true for insurance markets where nowadays a well-established terminology has in fact originated. An insurance company faces the problem of having to cover particular risks which are not fully known. The insurer then offers a set of contracts so as to allow each insured party to self-select from that set the contract that corresponds best to his own type of risks. This is the problem of *adverse selection*. Moreover, the insured party may also control a "self-protection" variable which is not observable by the insurer. The insured party's choice of a value for this variable affects his risk of accident, and may indeed be incompatible with the risk considered by the insurer. Such insufficient self-protection is called *moral hazard*. Other examples illustrating the problems of moral hazard are the management of collective property rights, or sharing a collective outcome (Alchian and Demsetz 1972). It is a priori in everyone's interest to cooperate. But if each participant can gain an individual advantage by acting as a "free-rider," then agreement may be impossible or unachievable.

This chapter aims to show how the two problems of moral hazard and adverse selection may be integrated and treated within the general theory of non-cooperative games. To ensure that the determination of players' behavior is understood by all and is not to impute irrationality to any player, the outcome of such games must be a non-cooperative (Nash) equilibrium. This is a state of individual strategies in which no participant is able to take his own alternative route and expect to achieve a better outcome. Moral hazard reflects situations of "bad equilibria," that is, equilibria which, due to strategic externalities, are socially unsatisfactory. But game theory can also propose a solution: some *cooperative transformation* may be introduced creating a new game with equilibria having better welfare properties. Such a transformation can come about through a "regulation," a "mediation" or an "audit." It may be obtained by "repeating" the game, by adding a "communication scheme," or by "contractually" modifying the original payoff structure. The latter is the aspect we shall concentrate on here. However, cooperative transformation can also give rise to multiple equilibria. It may thus be necessary to introduce an additional contractual structure not only to define new rules, but also to select a collectively efficient equilibrium.

One needs to be careful about the definition of an equilibrium in a model which integrates both adverse selection and moral hazard, and therefore relies on the analysis of "games with incomplete information." A model of a game with incomplete information must specifiy not only what each player "can do," that is the actions, but also what he "knows," that is his private information, and his "beliefs" as regards other players' information. In this framework, an individual strategy consists of a plan of action that defines what action is selected in every contingency, given the player's private information (or what he acquires during the course of the game). Thus, a

necessary requirement is the notion of a "Bayesian equilibrium" (Harsanyi 1967–8) generalizing the Nash equilibrium to situations of incomplete information: the players' "conjectures" about their mutual behavior are confirmed by the decisions taken by each on the basis of their private information.[1]

First, we will introduce a general model allowing the analysis of incentives constraints in situations where adverse selection and moral hazard are both present. We will however make a strong assumption: players assign to strategic issues utilities which are quasi-linear with respect to a numeraire (or money). This hypothesis facilitates the analysis in two ways. On the one hand, the cooperative transformation of the game can be limited to the introduction of a single instrument, namely monetary transfers. On the other hand, it considerably simplifies the definition of the welfare criterion in a world with incomplete information, by allowing us to retain the criterion of Pareto optimality in its most standard form.[2]

Next, we will obtain further simplification by applying the "revelation principle." The Bayesian equilibrium in a game with incomplete information, which is defined by the players' individual plan of action contingent on their private information, can indeed be viewed as the equilibrium of a "game of revelation" where each player's choice is related to the information he reveals. Incentive constraints characterize those issues which guarantee that it is optimal for each player to communicate his information sincerely: this is the meaning of the "revelation principle."[3] A game of revelation may involve an extremely complicated mechanism. From the viewpoint of incentive constraints, we will ignore these complexities in order to concentrate on the issues which are compatible with the sincere revelation of private information by participants. This framework allows us to relate our general model to examples in the literature. It will also allow us finally to define in the third section, a large class of situations where individual incentives and collective efficiency are compatible. In conclusion, we will discuss other ways of integrating moral hazard and adverse selection, as well as other criteria of collective efficiency.

A general model

We consider strategic situations where players have imperfect information with respect to the choices of possible courses of action and also have incomplete information regarding their individual characteristics by other players. The problem of moral hazard derives from imperfect information. Adverse selection is due to incomplete information. In such situations, players strategies have two components: an "action" which is non-observable and a "signal" which is observable. A vector of players' strategies generates a probability distribution over outcomes. The assumption that individual utilities are quasi-linear with respect to money allows us to analyze the conditions for a Bayesian equilibrium to be collectively efficient in terms of "transfers" between players.

The environment: imperfect and incomplete information

There is a finite set N of players facing a finite set X of collective *outcomes*, which are publicly observable. The *utility* (or payoff) of player $i \in N$ derived from outcome $x \in X$ is given by the real-valued function $u_i(x, d_i, \alpha_i)$, where d_i is a choice of action from a finite set D_i, and $\alpha_i \in A_i$ is a state of *private information* also in a finite set. Players make their choice of actions simultaneously, so that no one can observe the complete chosen profile of actions $d = (d_1, \ldots, d_i, \ldots, d_n) \in D \stackrel{\text{def}}{=} X_{i \in N} D_i$. The information at the disposal of player i only concerns his own characteristics. A_i can also be viewed as the set of all possible *types* of players. The sets A_i and D_i, as well as the set X_i are common knowledge. On the other hand, only player i knows the type in A_i which describes his true characteristics. Since the types are not observable, no player i can tell which state

$$\alpha_{-i} = (\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \alpha_n) \in A_{-i} \stackrel{\text{def}}{=} \underset{j \neq I}{\mathrm{X}} A_j$$

describes the characteristics of the other players.[4]

We restrict ourselves to the case where an individual's utility depends only on his choice of action and on his type, but is independent of the full profile of actions and of the complete state of types. In other words, we consider here only the case of *private values* with *individual concern*.[5] Furthermore, individual utilities are measured in units which can be freely transferred between players: this is the assumption of quasi-linearity.

Incomplete information implies that each $i \in N$ is uncertain about the characteristics of others. In a bayesian framework however, he acts on the basis of the likelihood he assigns to the others. Thus, to each player i is associated a function p_i; $A_i \rightarrow \mathrm{D}(A_{-i})$ where $\mathrm{D}(A_{-i})$ is the set of all probability distributions defined over A_{-i}. It gives the probability $p_i(\alpha_{-i}|\alpha_i)$ attached by $i \in N$ of type $\alpha_i \in A_i$ to other players being of type $\alpha_{-i} \in A_{-i}$. The p_i functions are common knowledge; but the α_i types are private information, and the same is true for the "beliefs" $p_i(.|\alpha_i)$. We assume for simplicity[6] that all individual beliefs are derived from the same joint probability distribution p defined over A, such that:

$$p_i(\alpha_{-i}|\alpha_i) = p(\alpha_{-i}|\alpha_i) = \frac{p(\alpha_i, \alpha_{-i})}{\sum_{\alpha - i} p(\alpha_i, \alpha_{-i})}, \; i \in N.$$

(We also assume that p has the full support of A.)

The distribution p gives the *structure of beliefs*. The environment is therefore characterized by two elements: the profile of utilities $(u_i)_{i \in N}$ and the structure of beliefs p.

The strategic description is however not yet exhausted. We further assume

that each player i can select a *signal* s_i from a finite set S_i. Unlike the action d_i, s_i is made public. Individual behaviour thus has a dual dimension: while the configuration of actions $d \in D$ is not observable, that of the signals $s \in S \overset{\text{def}}{=} \mathrm{X}_{i \in N} S_i$, is public information. The assumptions underlying the game are then incorporated into a function g defined over $D \times S$ and with value in the set $\mathrm{D}(X)$ of all probability distributions over X. This function which is called the *outcome function* gives the probability $g(x|d, s)$ of getting result x conditional on the players having chosen the configuration (d, s) of actions and signals.

Player i of type α_i evaluates a configuration (d, s) on the basis of his payoffs which are given by:

$$U_i^g(d, s, \alpha_i) = \sum_x U_i(x, d_i, \alpha_i) g(x|d, s) \tag{1}$$

However, under incomplete information and in the absence of any other communication than what is carried (implicitly) by the signals, a player only knows the characteristics which are attached to his type. Player i, who knows his type α_i but doesn't know others types $\alpha_j, j \neq i$, imputes to each j a "decision rule," that is a plan or a function which we define as $(\underset{\sim}{d}_j, \underset{\sim}{s}_j) : A_j \to D_j \times S_j$.[7] This function determines a strategy $(\underset{\sim}{d}_j(\alpha_j), \underset{\sim}{s}_j(\alpha_j))$ giving the action and the signal implemented by player j if his type were α_j. Such a decision rule is called a *normalized strategy*. The set of all normalized strategies that can be imputed to j is $\underset{\sim}{D}_j \times \underset{\sim}{S}_j$ with $\underset{\sim}{D}_j \overset{\text{def}}{=} D_j^{A_j}$ and $\underset{\sim}{S}_j \overset{\text{def}}{=} S_j^{A_j}$. Player i of type α_i who has to choose a strategy $(d_i, s_i) \in D_i \times S_i$ will conjecture normalized strategies $(\underset{\sim}{d}_{-i}, \underset{\sim}{s}_{-i}) \in \underset{\sim}{D}_{-i} \times \underset{\sim}{S}_{-i}$, $\underset{\sim}{D}_{-i} \overset{\text{def}}{=} \mathrm{X}_{j \neq i} \underset{\sim}{D}_j$, $\underset{\sim}{S}_{-i} \overset{\text{def}}{=} \mathrm{X}_{j \neq i} \underset{\sim}{S}_j$, and evaluate the situation on the basis of conditional expected payoffs, also called *interim payoffs*:

$$\begin{aligned} & U_i^g((d_i, s_i), (\underset{\sim}{d}_{-i}, \underset{\sim}{s}_{-i}), \alpha_i) \\ &= \sum_{\alpha_{-i}} U_i^g(d_i, \underset{\sim}{d}_{-i}(\alpha_{-i}), s_i, \underset{\sim}{s}_{-i}(\alpha_{-i}), \alpha_i)\, p(\alpha_{-i}|\alpha_i) \\ &= \sum_{\alpha_{-i}} \left(\sum_x u_i(x, d_i, \alpha_i) g(x|d_i, \underset{\sim}{d}_{-i}(\alpha_{-i}), s_i, \underset{\sim}{s}_{-i}(\alpha_{-i})) \right) p(\alpha_{-i}|\alpha_i). \end{aligned} \tag{2}$$

We therefore have a "Bayesian game," as introduced by Harsanyi (1967–8). A choice of normalized strategies $(\underset{\sim}{d}^*, \underset{\sim}{s}^*) \in \underset{\sim}{D} \times \underset{\sim}{S}$, where $\underset{\sim}{D} = \mathrm{X}_{i \in N} \underset{\sim}{D}$, $\underset{\sim}{S} = \mathrm{X}_{i \in N} \underset{\sim}{S}$, is called a Bayesian equilibrium if: $\forall i \in N$, $\forall \alpha_i \in A_i$, $\forall d_i \in S_i$, $\forall s_i \in S_i$,

$$U_i^g((d_i, s_i), (\underset{\sim}{d}^*_{-i}, \underset{\sim}{s}^*_{-i}), \alpha_i) \leq U_i^g((\underset{\sim}{d}_i^*(\alpha_i), \underset{\sim}{s}_i^*(\alpha_i)), (\underset{\sim}{d}^*_{-i}, \underset{\sim}{s}^*_{-i}), \alpha_i).$$

The strategy $(\underset{\sim}{d}_i^*(\alpha_i), \underset{\sim}{s}_i^*(\alpha_i))$ guarantees that, in equilibrium, player i of type α_i will get the highest conditional payoff $\bar{U}_i^g(\cdot, (\underset{\sim}{d}^*_{-i}, \underset{\sim}{s}^*_{-i}), \alpha_i)$, given his

conjectures $(\underline{d}^*_{-i}, \underline{s}^*_{-i})$ on the strategies of others. In equilibrium, these conjectures are automatically satisfied. This concept generalizes to the case of incomplete information, the notion of a Nash equilibrium while incorporating a principle of "rational expectations."[8] Since a configuration $(\underline{d}^*, \underline{s}^*)$ is a Bayesian equilibrium on the basis of the *interim* payoffs given by (2), it is also a Nash equilibrium on the basis of *ex ante* payoffs. Indeed, if we denote by p_{A_i} the marginal distribution of p with respect to A_i, we have:

$$\forall i \in N, \forall \underline{d}_i \in \underline{D}_i, \forall \underline{s}_i, \in \underline{S}_i,$$

$$\sum_{\alpha_i} \bar{U}^g_i((\underline{d}_i(\alpha_i), \underline{s}_i(\alpha_i)), (\underline{d}^*_{-i}, \underline{s}^*_{-i}), \alpha_i) p_{A_i}(\alpha_i)$$

$$\leq \sum \bar{U}^g_i((\underline{d}^*_i(\alpha_i), \underline{s}^*_i(\alpha_i)), (\underline{d}^*_{-i}, \underline{s}^*_{-i}), \alpha_i) p_{A_i}(\alpha_i).$$

Collective efficiency: introducing transfers

For a given state configuration of types α, a profile $(\underline{d}, \underline{s})$ of normalized strategies in $\underline{D} \times \underline{S}$ leads to a strategic outcome $(\underline{d}(\alpha), \underline{s}(\alpha)) = (\underline{d}_i(\alpha_i), \underline{s}_i(\alpha_i))_{i \in N}$ which, through the outcome function, generates a probability distribution over the outcomes denoted $g(\cdot|\underline{d}(\alpha), \underline{s}(\alpha))$. A classical notion of collective efficiency consists of selecting a profile $(\underline{d}^*, \underline{s}^*)$ of normalized strategies such that, irrespective of the state α, the outcome $(\underline{d}^*(\alpha), \underline{s}^*(\alpha))$ is Pareto optimal for that state. Under incomplete information, other notions are possible.[9] With the hypothesis that utilities are quasi-linear, it would simply require the maximization of the collective surplus, given the configuration of the utilities $(u_i)_{i \in N}$ and the outcome function g.

Thus, a profile (of normalized strategies) $(\underline{d}^*, \underline{s}^*) \in (\underline{D} \times \underline{S})$ is said to be *collectively efficient* (with respect to $(u_i)_{i \in N}$ and g) if:

$$\forall \alpha \in A, (\underline{d}^*(\alpha)\, \underline{s}^*(\alpha)) \in \arg\max_{(d,\, s)} \sum_i U^g_i(d, s, \alpha_i) \tag{3}$$

The collective surplus, evaluated *ex post* relative to the private information (although *ex ante* with respect to the outcome) is given by:

$$\sum_{\text{base}}^{\text{top}} U^g_i(\underline{d}^*(\alpha), \underline{s}^*(\alpha), \alpha_i)$$

$$= \sum_x \left(\sum_i u_i(x, \underline{d}^*_i(\alpha_i), \alpha_i)) g(x|\underline{d}^*(\alpha), \underline{s}^*(\alpha) \right). \tag{4}$$

For the environments which are considered here, a Bayesian equilibrium does

not in general yield a collectively efficient configuration of actions and signals. This conflict of rationality leads us to question the possibility of defining "new rules of the game." Since we limit ourselves to the static case, and without introducing new possibilities of communication between players, we will only consider the instruments given by transfers of utilities in numéraire. Such transfers can be managed by a mediator. They have also to be based on observable variables only: physical results or signals, but not actions or types.

A *transfer scheme* is a function t: $X \times S \to \mathbb{R}^n$ which determines the amount of numéraire $t_i(x, s)$ which is paid or received (according to its sign) by player $i \in N$, if the outcome $x \in X$ is observed and when the signals $s \in S$ have been made public. The introduction of a transfer scheme, accepted by all players, allows us to introduce a new game with incomplete information, which is a cooperative transformation of the original game, with the payoff functions given by:

$$\sum_x (u_i(x, d_i, \alpha_i) + t_i(x, s))g(x|d, s) = U_i^g(d, s, \alpha_i) + T_i(d, s) \tag{5}$$

where we use (1) and $T_i(d, s) \stackrel{\text{def}}{=} \Sigma_x\, t_i(x, s)g(x|d, s)$. It is assumed here that the structure of beliefs is unchanged.

A transfer scheme must fulfill admissibility constraints. To achieve collective efficiency, the strongest constraint is that the whole surplus be divided amongst the players. Thus, a transfer scheme t is said to satisfy *budget balance* if:

$$\forall\, x \in X, \forall\, s \in S, \sum_i t_i(x, s) = 0. \tag{6}$$

In the rest of this chapter, we assume that this condition holds.[10]

Given a configuration of utilities $(u_i)_{i \in N}$, a structure of beliefs p and an outcome function g, a (balanced) transfer scheme t is said to implement or *support* a profile of (collectively efficient) normalized strategies $(\underline{d}^*, \underline{s}^*)$, if this profile is a Bayesian equilibrium with respect to the expected interim payoffs deduced from (5), in such a way that the following *self-selection constraints* are satisfied:

$$\forall\, i \in N, \forall\, \alpha_i \in A_i, \forall\, d_i \in S_i, \forall\, s_i \in S_i,$$

$$\sum_{\alpha_{-i}} \left(\sum_x [u_i(x, d_i, \alpha_i) + t_i(x, s_i, \underline{s}^*_{-i}(\alpha_{-i}))]g(x|d_i, \underline{d}^*_{-i}(\alpha_{-i}), s_i, \underline{s}^*_{-i}(\alpha_{-i})) \right) p(\alpha_{-i}|\alpha_i)$$

$$\leq \sum_{\alpha_{-i}} \left(\sum_x [u_i(x, \underline{d}^*_i(\alpha_i), \alpha_i) + t_i(x, \underline{s}^*(\alpha))]g(x|\underline{d}^*(\alpha), \underline{s}^*(\alpha)) \right) p(\alpha_{-i}|\alpha_i). \tag{7}$$

In what follows, we try to identify belief structures and outcome functions such that a collectively efficient structure is achieved (as a Bayesian equilibrium) by balanced transfers, whatever be the configuration of utilities. In the next section however, we see how the discussion is made simple by reducing self-selection constraints to "incentives constraints."

We have not mentioned up to now the possibility open to each agent to refuse to participate in the mechanism and be satisfied with an alternative utility level. Assuming that this level is equal to zero irrespective of the agent's type, and that the a priori expected total surplus is non-negative, it is possible to add *ex ante participation (or individual rationality) constraints*:

$$\forall i \in N, \sum_{\alpha} \left(\sum_{x} [u_i(x, \underset{\sim}{d}_i^*(\alpha_i), \alpha_i) + t_i(x, \underset{\sim}{s}^*(\alpha))] g(x | \underset{\sim}{d}^*(\alpha), \underset{\sim}{s}^*(\alpha)) \right) p(\alpha) \geq 0 \quad (8)$$

Indeed, if transfers t satisfy (6) and (7) but not (8), we can always define new transfers t' such that: $\forall x \in X$, $\forall \alpha \in A$,

$$t_i'(x, \underset{\sim}{s}^*(\alpha)) = t_i(x, \underset{\sim}{s}^*(\alpha)) - \sum_{\alpha} [U_i^g(\underset{\sim}{d}^*(\alpha), \underset{\sim}{s}^*(\alpha), \alpha_i) + T_i(\underset{\sim}{d}^*(\alpha), \underset{\sim}{s}^*(\alpha))] p(\alpha)$$

for all $i \neq 1$, and

$$t_1'(x, \underset{\sim}{s}^*(\alpha))$$

$$= t_1(x, \underset{\sim}{s}^*(\alpha)) + \sum_{i \neq 1} \sum_{\alpha} [U_i^g(\underset{\sim}{d}^*(\alpha), \underset{\sim}{s}^*(\alpha), \alpha_i) + T_i(\underset{\sim}{d}^*(\alpha), \underset{\sim}{s}^*(\alpha))] p(\alpha).$$

The transfers t' clearly satisfy (6), (7) and (8). However, the constraints (8) apply to payoffs which are *ex ante* relative to outcomes and types. Other constraints of participation (for example *interim constraints*, that are conditional on each individual's private information) are harder to satisfy. We will return to this issue.

Incentives constraints

We have so far adopted a very general formulation of the incentives problem. Given a configuration of utilities, a structure of beliefs and an outcome function, a mediator defines balanced transfers in such a way that a configuration of normalized strategies, leading to a Pareto optimal outcome in each state, be self-selected (as defined by constraints (7)). Such a problem may prove difficult to solve and the mediator is forced to rely a priori on everything that is common knowledge. It is however possible to simplify the problem, using the "revelation principle." This principle implies that players exchange information about their types, thus changing the self-selection

constraints into incentive constraints. Once we define the canonical form of the incentive constraints, the revelation principle also allows us to discuss two polar cases: models with pure moral hazard and models with pure adverse selection.

The revelation principle

Let us assume that, to determine the transfers, the mediator can collect information about the players' private characteristics. These data, which are obtained from all players simultaneously, are substitutes for the public signals. We assume that A_i itself, is the set of all messages that player i can communicate regarding his type. In this new setup, the outcome is given by an outcome function γ: $D \times A \to \mathrm{D}(X)$, where $\gamma(x|d, a)$ is the probability of observing the outcome $x \in X$, given that $d \in D$ is the course of actions that is selected and $a = (a_1, \ldots, a_i, \ldots, a_n) \in A$ is the profile of messages that are communicated. We obtain a "*revelation game*."

In this kind of game, each player $i \in N$ selects at the same time a *decision rule* $\underset{\sim}{d}_i$ which determines his action conditional on his type, and a *strategy of announcement* which is a function $\underset{\sim}{a}_i$: $A_i \to A_i$, where $\underset{\sim}{a}_i(\alpha_i) \in A_i$ is the type he declares given that $\alpha_i \in A_i$ is his true type. We denote by $\underset{\sim}{A}_i = A_i^{A_i}$ the set of all possible strategies of announcement of the player. Given a profile $(\underset{\sim}{d}_{-i}, \underset{\sim}{a}_{-i}) \in \underset{\sim}{D}_{-i} \times \underset{\sim}{A}_{-i}$, player $i \in N$ of type $\alpha_i \in A_i$ selects an action $d_i \in D_i$ and a message $\alpha_i \in A_i$, on the basis of his expected interim payoffs:

$$U_i^g((d_i, a_i), (\underset{\sim}{d}_{-i}, \underset{\sim}{a}_{-i}), \alpha_i)$$

$$= \sum_{\alpha_{-i}} \left(\sum_x u_i(x, d_i, \alpha_i)\gamma(x|d_i, \underset{\sim}{d}_{-i}(\alpha_{-i}), a_i, \underset{\sim}{a}_{-i}(\alpha_{-i})) \right) p(\alpha_{-i}|\alpha_i). \qquad (2')$$

One special announcement strategy, denoted by $\hat{a}_i \in \underset{\sim}{A}_i$, consists of player i "telling the truth," that is we have $\hat{a}_i(\alpha_i) = \alpha_i$ for all possible values of the type $\alpha_i \in A_i$ ($\hat{a}_i$ is the identity function of A_i into A_i). We now seek conditions under which $(\underset{\sim}{d}, \hat{a}) \in \underset{\sim}{D} \times \underset{\sim}{A}$ is a Bayesian equilibrium: each player is induced to reveal his private characteristics.

In a revelation game, collective efficiency leads to a profile of decision rules $\underset{\sim}{d}^* \in \underset{\sim}{D}$ satisfying:

$$\forall\, \alpha \in A,\ \underset{\sim}{d}^*(\alpha) \in \arg\max_d \sum_x \left(\sum_i u_i(x, d_i, \alpha_i) \right) \gamma(x|d, \alpha). \qquad (3')$$

A transfer scheme is now a function τ: $X \times A \to \mathbb{R}^n$ which is balanced if:

$$\forall\, x \in X,\ \forall\, \alpha \in A,\ \sum_i \tau_i(x, \alpha) = 0. \qquad (6')$$

Given a configuration of utilities $(u_i)_{i\in N}$, a structure of beliefs p and an outcome function γ, then a balanced transfer rule τ and a collectively efficient decision rule $\underset{\sim}{d}^*$ form an *incentive compatible mechanism* if the following incentives constraints hold:

$$\forall\, i \in N, \forall\, \alpha_i \in A_i, \forall\, a_i \in A_i, \forall\, d_i \in D_i,$$

$$\sum_{\alpha_{-i}} \left(\sum_x [u_i(x,\, d_i,\, \alpha_i) + \tau_i(x,\, a_i,\, \alpha_{-i})]\gamma(x|d_i,\, \underset{\sim}{d}^*_{-i}(\alpha_{-i}),\, a_i,\, \alpha_{-i}) \right) p(\alpha_{-i}\,|\alpha_i)$$

$$\leq \sum_{\alpha_{-i}} \left(\sum_x [u_i(x,\, \underset{\sim}{d}^*_i(\alpha_i),\, \alpha_i) + \tau_i(x,\, \alpha)]\gamma(x|\underset{\sim}{d}^*(\alpha),\, \alpha) \right) p(\alpha_{-i}|\alpha_i). \qquad (7')$$

In other words, $(\underset{\sim}{d}^*, \hat{a})$ is a Bayesian equilibrium of the associated game.

We can now define the *revelation principle*. Given a utility profile $(u_i)_{i\in N}$ and a structure of beliefs p, if for a given outcome function g there exists a balanced transfer scheme t which leads to a collectively efficient configuration $(\underset{\sim}{d}^*, \underset{\sim}{s}^*)$, then the assumptions that $S_i = A_i$, $\forall_i \in N$, and $g^*(\cdot|d, \alpha) = g(\cdot|d, \underset{\sim}{s}^*(\alpha))$ imply that we can find a balanced transfer scheme τ which together with $\underset{\sim}{d}^*$ constitutes an incentive compatible mechanism. Furthermore, $\underset{\sim}{d}^*$ remains collectively efficient with respect to g^*. Therefore, if the self-selection constraints are satisfied, the mediator can evaluate, on the basis of what is common knowledge, the signals that players would have communicated in equilibrium depending on their type. The mediator would obtain an incentive compatible mechanism by choosing the transfers $\tau_i(x, \alpha) = t_i\,(x_i, \underset{\sim}{s}^*(\alpha))$, proposing the outcome function g^* and recommending the decision rule $\underset{\sim}{d}^*$. No player will have an advantage in lying about his type or in disobeying secretly the recommended course of action. In fact, in a revelation game where the incentive constraints do not hold, at least one player would find it to his advantage to lie to himself before implementing his signaling strategy, or to deviate from the implementation of his individual action. The self-selection constraints would no longer be binding either.

Moral hazard and adverse selection: two polar cases

Our general model involves both incomplete information (with respect to types) and imperfect information (with respect to actions). A model of pure moral hazard corresponds to complete information and a model of pure adverse selection to perfect information.

Let us first consider the case of complete information: here, the space of types is reduced to a single element $A_i = \{\alpha_i^0\}$ for all players $i \in N_i$, and the certainty (i.e. $p(\alpha^0) = 1$) that the true types are $\alpha^0 = (\alpha_i^0, \ldots, \alpha_i^0, \ldots, \alpha_i^0)$ becomes common knowledge. Since information is complete, the expression of types can be implicitly incorporated into the utilities. The

latter can be written as functions $u_i(x_i, d_i)$ of the observable result x and the non-observable action d_i. Even the signaling strategies may be eliminated. Only the actions D remain significant. The structure of the game is now reduced to an outcome function g: $D \to \Delta(X)$ which gives the probability of getting an outcome $x \in X$ conditional on the players implementing a profile $d \in D$ of actions. *The problem of moral hazard is characterized by the remaining imperfect information with regard to actions.*

A balanced transfer scheme for a problem of moral hazard is written as:

$$t: X \to \mathbb{R}^n, \forall x \in X, \sum_i t_i(x) = 0,$$

which involves only the observable outcomes. Let us now consider when a transfer scheme t can sustain a configuration of actions $d^* \in D$, by satisfying the self-selection constraints:

$$\sum_x (u_i(x, d_i) + t_i(x))g(x|d_i, d^*_{-i}) \leq \sum_x (u_i(x, d^*_i) + t_i(x))g(x|d^*).$$

Furthermore, collective efficiency would require choosing a profile of actions such that:

$$d^* \in \arg\max_d \sum_x \left(\sum_i u_i(x, d_i)\right) g(x|d).$$

Using the arguments of the previous section, it is easy to verify that one can impose individual rationality.

A canonical example of the problem of pure moral hazard has been exhibited by Holmstrom (1982) and discussed by Radner, Myerson, and Maskin (1986), Radner, Williams (1999) or Legros and Matsushima (1991) and more recently by Fudenberg, Levine and Maskin (1994) and d'Aspremont and Gérard-Varet (1998). The n players form a team (or a partnership) and together contribute to achieve a collective outcome. This outcome is an observable "output," whereas the individual actions are non-observable "inputs." The outcome function is treated like a "production function." Members of the team must share out the monetary output of the collective action, and the division must be totally balanced. One can then derive (generic) conditions on the outcome function g which guarantee that a collectively efficient profile of actions can be based on balanced transfers, irrespective of the utilities $(u_i)_{\in N}$. The conditions given by d'Aspremont and Gérard-Varet (1998) are more general than the ones introduced by Fudenberg, Levine and Maskin (1994).

The second polar case arises when imperfect information about the

actions is removed, so that only incomplete information remains, characterizing the problem of adverse selection. Let us suppose that players' choice of actions are predetermined such that for all $i \in N$ we have $D_i = \{d_i^0\}$. The utility functions $u_i(x_i, \alpha_i)$ now implicitly incorporate the actions d_i^0 and are therefore functions of the public outcome $x \in X$ and the type $\alpha_i \in A_i$. Furthermore, the application of the revelation principle brings us to a situation where, for each player $i \in N$, the set of signals coincides with that of types, i.e. $S_i = A_i$. A mechanism, composed of an outcome function γ: $A \to \mathrm{D}(X)$ and a transfer scheme t: $X \times A \to \mathbb{R}^n$, is incentive compatible if:

$$\forall\, i \in N_i\ \forall\, \alpha_i \in A_i,\ \forall\, \alpha_i \in A_i,$$

$$\sum_{\alpha_{-i}} \left(\sum_x [u_i(x, \alpha_i) + t_i(x, \alpha_i, \alpha_{-i})]\gamma(x|\alpha_i, \alpha_{-i}) \right) p(\alpha_{-i}|\alpha_i)$$

$$\leq \sum_{\alpha_{-i}} \left(\sum_x [u_i(x, \alpha_i) + t_i(x, \alpha)]\gamma(x|\alpha) \right) p(\alpha_{-i}|\alpha_i).$$

Collective efficiency as defined above is satisfied, but can be strengthened by choosing the function γ which stipulates:

$$\forall\, \alpha \in A,\ \forall\, x \in X,\ \gamma(x|\alpha) > 0 \text{ implies } \sum_i u_i(x', \alpha_i) \leq \sum_i (x, \alpha_i),\ \forall\, x' \in X.$$

One canonical example of the problem of adverse selection is the case where X is a set of public projects and $u_i(x, \alpha_i)$ represents what player i is willing to pay for project x (Clarke (1971); Groves (1973); d'Aspremont and Gérard-Varet (1976), Green and Laffont (1970). In d'Aspremont and Gérard-Varet (1976); d'Aspremont, Crémer, and Gérard-Varet (1990), (1995) and, Pratt and Johnson Zeckhauser (1990) (generic) conditions are put on the structure of beliefs p which entail the existence of balanced transfers, irrespective of the profile of utilities and the *deterministic* rule[11] of selecting collectively efficient outcomes. Such transfers, together with the deterministic rule, form an incentive compatible mechanism.

Problems of bargaining with incomplete information provide other interesting examples (see Kennan and Wilson (1993) for a review). Let us consider two players $N = \{1, 2\}$. Player 1 has one unit of a good which he believes is worth $\alpha_1 > 0$ units of money. He wants to sell the good. Player 2 attributes the value $\alpha_2 > 0$ to the ownership of the good, and is the buyer. Both α_1 and α_2 are private information. Utilities[12] are given by $u_1(1, \alpha_1) = \alpha_1$ and $u_2(1, \alpha_2) = \alpha_2$ (with $u_1(0, \alpha_1) = u_2(0, \alpha_2) = 0$). A mechanism is a function γ: $A_1 \times A_2 \to [0, 1]$ giving the probability $\gamma(\alpha_1, \alpha_2)$ that a trade will take place, given the players'

announced valuations α_1 and α_2 respectively. The mechanism is completed[13] with a transfer scheme t: $A_1 \times A_2 \to \mathbb{R}$, where $t(a_1, a_2)$ is the monetary transfer from the buyer to the seller, given their announced valuations α_1 and α_2. The incentive constraints are:

$$\forall\, \alpha_1 \in A_1,\ \forall\, a_1 \in A_1,$$

$$\sum_{\alpha_2} (t(a_1, \alpha_2) - \gamma(a_1, \alpha_2)\alpha_1) p(\alpha_2|\alpha_1) \leq \sum_{\alpha_2} (t(\alpha_1, \alpha_2) - \gamma(\alpha_1, \alpha_2)\alpha_1) p(\alpha_2|\alpha_1)$$

$$\forall\, \alpha_2 \in A_2,\ \forall\, a_2 \in A_2,$$

$$\sum_{\alpha 1} (\gamma(\alpha_1, a_2)\alpha_2 - t(\alpha_1, a_2)) p(\alpha_1|\alpha_2) \leq \sum_{\alpha_1} (\gamma(\alpha_1, \alpha_2)\alpha_2 - t(\alpha_1, \alpha_2)) p(\alpha_1|\alpha_2).$$

Furthermore, collective efficiency imposes $\gamma(a_1, a_2) = 1$ (resp. $= 0$) if and only if $a_2 - a_1 > 0$ (resp. < 0).

This simple model of bargaining between a buyer and a seller highlights an inherent difficulty with any problem of adverse selection. Even under private values and with transferable utilities, individual incentives can be inconsistent with the *interim* individual rationality constraints:

$$\forall\, i \in N,\ \forall\, \alpha_i \in A_i,$$

$$\sum_{\alpha_{-i}} \left(\sum_x [u_i(x, \alpha_i) + t_i(x, \alpha)] \gamma(x|\alpha) \right) p(\alpha_{-i}|\alpha_i) \geq u^0{}_i(\alpha_i).$$

where $u_i^0(\alpha_i)$ is the reservation valuation of player i of type α_i. This is in fact the core of the argument in Myerson and Satterthwaite (1983). Examples in other contexts are provided in d'Aspremont and Gérard-Varet (1979b) or Laffont and Maskin (1979).

Environments where individual incentives and collective efficiency are compatible

The general model we consider here is an extension of the classical problems of moral hazard and adverse selection. These two problems are special cases of our model. It is possible to identify certain conditions on beliefs (which characterize adverse selection) as well as the outcome function (characterizing moral hazard) which, at least given transferable individual utilities and private values, guarantee the compatibility of collective efficiency and individual incentives. In this section, we present these conditions, using duality arguments. The discussion of the previous section allows us to restrict to a revelation game.

A necessary and sufficient condition

Let us consider a utility profile $(u_i)_{i\in N}$ and a structure of beliefs p. Let $\underset{\sim}{d}^* : A \to D$ be a decision rule and $\gamma : D \times A \to \Delta(X)$ an outcome function. Saying that there exist balanced transfers which form an incentive compatible mechanism together with $\underset{\sim}{d}^*$ is equivalent to solving a linear system of inequalities in the transfers derived from (6′) and (7′), that is:

$$\exists\, \tau \in \mathbb{R}^{X\times A} \text{ such that } \forall\, i \in N,\ \forall \alpha_i \in A_i,\ \forall a_i \in A_i,\ \forall \mathrm{d}_i \in \mathrm{D}_i|$$

$$\sum_{\alpha_{-i}} \left(\sum_{x} [\tau_i(x, \alpha)\gamma(x|\underset{\sim}{d}^*(\alpha),\alpha) - \tau_i(x, a_i, \alpha_{-i})\gamma(x|d_i, \underset{\sim}{d}^*_{-i}(\alpha_{-i}), a_i, \alpha_{-i})] \right) p(\alpha_{-i}|\alpha_i)$$

$$\geq \sum_{\alpha_{-i}} \left(\sum_{x} [u_i(x, \underset{\sim}{d}_i(\alpha_i), \alpha_i)\gamma(x|\mathrm{d}_i|, \underset{\sim}{d}^*_{-i}(\alpha_i), a_i, \alpha_{-i}) - u_i(x, \underset{\sim}{d}^*_i, (\alpha_i), \alpha_i)\gamma(x|d^*(\alpha), \alpha)] \right) p(\alpha_{-i}|\alpha_i)$$

and

$$\forall\, \alpha \in A,\ \forall\, x \in X,\ \sum_i \tau_i(x, \alpha) = 0.$$

Duality theorems (or the "theorem of the alternative," see Fan (1956) for example) can be used to derive a necessary and sufficient condition.

Denote by $\lambda_i(\alpha_i, a_i\, \mathrm{d}_i) \geq 0$, $i \in N$, $\mathrm{d}_i \in \mathrm{D}_i$, the dual variables associated with the incentive constraints, and $\mu(\alpha, x)$, $\alpha \in X$, $x \in A$, the dual variables associated with the budget constraints. The necessary and sufficient condition can be written as: either there is no $\lambda \in \mathrm{X}_{i\in N}\, \mathbb{R}^{A_i\times D_i}$, $\lambda \neq 0$, and no $\mu \in \mathbb{R}^{X\times A}$ such that:

$$\forall\, i \in N,\ \forall\, \alpha \in A,\ \forall\, x \in X,$$

$$\gamma(\mathrm{x}|\underset{\sim}{d}^*(\alpha),\alpha)p(\alpha_{-i}|\alpha_i) \sum_{a_i} \sum_{d_i} \lambda_i(a_i, \alpha_i, d_I)$$

$$- \sum_{a_i} \sum_{d_i} \lambda_i(\alpha_i, \alpha_i\, d_i)\gamma(x|d_i, \underset{\sim}{d}^*_{-i}(\alpha_i)\, \alpha)p(\alpha-_i|\alpha i) + \mu(x, \alpha) = 0 \qquad (8')$$

or we have:

$$\sum_i \left\{ \sum_{\alpha} \cdot \sum_{x} \sum_{a_i} \sum_{d_i} \lambda_i(a_i, \alpha_i, d_i)[u_i(x_i, d_i, \alpha_i)\gamma(x|d_i, \underset{\sim}{d}^*_{-i}(\alpha_{-i}), \alpha_{-i},\alpha_{-i}) \right.$$
$$\left. -u_i(x, \underset{\sim}{d}^*_i(\alpha_i),\alpha_i)\gamma(x|\underset{\sim}{d}^*(\alpha),\alpha)]p(\alpha_{-i}\,|\alpha_i) \right\} \leq 0. \qquad (9')$$

This necessary and sufficient condition, for the existence of balanced and incentive compatible transfers is about all variables simultaneously (utilities, beliefs, the outcome function). We shall now concentrate on beliefs and the outcome function.

An informational based condition

Apart from covering all the elements of the problem, the last condition does not take advantage of the requirement of collective efficiency. Such a requirement, given a configuration of utilities $(u_i)_{i\in N}$, implies in the present context the choice of a decision rule $\underset{\sim}{d}^*$ satisfying (3′).

We will now consider a condition on beliefs and the outcome function which will ensure the possibility of setting up an incentive compatible and collectively efficient mechanism, with balanced transfers, irrespective of the utility profile $(u_i)_{i\in N}$.

> Condition $C^*_{\underset{\sim}{d}^*}$:
> Let $\underset{\sim}{d}^* \in \underset{\sim}{D}$ be a a profile of decision rules. A structure of beliefs p and an outcome function γ satisfy the following condition:
> either there is no $\lambda \in \mathrm{X}_{i\in N}\, \mathbb{R}^2_i$, $\lambda \neq 0$, such that (8′) is satisfied,

or we have:

$$\forall\, i \in N, \forall\, \alpha \in A, \forall\, x \in X$$

$$\gamma(x|\underset{\sim}{d}^*(\alpha), \alpha)p(\alpha_{-i}|\alpha_i) \sum_{(\alpha_i,\alpha_i)} \sum_{d_i} \lambda_i(a_i, \alpha_i, d_i)$$

$$-\sum_{(\alpha_i,\alpha_i)} \sum_{d_i} \lambda_i(\alpha_i,\alpha_i, \mathrm{d}_i)\gamma(x|d_i, \underset{\sim}{d}^*_{-i}(\alpha_{-i}),\alpha)p(\alpha_{-i}|\alpha_{-i}) = 0 \qquad (10)$$

We have the following:

> *Theorem*: Assume that the structure of beliefs p and the outcome function γ satisfy condition $C^*_{\underset{\sim}{d}^*}$ for a profile of decision rules $\underset{\sim}{d}^* \in \underset{\sim}{D}$ and that $\gamma\,(x|\mathrm{d},\alpha)$ is constant in α. Then for any utility profile $(u_i)_{i\in N}$ for which the rule $\underset{\sim}{d}^*$ is collectively efficient given γ, there exist balanced transfers which, together with this rule, form an incentive compatible, balanced, ex ante individually rational and collectively efficient mechanism.

This result may be proved in many ways. One is to show that condition $C^*_{\underset{\sim}{d}^*}$ implies the previous necessary and sufficient condition, using collective efficiency.[14] Another is to use a second formulation,[15] which is equivalent to the condition $C^*_{\underset{\sim}{d}^*}$, but is now its "primal" version.[16]

Condition $C_{\underset{\sim}{d}^*}$: Let $\underset{\sim}{d}^* \in \underset{\sim}{D}$ be a profile of decision rules. For any function

$$\beta : X \times A \to \mathbb{R}, \exists \tau \in \mathbb{R}^{X \times A}, \text{ such that } \forall i \in N, \forall a_i \in A_i, \forall \alpha_i \in A_i, \forall d_i \in D_i,$$

$$\sum_{\alpha_{-i}} \sum_{x} [\tau_i(x, \alpha)\gamma(x|\underset{\sim}{d}^*(\alpha),\alpha) - \tau_i(x,a_i, \alpha_{-i})\, \gamma(x|d_i, \underset{\sim}{d}^*_{-i}(\alpha_{-i}), a_i\, \alpha_{-i})]p\,(\alpha_{-i}|\alpha_i) \geq 0$$

$$\text{and, } \forall \alpha \in A, \forall x \in X, \sum_i \tau_i(x,\alpha) = \beta(x,\alpha).$$

The proof goes as follows. Let $(u_i)_{i \in N}$ be a profile of utility functions and $\underset{\sim}{d}^* \in \underset{\sim}{D}$, a profile of decision rules which is collectively efficient for an outcome function γ. Following Clarke, Groves and Vickrey, we introduce the transfers given by:

$$t_i^0(x,\alpha) \overset{\text{def}}{=} \sum_{j \neq i} u_j(x, \underset{\sim}{d}^*_j(\alpha_j), \alpha_j)\tilde{\gamma}(x|\underset{\sim}{d}^*(\alpha)),$$

with $\tilde{\gamma}(x|\underset{\sim}{d}^*(\alpha)) = \gamma\,(x|\underset{\sim}{d}^*(\alpha), \alpha)$, assumed to be constant in α.

Collective efficiency means that such transfers generate an incentive compatible mechanism.[17]

Clearly, there is no guarantee that the transfers are balanced: $\Sigma_i t_i^0(x, \alpha)$ may be either positive or negative. The condition $C_{\underset{\sim}{d}^*}$ in fact states that the total, be it a surplus or a deficit, may be allocated amongst all players in a way which is incentive compatible. All we need is the requirement that:

$$\beta(x, \alpha) = -\sum_i t_i^0(x, \alpha),$$

and the application of the condition: the transfers $t = \tau + t^0$ are both balanced and incentive compatible. Finally, the earlier discussion may be repeated to obtain ex ante individual rationality.

Thus, condition $C_{\underset{\sim}{d}^*}$ is sufficient to obtain balanced transfers that allow an efficient and incentive compatible mechanism for an efficient profile of decision rules. It is possible to strengthen this property by considering a stronger condition that holds *irrespective* of the decision rule $\underset{\sim}{d}^*$ likely to be selected by introducing a condition similar to condition B̄ in d'Aspremont and Génard-Varet (1982, 1998).

The general model presented here can be used to analyze many particular situations, relevant to markets and organizations. Examples include relations between the regulator and the management of a public enterprise (or between

the shareholders and the management of a private enterprise), those between producers and consumers of a public good, tendering or auctioning procedures, insurance contracts, and even employment contracts between an employer and an employee. Challier and Gérard-Varet (1994) follow the same lines of argument for a discussion in the context of employment contracts.

It is however important to recognize that the conditions under which incentive compatible and collectively efficient (in its classical sense) mechanisms exist in the presence of moral hazard and adverse selection, are crucially linked to two central assumptions: "transferability of utilities" and "private values."

To begin with, even when utilities are transferable, the application of the condition $C_{\mathcal{E}^*}$ and the existence of incentive compatible and collectively efficient mechanisms are put into question by the introduction of common values. This is indeed the main lesson to be drawn from the famous example in Akerlof (1970). However, d'Aspremont and Gérard-Varet (1982, 1998) suggest a stronger condition for the existence of incentive compatible mechanisms in this case (condition B). The ensuing mechanisms do not have to satisfy efficiency in its traditional sense. As suggested in Forges (1994), one may indeed question whether the classical definition of efficiency is relevant outside a world of private values.

The assumption of transferable utilities opens the possibility of using transfers between players as an instrument to fulfill individual incentives. To abandon it, would not only eliminate this possibility, but also introduce all the difficulties, which we have ignored here, and which relate to the combination of incentives and risk sharing. One may envisage adopting a sequential approach. In the case of pure moral hazard, "repetition" helps reconcile collective efficiency with the principles of non-cooperative individual behaviour (see Fudenberg, Levine, and Maskin 1994, or d'Aspremont and Gérard-Varet 1998). However, with adverse selection as well as moral hazard, the question is no longer so simple, mainly due to problems of individual rationality under incomplete information. Furthermore, if there is repetition, it would be difficult to ignore the selection of equilibrium on the basis of a criterion of "sequential rationality" which implies that each player's strategy is the best response to the strategies of others, not only during the course of the game, but also for every possible contingency he may find himself in and for the rest of the game.

A minimal sequential approach, without going as far as repetition, combined with some coordination can be introduced through "communication mechanisms" as suggested in Forges (1986) and Myerson (1985, 1991). In this case, one has to consider a (finite) set of messages R_i that each player $i \in N$ receives from the mechanism. These messages are distinct from the signals S_i the player transmits. Let $R = \mathrm{X}_{i \in N} R_i$ be the set of all possible configurations of messages, and $S = \mathrm{X}_{i \in N} S_i$ the set of all possible configurations of signals. A *communication mechanism* is a function π of S to the set $\mathrm{D}(R)$ of all possible probability distributions over R.

We can add communication mechanisms to our model. Let D_i: $R_i \to D_i$ be a strategy that player i chooses (privately) on the basis of a message τ_i communicated to him (secretly). The resulting action is denoted by $\delta_i(r_i)$. Let $D_i \stackrel{\text{def}}{=} D_i^{R_i}$ be the set of all such strategies and assume $D = X_{i \in N} D_i$. Given a communication scheme π and an outcome function g, we can form a new outcome function:

$$g^{\pi}(x|\delta, s) = \sum_{r} g(x|\delta\ (r), s)\ \pi\ (r|s)$$

The analysis may then continue with this outcome function. The self-selection constraints now apply to decision rules $\underset{\sim}{s}$: $A \to S$ and $\underset{\sim}{d}$: $A \to \delta$, and the "revelation principle" allows the identification of A_i and Δ_i on the one hand, and R_i and D_i on the other. We are then back to the incentive constraints discussed by Myerson (1991). In fact, we have considered here communication relative to "inputs" (signals), whereas Myerson (1991) argues the case for extending communication to "outputs" (messages).

A basic conflict lies at the heart of the question of combining collective efficiency with individual incentives. Individual incentives depend on what each person knows privately, whereas collective efficiency assumes the pooling of this information. A second best approach, called "interim incentive efficiency," has been put forward to deal with this difficulty (Holmstrom and Myerson 1983). An incentive compatible mechanism (in general a mechanism which is supported by a non-cooperative equilibrium) is efficient according to the latter criterion if there exists no other incentive compatible mechanism (supported by a non-cooperative equilibrium) which dominates it uniformly as far as the interim payoffs are concerned. However, it is possible for such mechanisms to select Pareto dominated outcomes, once the intrinsic communication has taken place (Forges 1994).

We must underline finally the limitations of the revelation principle upon which we have relied so much. This principle indicates that, by studying the revelation mechanisms, it is possible to determine an efficiency frontier in the set of all feasible mechanisms for which, in equilibrium, each agent reveals sincerely his private information. It does not however preclude the dependence of efficient mechanisms on all elements which are common knowledge, but often contingent, such as the form of beliefs or individual utilities, maybe even the number of participants. In practice, the question is to find a procedure – either static like an auction or dynamic such as a negotiation – and equilibrium strategies which allow an efficient mechanism to be sustained (or "implemented"). The procedures which are actually used, often lack such contingency. Forges (1990) defines a general class of "universal mechanisms." As underlined by Wilson (1985), revelation mechanisms use complex rules, taking contingencies into account, but lead to an equilibrium which is easy to calculate ("to tell the truth"), whereas procedures used in practice have simple

rules, though the players on the other hand must have complex equilibrium strategies, which integrate all contingencies they know about. Theory will gain much by taking more into consideration such "games" where practical organizational procedures determine the rules.

Notes

1 The notion of "conjecture" may have different meanings. We take here the more restrictive where the information structure of the game is common knowledge, and where each player can predict other players' strategies at equilibrium.
2 We will come back to this issue. See Forges (1994) for a more detailed discussion.
3 See Myerson (1985) for a discussion and more references.
4 The literature refers often to models with "hidden information" or "hidden action." A chosen action d_i is hidden if it is not publicly observable. Similarly, a type $\alpha_i \in A_i$ is hidden if it cannot be verified on the basis of common knowledge.
5 In the case of *common values* with *collective concern*, utilities would take the form $u_i(x, d, \alpha)$. The question of collective concern (as regards actions) is less problematic than that of common values (as regards types).
6 The properties derived from the model do not depend on this assumption.
7 We assume, without loss of generality, deterministic decision rules. One could consider functions of A_j mapped into the set D (D_j ò S_j) of probability distributions over D_j ò S_j.
8 See the discussion on this framework in d'Aspremont, Dos Santos Ferreira, and Gérard-Varet (1995).
9 See Holmström and Myerson (1983), Forges (1994).
10 An alternative condition, called *expected budget balance* postulates that: for every d and every s, $\Sigma_i T_i(d, s) = 0$. This is a weaker condition, because it only balances the budget on average with respect to the distribution given by g. It does allow the reinforcement of later results. See d'Aspremont and Gérard-Varet (1998).
11 Unlike our general model where *stochastic* mechanisms are considered, most of the literature has discussed *deterministic* mechanisms, i.e. functions $g: A \to X$. In that case, transfers are given by a function $t: A \to \mathbb{R}^n$. Johnson *et al.* (1990) introduce observable actions and discuss the case of common values.
12 We follow here Myerson and Satterthwaite (1983) in taking a model with private values. Akerlof (1970) considers a very different model with common values, where $u_2(1, \alpha_1) = \alpha_1$ and $u_2(1, \alpha_2) = \alpha_1$, α_2 being irrelevant.
13 This model can also be applied to sequential mechanisms. The probability $\gamma(\alpha_1, \alpha_2)$ is interpreted as the probability of reaching arrangement discounted on the basis of the time period in which it takes place and $t(\alpha_1, \alpha_2)$ is the expected present value of the transfers. Thus, when a procedure is likely to go onto infinity, and if after exchange the gains are discounted at the rate δ, we have $\gamma(\alpha_1, \alpha_2) = \delta^{\tau(\alpha_1, \alpha_2)}$, with $\tau(\alpha_1, \alpha_2)$ being the period during which transaction takes place (when $\tau(\alpha_1, \alpha_2) = ¥$, exchange never takes place). Such an interpretation is acceptable when the negotiating parties continue to bargain as long as there are unexploited gains from trade.
14 This method is used in d'Aspremont and Gérard-Varet (1979, 1998) respectively for the problem of pure adverse selection and that of pure moral hazard.
15 See d'Aspremont, Crémer and Gérard-Varet (1995).
16 To see this, simply apply the theorem of the alternative.
17 In fact a strongly incentive compatible mechanism: this property continues to hold with dominant strategies. This is the point made by Groves (1973).

References

Akerlof, G. (1970). The market for lemons: qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84, 488–500.

Alchian, A. A. and H. Demsetz (1972). Production, information costs and economic organization. *American Economic Review*, 62, 777–85.

d'Aspremont, C. and L. A. Gérard-Varet (1976). Un modèle de négociation pour certains problèmes internationaux de pollution. *Revue d'Economic Politique*, 4, 547–62.

d'Aspremont, C. and L. A. Gérard-Varet (1979). Incentives and incomplete information. *Journal of Public Economics*, 11, 25–45.

d'Aspremont, C. and L. A. Gérard-Varet (1979b). On Bayesian incentive compatible mechanisms. In J. J. Laffont (ed.), *Aggregation and Revelation of Preferences*. North Holland, Amsterdam, 269–88.

d'Aspremont, C. and L. A. Gérard-Varet (1990). Mécanismes de marchés, concurrence et contraintes d'incitation. In J. Cartelier (ed.), *La formation des grouleurs économiques*. PUF, Paris, 83–126.

d'Aspremont, C., Crémer, J., and L. A. Gérard-Varet (1990). Incentives and the existence of Pareto-optimal revelation mechanisms. *Journal of Economic Theory*, 51, 233–54.

d'Aspremont, C., Crémer, J., and L. A. Gérard-Varet (1995). Correlation, independence and Bayesian incentives. Mimeo.

d'Aspremont, C., Dos Santos Ferreira, R., and L. A. Gérard-Varet (1995). Fondements stratégiques de l'équilibre en économie. Coordination, rationalité individuelle et anticipations. In L. A. Gérard-Varet and J. C. Passeron (eds), *Le modèle et l'enquête*. Editions de l'EHESS, Paris, 447–68.

d'Aspremont, C. and L. A. Gérard-Varet (1998). Linear methods to enforce partnerships under uncertainty: an overview. *Games and Economic Behavior*, 25, 311–36.

Challier, M. C. and L. A. Gérard-Varet (1994). Incentive payment schemes in teams with moral hazard and adverse selection. Mimeo.

Clarke, E. (1971). Multi-part pricing of public goods. *Public Choice*, 11, 17–33.

Drèze, J. H. and D. de la Vallée Poussin (1971). A tatonnement process for public goods. *Review of Economic Studies*, 38, 133–50.

Fan, K. (1956). On systems of inequalities. In H.W. Kuhn and A.W. Tucker (eds), *Linear inequalities and related systems*. Princeton University Press, Princeton, NJ, 99–156.

Fozges, F. (1986). An approach to communication equilibria. *Econometrica*, 54, 1375–85.

Forges, F. (1990). Universal mechanisms. *Econometrica*, 58, 1341–64.

Forges, F. (1994). Posterior efficiency. *Games and Economic Behavior*, 6, 238–61.

Fudenberg, D., Levine, D., and E. Maskin (1994). The Folk theorem with imperfect public information. *Econometrica*, 62(5), 997–1039.

Green, J. and J. J. Laffont (1979). *Incentives in Public Decision Making*. North Holland, Amsterdam.

Grossman, S., and O. D. Hart (1983). An analysis of the principal agent problem. *Econometrica*, 51, 7–45.

Groves, T. (1973). Incentives in teams. *Econometrica*, 31, 617–63.

Harsanyi, J. C. (1967–8). Games with incomplete information played by Bayesian players. *Management Science*, 14, 159–89, 320–34, 486–502.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 13, 324–40.
Holmstrom, B. and R. B. Myerson (1983). Efficient and durable decision rules with incomplete information. *Econometrica*, 51, 1799–819.
Hurwicz, L. J. (1960). Optimality and informational efficiency in resource allocation process. In K. Arrow, S. Karlin, and P. Suppes (eds), *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford, CA, 27–46.
Hurwicz, L. J. (1972). On informationally decentralized systems. In R. Radner and C. B. McGuire (eds), *Decision and Organization. A Volume in Honor of Jacob Marschak*. North Holland, Amsterdam, 297–336.
Johnson, S., J. W. Pratt and R. Zeckhauser (1990). Efficiency despite mutually payoff-relevant private information: the finite case. *Econometrica*, 4, 873–900.
Kennan, J. and R. Wilson (1993). Bargaining with private information. *Journal of Economic Literature*, 31, 45–104.
Laffont, J. J. and E. Maskin (1979). Differential approach to expected utility maximizing mechanisms. In J. J. Laffont (ed.), *Aggregation and Revelation of Preferences*. North Holland, Amsterdam, 289–305.
Legros, P. and H. Matsushima (1991). Efficiency in partnership. *Journal of Economic Theory*, 55, 296–322.
Malinvaud, E. (1971). A planning approach to the public good problem. *Swedish Journal of Economics*, 173, 96–112.
Myerson, R. B. (1985). Bayesian equilibrium and incentive compatibility. In L. Hurwicz, D. Schmeidler and H. Sonnenschein (eds), *Social Goals and Social Organization*. Cambridge University Press, Cambridge, 229–59.
Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*, Harvard University Press. Cambridge, MA.
Myerson, R. B. and M. A. Satterthwaite (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29, 265–81.
Radner, R., R. Myerson and E. Maskin (1986). An example of a repeated partnership game with discounting and with uniformly inefficient equilibria. *Review of Economic Studies*, 53, 59–69.
Radner, R. and S. R. Williams (1994). Efficiency in partnership when the joint output is uncertain. In J. Ledyard (ed.), *The Economics of Informational Decentralization*. Kluwer, London.
Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, 36, 387–9.
Vickrey, W. (1961). Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance*, 16, 8–37.
von Hayek, F. (1945). The use of knowledge in society. *American Economic Review*, 35, 519–30.
Wicksell, K. (1896). A new principle of just taxation. *Finanz Theoretische, untersuclungen und das steuerwesnschweden's.* Iena, Germany. Reprinted in: R. A. Musgrove and A. T. Peacock (eds), *Classics in the Theory of Public Finance*. St Martin's Press, New York, 1958, 72–118.
Wilson, R. (1985). Efficient trading. In G. Feiwel (ed.), *Issues of Contemporary Microeconomics and Welfare*. State University of New York Press, Albany, 169–208.

9 Team models as a framework to analyze coordination problems within the firm

Jean-Pierre Ponssard, Sébastien Steinmetz, and Hervé Tanguy

Introduction

Team theory is a meaningful way to address important questions about coordination activities within the firm. This chapter elaborates on the issue of defining an efficient plan of decentralized actions in a firm organized according to a functional structure.

The first team-theoretic models were historically introduced by Schelling (1960) as pure coordination games along with the notion of focal point. Kreps (1984) used this notion to emphasize the role of a common culture in solving coordination problems within an organization. In order for a focal point to emerge among decentralized agents, the coordination rules representing Kreps's vision of corporate culture need to present some properties such as simplicity, generalizability, and transmissibility. Following these considerations, formal developments led to advances in two directions: the dynamic construction of a focal point (Crawford and Haller 1990) and the role of conventions to overcome the eventual ambiguities of this construction (Kramarz 1994). The relevance of these formalizations is confirmed by some empirical works (Broseta 1993; Mayer *et al.* 1992). The experimental results reported in this chapter go in that direction as well.

In a more technical vein, the theory of teams as formally introduced by Marschak and Radner (1972) provides a general framework to analyze the role of incomplete information in the firm coordination problems. Using this framework, Crémer (1980) formalizes the choice of an organizational structure as originally discussed in Chandler (1962). For a given organizational structure, Crémer's model allows us to distinguish between ex post transfers between shops belonging to the same unit and ex ante transfers between these units. Aoki (1986) uses the same formalism in his discussion of vertical and horizontal coordination. This chapter will also draw on this formalism to elaborate a model of a functional structure.

Such formal contributions to the theory of the firm do not attribute to each agent a utility function of its own. They are, in a way, non-standard approaches to most economists. This is certainly one reason why, when compared with the abundant literature concerning incentives in organizations (see

for instance Hart and Homstrom 1987), these works are still relatively isolated in economics textbooks.

On the contrary, they would appear natural in management science where the problem of internal coordination within the organization is featured as one of coordinating several departments or services sharing the same objective function.[1] According to this view, when it faces an unpredictable environment, the goal of the team is to conciliate two conflicting objectives: on the one hand, high decentralized reactivity and, on the other hand, global coherence between local decisions. The main difficulty of such a task arises from the dispersion of information within the organization, together with the technical constraints inherent to a joint production system. Delay, lack of focus and imperfections in the reported information explain the loss of optimality rather than pure opportunistic behavior of the agents.

This view does not mean that the question of incentives is completely forgotten: It is in fact possible to adopt a basic principle recalled in Milgrom and Roberts (1992: ch. 4) according to which a good understanding of coordination requirements is a prerequisite to the design of any coherent incentive scheme. Indeed, there are many empirical studies describing the perverse effects that may accompany the use of simplistic indicators which neglect the underlying joint coordination constraints (see for instance Johnson and Kaplan 1987, concerning the limits of cost accounting).

Within a functional structure, these coordination processes usually follow physical flows. Each division (purchasing, production, sales, etc.) is constrained in throughputs by other internal divisions while adapting to local uncertainties (labor, breakdown of equipment, input and output changes with regard to the firm environment, etc.). Routines are used to balance ex ante budgets and define buffer inventories. As discussed in detail in Chandler about the successive organizational changes which occurred in Du Pont de Nemours from 1907 to 1920, it is only when coordination within a functional firm becomes clearly inefficient that it is broken down into parts and that a multidivisional structure emerges. Then much simpler coordination procedures can be implemented under the authority of the division manager while the benefits of the specialization by function are lost (economies of scope).

The main objective of this chapter is to capture the difficulties associated with the coordination of activities in a functional structure. Consequently, due attention is given to the underlying production process, introducing relevant specific details. The proposed model is in fact inspired by an analysis of actual coordination processes in companies operating in the Champagne sector (Ponssard and Tanguy 1993). Such firms have simple functional structures (purchasing grapes, elaboration of the Champagne wines through mixing production of different vintages and of different years, marketing and sales, etc). They are confronted with important uncertainties (prices and quantities vary considerably from year to year, etc). Empirical observation leads to the fact that two main generic policies can be used to efficiently coordinate

decentralized actions in such firms: a "growth policy" and a "speculation policy." The first policy relies on internal cost minimization to absorb external uncertainties, while the second one adopts more flexible production schedules to benefit from favorable market conditions. The proper identification and the assessment of the relative efficiency of these two policies appear as an important managerial question.

The proposed model illustrates well the existence of the two generic policies and the difficulty in ranking them. These can be associated with two Nash equilibria of a related team model. It is also proved that there are many other Nash equilibria, but these are all dominated by one of the two generic policies. An experimental game has also been elaborated in connection with this model (de Jaegere and Ponssard 1990). The results of this experimental game make it clear that the mere identification of a generic policy is far from obvious, emphasizing the general requirements needed to properly identify a focal point. More specifically, it is argued that model building can precisely help in this matter, enhancing a shared management culture within the firm.

The rest of the chapter is organized as follows. First, we present the model and solve it. Second, we provide a comparison between the theoretical results of the model and the outcomes observed in the experimental game.

A team model for simple functional structures

Consider a stylized firm with a functional structure consisting of a purchasing, a manufacturing, and a selling division. It is assumed that manufacturing involves fixed costs, so that it pays to increase the volume of throughput. However, the environment is assumed to be uncertain both in the downstream and upstream markets. This suggests that corresponding decisions should be flexible in order to benefit from favorable market conditions. Altogether these conflicting objectives may endanger the financial position of the firm through over-stocking and/or high cost inefficiency. How can one decentralize decisions in such an organization?

The model

The model implements some drastic simplifications while retaining the difficulties of the coordination process. Only two divisions are introduced: purchasing/manufacturing on one hand, and selling on the other hand. These two divisions are respectively denoted by M and S. Further simplications are now detailed.

Production function and technical constraints

The activity of the firm consists in buying an input at market price p_m, transforming it into a final good which can be sold at market price p_s. The production process is characterized by a fixed cost F. It transforms one unit of input

into one unit of output[2] and requires a production cycle of, say, one period. The whole quantity of input bought at the price p_m is immediately engaged in the production process: There are no stocks of inputs and no rejects. At the end of the production cycle, the final good can either be sold in the retail market at the current price p_s, or be kept in stock in order to be sold at a future time.

Due to the production cycle, at a given moment the firm can only sell the quantity s of final good that was produced in previous periods. Besides this physical stock, the firm is also endowed with a financial stock denoted t, that it can use to finance the production of the period. Contrary to physical transfers, financial transfers inside the firms are supposed to be instantaneous (no customer or supplier credit), so that the returns from the sales of the current period can also be used to finance production. The couple (s, t) represents thus the initial state of the organization. We assume that the firm operates on the sole basis of its physical and monetary wealth.

These assumptions restrict the operational decisions that the firm can make. Let Q_m be the quantity of input bought at price p_m and Q_s the quantity of output sold at price p_s. The first technical constraint reflects that the firm has no access to an outside market for the final good:

$$0 \leq Q_s \leq s. \tag{1}$$

The second constraint is that the firm can only use its internal financial resources. The production decision is then solely based on the cash currently available, t, and the returns from the sales of the current period $p_s Q_s$. For $Q_m > 0$, this writes as:

$$p_m Q_m + F \leq t + p_s Q_s. \tag{2}$$

Given the decisions Q_s and Q_m, then at the end of the period the firm is in a new state characterized as:

$$s' = s - Q_s + Q_m,$$
$$t' = t + p_s Q_s - p_m Q_m - F.$$

Information and decision structures of the firm

The information and decision structures of the firm are decentralized: knowing only the price p_m, division M decides on the quantity of input that is bought (and transformed) during the period, and division S, knowing only the price p_s, decides on the quantity of final good that will be retailed. We assume that the market prices p_m and p_s are independent random variables, with respective laws and supports $\mathcal{L}_m$ over $[p_m^l, p_m^H]$ and $\mathcal{L}_s$ over $[p_s^l, p_s^H]$. The average selling price is $\bar{p}_s$.

Payoff function of the team

At the end of the period, the firm is evaluated on the basis of the wealth created and the depreciation of the assets which are not utilized. Analytically, the payoff function of the game is thus defined as follows.

Recall that s' and t' are the values of the physical and financial assets at the end of the period. Then, with δ as the discount rate, the ex post payoff function is:[3]

$$\Pi = \bar{p}_s s' + t' - (1-\delta)\,\bar{p}_s(s - Q_s) - (1-\delta)(t - p_m Q_m - F)^+. \tag{3}$$

If inequality (1) or (2) is violated, the payoff is set to minus infinity.

The ex ante payoff function is the expectation of the ex post payoff with respect to the probability laws of the prices, that is[4] $\mathcal{L}\,(\Pi\,(Q_m, Q_s))$, where $\mathcal{L}$ designates the product law of $\mathcal{L}_m$ and $\mathcal{L}_s$. The game is then in normal form. Due to the infinite penalty associated with bankruptcy, the expected payoff takes a finite value only for the strategies that respect the technical constraints for every realization of the prices. Those strategies are called *admissible*.

Solving the model

Finding the optimal policy comes down to determining the best Nash equilibrium of the incomplete information game, that is, the one which gives the highest ex ante payoff to the team. First we show that the equilibria of the game can be parametrized by the minimum revenue of the sales (proposition 2). This parameter can be interpreted as an ex ante financial transfer from division *S* to division *M*. Second, we determine the optimal way of fixing such an ex ante transfer (proposition 3). Due to the increasing returns to scale, the optimal policy is either no transfer or maximum transfer between the two divisions. Proofs of the propositions are given in Appendix 9.1.

Decentralizing the technical interface

One way to decentralize the technical interface corresponding to the constraint (2) on the financial flows within the organization is to assume that the sales department guarantees at the beginning of the period a minimal return from the sales to the manufacturing department. This certain financial transfer, the value of which is decided upon before any realization of the prices, allows us to define a rational local decision rule for each player that is compatible with their common objective.

Denote by θ the monetary transfer the seller is committed to. One can then define a non-bankruptcy constraint by $\theta \le p_s^l s$: In that case, provided the behaviors of the players are consistent with the ex ante transfer θ, the risk of bankruptcy in the decentralized organization is zero. The corresponding strategies are then admissible and the values of θ which verify the non-bankruptcy constraint are also called *admissible*.

Solving the decentralized game is in two steps. First, for any admissible value of the transfer, we compute the locally optimal decision rule for each department, given that they are committed to respect the agreed upon transfer. Then, taking into account the probability laws of the prices, the local optimal rules can be evaluated in a central ex ante basis in order to determine the optimal value of the transfer.

Locally optimal decision rules and Nash equilibria

We define a locally optimal decision rule for the seller to be a quantity $Q_s(p_s, \theta)$ which maximizes the revenues from the sales at the price p_s under the constraint that the transfer θ is binding. This constraint imposes a minimal value to the quantity sold, Q^{min}_s, such that $p_s Q^{min}_s = \theta$.

Similarly we define a locally optimal decision for the manufacturer to be a quantity $Q_m(p_m, \theta)$ which maximizes the revenues of the manufacturing department when the price of the input is p_m and when the value of the anticipated transfer is θ. The financial constraint imposes then an upper bound on the quantity that is manufactured within the firm, $Q^{max}_m(p_m, \theta)$, defined by $p_m Q^{max}_m = t + \theta - F$.

The decentralized decisions $Q_m(p_m, \theta)$ and $Q_s(p_s, \theta)$ must thus respectively verify:

$$0 \leq Q_m \leq Q^{max}_m$$

and

$$Q^{min}_s \leq Q_s \leq s.$$

In order to study a real problem of production, we make the following assumption. When it does not hold, policies will merely reflect a trade-off beween financial costs and production costs.

Assumption 1 (*strongly increasing returns to scale*) $F > (1 - \delta)t$.

Under assumption 1, the following proposition gives the optimal decision rules $Q^*_m(p_m, \theta)$ and $Q^*_s(p_s, \theta)$. They are defined using threshold values of the prices p^0_m and p^0_s, respectively equal to $\bar{p}_s(t + \theta - F)/(\delta t + \theta)$ and $\delta \bar{p}_s$.

Proposition 1 *Under assumption 1:*

$$\begin{cases} Q^*_m = Q^{max}_m \text{ for } p_m \leq p^0_m \\ Q^*_m = 0 \text{ for } p_m > p^0_m, \end{cases}$$

and

$$\begin{cases} Q_s^* = Q_s^{min} \text{ for } p_s < p_s^0 \\ Q_s^* = s \text{ for } p_s \geq p_s^0. \end{cases}$$

The bang-bang characteristic of the optimal policies is due to the increasing returns to scale assumption.

The coordination problem within the firm is maximum when the values of the threshold prices lies inside the domain of variation of the prices. This is what we assume from now on.

Assumption 2 *For every admissible value of the transfer* θ, the respective threshold prices lie in the interior of the supports of the probability laws $\mathcal{L}_m$ and $\mathcal{L}_s$.

Let $\theta^{\max}$ be the maximal transfer which verifies the non-bankruptcy constraint, i.e. such that $\theta^{\max} = p_s^l s$. It can be shown that, under assumption 1, p_m^0 is an increasing function of θ (see Appendix 9.1). Then assumption 2 is equivalent to the following inequalities:

$$p_s^l < \delta \bar{p}_s,$$

$$p_m^l < \bar{p}_s(t - F)/\delta t,$$

$$\bar{p}_s\,(t + \theta^{\max} - F)/(\delta t + \theta^{\max}) < p_m^H.$$

Note that under these assumptions the average selling price verifies $p_m^l < \bar{p}_s < p_m^H$.

By focusing on the transfer θ from the seller to the manufacturer, we have in fact constructed Nash equilibria of the incomplete information game: if the minimum transfer that the manufacturer anticipates is equal to the minimum transfer that the seller's policy provides, then the strategies $Q_m^*(p_m, \theta)$ and $Q_s^*(p_s, \theta)$, where θ is the common value of this minimum transfer, form a Nash equilibrium of the game. The following proposition shows that at any Nash equilibrium of the game the values of the transfers anticipated by each department are the same and so our construction gives in fact all the equilibria of the incomplete information game.

Proposition 2 *Under assumptions 1 and 2, the Nash equilibria of the incomplete information game are described by a single parameter, namely the necessarily common value of the transfer the manufacturer anticipates and the seller guarantees. For any admissible value of the transfer, the equilibrium strategies are given by proposition 1.*

Determining the best Nash equilibrium of the decentralized game comes down then to finding the optimal minimum transfer within the organization. This is what is done now.

The optimal minimum transfer

The optimal minimum transfer is the one that maximizes the expected result $\mathcal{E}(\Pi)$. The following proposition shows that the optimal transfer is either zero or binds the non-bankruptcy constraint. Of course, which of the two is the optimal one depends on the values of the parameters.

Proposition 3 *For any bounded and continuous distribution of the prices, the optimal transfer is unique. Depending on the values of the parameters of the model, the optimal policy is either* $\theta = 0$ *or* $\theta = \theta^{max}$.

Interpretation of the results

The solution of this model captures some interesting features about coordination procedures in functional organizations. Two policies are identified. Observe that they result from corner solutions of the optimization problem. This means that the solution will not change continuously with a change in the values of the parameters, emphasizing the difficulty of adjusting in practice.

This difficulty will be even higher if one realizes that the two policies are completely different. To see this suppose now that the firm is operating in a multiperiod environment. Then, the policy which relies on a maximal transfer builds on sales today to reduce current costs, so as to make higher margins tomorrow. This policy may be called a "growth" policy in the sense that it focuses on internal efficiency almost irrespectively of the underlying uncertainty in the environment. The second policy relies exactly on the reverse, it bets on favorable market conditions to buy inputs at low prices and then to sell output at high prices. It may be called a "speculation" policy.

Relative efficiency of the two policies depends on the evolution of market conditions and on the initial wealth of the firm. If the market conditions become tighter (through convergence of output and output prices) then the growth policy may lose relevance. On the other hand, with a favorable spread between output and input prices and a relatively small initial wealth, the growth policy can be extremely efficient.

Suppose that the values of the parameters change from period to period, then it may become difficult to reorganize the information system of a firm currently organized along, say, a growth policy. This may require major discussions about what are the trends in the market conditions. It will also clearly change the internal dependencies, and the relative power, between the divisions. With a growth policy, the sales division is partly constrained to the manufacturing division while with a speculation policy it is totally free to move in its own best interest.

To illustrate the empirical relevance of this discussion we report the outcomes of an experimental game built along the line of the previous model.

Discussion with respect to the outcomes of a related experimental game

The experimental game

The experimental game (Figure 9.1) makes the following further assumptions.

1 The game is played during 15 periods, from period 6 to period 20. The first 5 periods are pre-played, so that the players have at their disposal historical accounting data and can get familiar with the rules.
2 The different periods are interrelated: the final stocks in financial and physical assets of a period become the initial state of the next period.
3 The probability laws $\mathcal{L}_m$ and $\mathcal{L}_s$ are not given to the players. The players must use their data to infer subjective probabilities on the prices p_m and p_s. The prices in the game have been drawn in advance once for all (they are depicted in Figure 9.1). Note that the selling price (in dashed line) tends to decrease over time. It is important that the players do not know in advance the time dependent probability laws that have been used to determine the prices. It is a crucial part of the experiment to observe whether or not the players adjust their subjective assessments and why.
4 The objective of the players is to maximize the value of the firm in period 20. At each period the value of the firm corresponds to the sum of the stock of the physical asset evaluated on the basis of an average selling price (the arithmetic mean of the last five observed prices) and of the available cash. Bankruptcy at any period during the play leads to the immediate termination of the game.

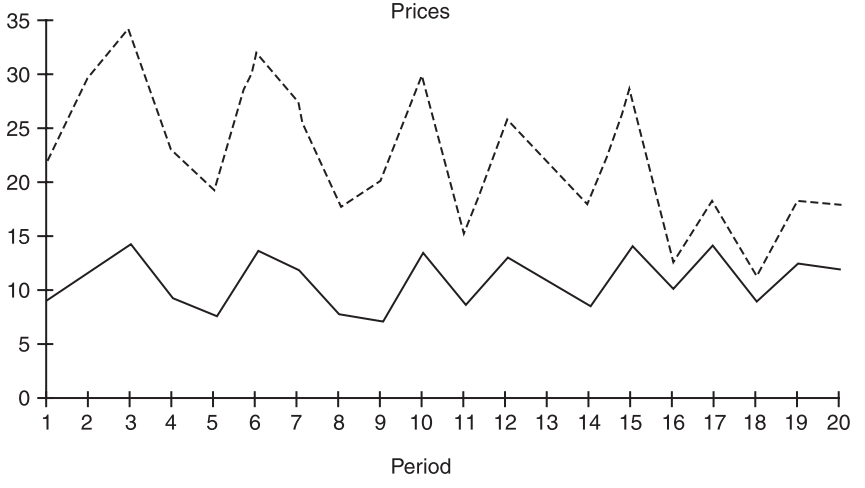

Figure 9.1 Prices in the experimental game

There are some differences between the model and the experimental framework, which play a minor role in our discussion (for instance the fixed cost is replaced by a production cost proportional to 120 times the square root of the quantity that is produced; depreciation of physical inventory occurs through a linear cost function). See Appendix 9.2 for the spread sheet which details the assumptions made for the experimental game.

This experimental game is played as follows. Prior to each period, there is a planning session in which discussion is completely open. Yet, the prices of the current period remain unknown. Then, the two players move apart and cannot communicate any longer. The different prices are given to the respective players who then decide on their move independently. As soon as these decisions are made the results of the period are computed and the next planning session can start.

Ordinarily, the participants consider that this framework is not too artificial and in fact is quite representative of how decisions are made in actual firms.

This experimental game has been played with managers as well as with graduate students from diverse origins such as maths, engineering, business, or economics.

Theoretical results and experimental outcomes

The two generic policies that appeared in the theoretical framework can be simulated in the experimental game.

The "growth" policy corresponds to the maximal transfer ($\theta = \theta^{\max}$):

1 sell the maximum available quantity (except in periods 11, 16 and 18 where one can consider that $p_s < p_s^0$),
2 buy on the basis of the sale of the whole stock, anticipating a lower selling price determined ex ante taking into account the previous realized prices.

The "speculation" policy corresponds to a zero transfer ($\theta = 0$):

1 when this policy is applied, the firm rapidly operates either only on cash or only on physical stocks,
2 at each period, only one player is active and makes his decision essentially on the basis of his observed price.

Figure 9.2 summarizes the evolution of the value of the firm for each policy from period 6 to 20. Observe that the values of the parameters of the game are such that the two policies lead to similar results in period 20.

Compared to these results, the majority of the observed outcomes when the game is played by real teams is extremely poor (Figure 9.2 gives the standard result of an experimental play of the game). Empirical scores of

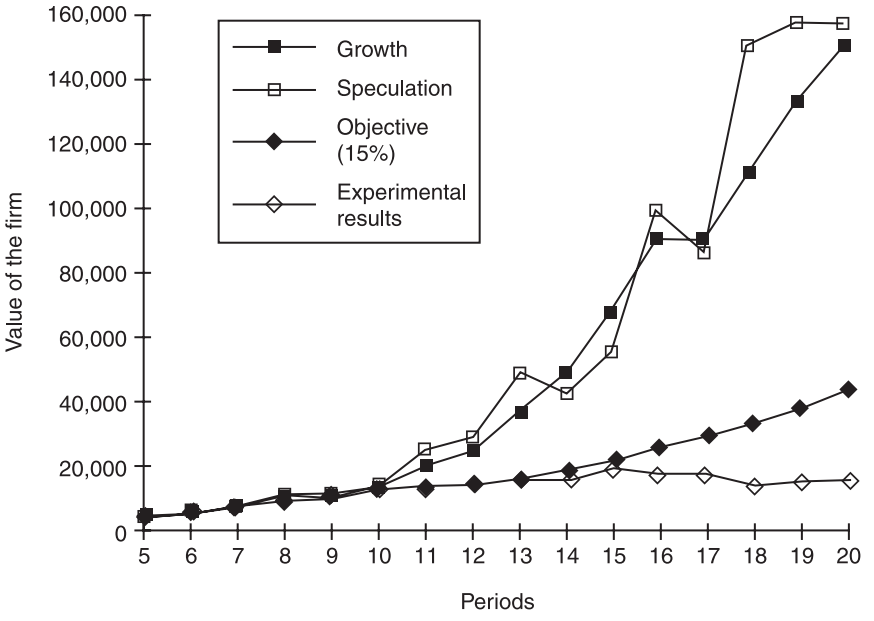

Figure 9.2 Numerical simulations and experimental outcomes

10,000 are a majority as compared with scores of 150,000 given by the generic policies.

A first lesson from the experiment is that identifying one of the two generic policies is already a major empirical problem. It is as if the players, forced to operate under important uncertainty, were completely losing their capacities for collective reasoning (recall that the players can explicitly discuss their strategies).

Still, a minority of teams do indeed identify one such policy, namely the growth policy. These teams do not go through a detailed analysis of the game like the one made above. In fact, such teams directly focus on the underlying line of reasoning which appears obvious to them just as a focal point: increase sales today to increase purchases to reduce cost to obtain higher margins tomorrow. Then they select a rule of thumb to determine what minimal transfer can be expected from the sales division. This routine is then applied repeatedly in spite of the fact that the effectiveness of such a policy declines over time with the corresponding decline in the average spread between prices. In fact, some teams may even go bankrupt because they remain too optimistic about the feasible transfer.

Thus the second lesson from this experiment appears to be that management tools have virtues and perverse effects as well. This is hardly a surprise but it is rarely taught in the same classes. In this sense this is an interesting contribution of this chapter. It emphazises the value of a cognitive map as well as the sociological and psychological difficulties to change it (Hall 1984).

While this game has been experimented several hundred times, it has never been the case that a team changes from a growth policy to a speculation policy toward the end of the play. It can be proved that such a time dependent strategy is the optimal one (see de Jaegere 1991) but this strategy seems completely unreachable.

What makes the difference between a team who identifies a generic policy and a team who does not has not been analyzed in detail in this chapter. This would certainly be a worthwile contribution to the notion of corporate culture as discussed in the introduction.

Notes

1 The reader is referred to de Groote (1994) for such a contribution.
2 So that the price p_m for inputs introduces also a variable cost proportional to the quantity of output produced.
3 X^+ is the positive part of real number X, i.e. X if $X > 0$ and 0 otherwise.
4 $\mathcal{L}(X)$ designates the expectation of the random variable X with respect to the law $\mathcal{L}$.

Appendix 9.1: Solving the incomplete information game

Proof of proposition 1

The seller

When θ is the financial transfer the seller guarantees to the manufacturer, the program of the seller is to maximize $\delta\bar{p}_s\ (s - Q_s) + p_s Q_s - \theta$ under the constraints $p_s Q_s \geq \theta$ and $Q_s \leq s$. Hence the optimal policy $Q_s^*\ (p_s; \theta)$;

$$\begin{cases} Q_s^* = Q_s^{min} \text{ if } p_s < p_s^0 = \delta\bar{p}_s \\ Q_s^* = s \text{ if } p_s \geq p_s^0, \end{cases}$$

with $p_s Q_s^{min} = \theta$.

The manufacturer

Due to the increasing returns to scale of the production process, the optimal policy for the manufacturer is either no production (A1), or production using only the financial stock t (A2), or production using the financial stock t and the transfer θ (A3). Given that the transfer θ, if it is not used, does not depreciate at the rate δ, the payoffs corresponding to these policies write as:

$$\begin{cases} \text{A1: } \delta t + \theta \\ \text{A2: } \bar{p}_s(t - F)/p_m + \theta \\ \text{A3: } \bar{p}_s(t + \theta - F)/p_m. \end{cases}$$

We now show that under the assumption $F > (1 - \delta)t$, policy A2 is always dominated.

Define two threshold prices p_m^0 and $\tilde{p}_m^0$ by:

$$\begin{cases} p_m^0 = \bar{p}_s(t + \theta - F)/(\delta t + \theta) \\ \text{and} \\ \tilde{p}_m^0 = \bar{p}_s(t - F)/(\delta t). \end{cases}$$

One has then:

$$\begin{cases} \text{A3 better than A1 for } p_m < p_m^0 \\ \text{A2 better than A1 for } p_m < \tilde{p}_m^0 \\ \text{A3 better than A2 for } p_m < \bar{p}_s. \end{cases}$$

Under the assumption $F > (1 - \delta t)$, one has $\tilde{p}_m^0 < \bar{p}_s$. This is sufficient to say that A2 is either dominated by A3 (for $p_m < \bar{p}_s$) or by A1 (for $p_m < \tilde{p}_m^0$). Hence:

$$\begin{cases} Q_m^* = Q_m^0 \text{ if } p_m \leq p_{\mathrm{m}}^0 \\ Q_m^* = 0 \text{ if } p_m > p_m^0, \end{cases}$$

QED.

Proof of proposition 2

Consider a Nash equilibrium $(Q_m(p_m);\ Q_s(p_s))$ of the team game with incomplete information. Let θ_s be the minimum transfer from the sales department, that is the lowest value of $p_s Q_s(p_s)$ over all the possible prices p_s. Let θ_m be the maximum transfer from the point of view of the manufacturing department, that is, the highest value of $p_m Q_m(p_m) + F - t$ over all the possible prices p_m. Notice first that to avoid bankruptcy, which is a necessary condition to get a Nash equilibrium in our game, the transfers must be admissible, that is, must take values between 0 and θ^{max}. We next show that at a Nash equilibrium, it must be true that $\theta_m = \theta_s$. First it is impossible that $\theta_m > \theta_s$: then there would exist prices configuration where constraint (2) would not hold, which would give an infinitely negative payoff. Second, it cannot be the case that $\theta_m < \theta_s$, since the manufacturer's best reply to a seller's strategy which guarantees a transfer θ_s, that is $Q_m^*(p_m;\ \theta_s)$ given by proposition 2, is under assumption 1 different from $Q_m(p_m)$. Thus one has $\theta_{\mathrm{m}} = \theta_{\mathrm{s}} = \theta$. Finally, proposition 2 ensures that the strategies $Q_m^*(p_m;\ \theta)$; $Q_s^*(p_s;\ \theta)$ are best replies to each other. QED.

Proof of proposition 3

When the probability density functions of the prices p_m and p_s are respectively ρ_{m} and ρ_{s} with respective supports $[p_m^l, p_m^H]$ and $[p_s^l, p_s^H]$, the ex ante profit function writes as:

$$\int_{p_m^l}^{p_m^0} \bar{p}_s(t+\theta-F)/p_m \cdot d\rho_{\mathrm{m}} + \int_{p_m^0}^{p_m^{\mathrm{H}}} (\delta \mathrm{t}+\theta)\cdot d\rho_{\mathrm{m}} + \int_{p_s^l}^{p_s^0} \delta\bar{p}_s\,(s-\theta/p_s)\cdot d\rho_{\mathrm{s}} +$$

$$\int_{p_s^0}^{p_s^{\mathrm{H}}} (p_s S-\theta)\cdot d\rho_{\mathrm{s}}.$$

Denote by $\Pi(\theta)$ this function. One can write:

$$\Pi(\theta) = \int_{p_m^l}^{p_m^0} \{(\bar{p}_s/p_m - 1)\theta + \bar{p}_s(t-F)/p_m\}\cdot d\rho_{\mathrm{m}} + \int_{p_m^0}^{p_m^{\mathrm{H}}} \delta \mathrm{t}\cdot d\rho_{\mathrm{m}}$$

$$+ \int_{p_s^l}^{p_s^0} \{\delta\bar{p}_s S + (1-\delta\bar{p}_s/p_s)\theta\}\cdot d\rho_{\mathrm{s}} + \int_{p_s^0}^{p_s^{\mathrm{H}}} p_s S\cdot d\rho_{\mathrm{s}}.$$

By differentiating one gets (taking into account the definition of p_m^0):

$$\frac{d}{d\theta}(\Pi(\theta)) = \int_{p_m^l}^{p_m^0} (\bar{p}_s/p_m - 1)\cdot d\rho_{\mathrm{m}}$$

$$+ \int_{p_s^l}^{p_s^0} (1-\delta\bar{p}_s/p_s)\cdot d\rho_{\mathrm{s}}.$$

As intuition suggests, the first term is positive and the second one is negative. However:

$$\Pi''(\theta) = (\bar{p}_s/p_m^0 - 1).\frac{dp_m^0}{d\theta},$$

with

$$\frac{dp_m^0}{d\theta} = \bar{p}_s\,(F-(1-\delta)t)/(\delta \mathrm{t}+\theta)^2.$$

Under the strongly increasing returns to scale assumption ($F > (1-\delta)t$), this last expression is positive. Thus Π, which is a convex function, can only be maximized when $\theta = 0$ or θ^{max}. QED.

Appendix 9.2: Initial spreadsheet of the experimental game

Period	1	2	3	4	5	6	7	8
Physical flows								
Quantity bought	100	124	180	140	140			
Quantity sold	100	95	129	125	145			
In process	100	124	180	140	140			
Finished goods	0	5	0	55	50			
Prices								
Input price	8,9	11,8	13,9	8,7	7,3			
Output price	21,8	29,5	34	22,3	18,9			
Average input price*					10,1			
Average output price*					25,3			
Financial flows								
Purchases	890	1463	2502	1218	1022			
Sales	2180	2803	4386	2788	2741			
Opening inventory	890	890	1522	2502	1697			
Closing inventory	890	1522	2502	1697	1387			
Added value	1290	1972	2864	764	1409			
Production costs	1200	1336	1610	1420	1420			
Inventory costs	0	10	0	110	100			
Management costs	30	30	30	30	30	30	30	30
Total costs	1230	1376	1640	1560	1550			
Balance sheet								
Value of inventory	890	1522	2502	1697	1387			
Net income	60	595	1224	–796	–141			
Capital	950	1545	2769	1973	1832			
Cash	60	23	267	227	445			
Indicators								
Value of the firm					5252			
Objective (15%)					5252	6040	6946	7988

9	10	11	12	13	14	15	16	17	18	19	20

30	30	30	30	30	30	30	30	30	30	30	30

9186	10564	12149	13971	16067	18477	21249	24436	28101	32317	37164	42739

References

Aoki, M. (1986) "Horizontal versus vertical information structure of the firm," *American Economic Review*, 5, 971–83.

Broseta, R. (1993) "Strategic uncertainty and learning in coordination games," University of California, San Diego, mimeo, 93–34.

Chandler, A. D. Jr (1962) *Strategy and Structure*, Cambridge, MA: MIT Press.

Crawford, V. P. and Haller, H. (1990) "Learning how to cooperate: optimal play in repeated coordination games," *Econometrica*, 58, 3, 571–96.

Crémer, J. (1980) "A partial theory of the optimal organization of a bureaucracy," *The Bell Journal of Economics*, 11, 2.

de Jaegere, A. and Ponssard, J. P. (1990) "La comptabilité: genèse de la modélisation en économie d'entreprise," *Annales des Mines, Gérer et Comprendre*, March, 90–8.

de Jaegere, A. (1991) "Vers un contrôle de gestion stratégique: un exemple de mise en oeuvre," *Economie Rurale*, 206, 97–104.

Hall, R. I. (1984) "The natural logic of management policy making: its implications for the survival of an organization," *Management Science*, 30, 8, 30–8.

Hart, O. and Holmstrom, B. (1987) "The theory of contracts," in *Advances in Economic Theory: Fifth World Congress*, T. Bewley (ed.), Cambridge: Cambridge University Press.

Johnson, H. T. and Kaplan, R. S. (1987) *Relevance Lost: The Rise and Fall of Management Accounting*, Cambridge, MA: Harvard University Press.

Kramarz, F. (1994) "Dynamic focal points in N person coordination games," *Theory and Decision*, 40: 277–313.

Kreps, D. M. (1984) "Corporate culture and economic theory," in *Perspectives on Positive Political Economy*, J. Alt and K. Shepsle (eds), New York: Cambridge University Press, 90–143.

Marschak, J. and Radner, R. (1972) *Economic Theory of Teams*, New Haven, CT: Yale University Press.

Mayer, D. J. Van Huyck, J. B, Battalio, R., and Saving, T. R. (1992) "History's role in coordinating decentralized allocation decisions," *Journal of Political Economy*, 100, 292–316.

Milgrom, P. and Roberts, J. (1992) *Economics, Organization and Management*, Englewood Cliffs, NJ: Prentice-Hall.

Ponssard, J. P. and Tanguy, H. (1993) "Planning in firms as an interactive process," *Theory and Decision*, 34, 139–59.

Schelling, T. C. (1960) *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.