Diplomacy Games

Formal Models and International Negotiations

Springer

Diplomacy Games

Rudolf Avenhaus · I. William Zartman
(Editors)

Diplomacy Games

Formal Models and International Negotiations

With 37 Figures and 42 Tables

Springer

Professor Dr. Rudolf Avenhaus
University of the Federal Armed Forces Munich
Werner-Heisenberg-Weg 39
85577 Neubiberg
Germany
rudolf.avenhaus@unibw.de

Professor I. William Zartman
School of Advanced International Studies
The John Hopkins University
1740 Massachusetts Ave.,
Washington, D.C. 20036
USA
zartman@jhu.edu

Library of Congress Control Number: 2007925685

ISBN 978-3-540-68303-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Published by Springer-Verlag Berlin Heidelberg 2007

Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover-design: WMX Design GmbH, Heidelberg

SPIN 11948902 42/3180YL - 5 4 3 2 1 0 Printed on acid-free paper

To Howard Raiffa
The formal modeler of PIN

Preface

Why does all the research on international negotiations over the past 20 years make very little use of formal theory, despite statements like that of Peyton Young that the principal tool for analyzing negotiations is the theory of games? Formal analysis of negotiations centers on paradigmatic models like Nash's bargaining scheme or Stahl and Rubinstein's sequential models, yet case studies of important international negotiations use only the most elementary concepts. The state of the art was characterized some years ago by a controversy surrounding the role of formal models—notably rational choice models—in political science. Some political scientists claimed that formal models have contributed little to the progress of political theory as a whole. Others vigorously defended formal theory. Here we deal only with the international negotiations part of the dispute.

From the very beginning and until today formal models were part of the research of the Processes of International Negotiations (PIN) Program of the International Institute for Applied Systems Analysis (IIASA) in Laxenburg near Vienna, Austria. In fact, there was a continuing emphasis on the importance of research on formal models for international negotiations by Howard Raiffa, IIASA's first director and one of the founders of the PIN program, to whom this book is dedicated. But it was not until 2002 that the idea was put into effect by initiating a PIN project devoted exclusively to the purpose of analyzing the role of formal models of, for, and in international negotiations and answering such questions as: How can rational choice theorists be induced to make greater efforts to render their abstract concepts and results more understandable to political and social scientists? What can be done to encourage social scientists to use formal approaches in their analysis of real-world problems? How can practitioners of international negotiations be induced to develop interest in and be taught to apply formal models to their more important problems?

This book contains contributions from leading experts in the field of formal modeling and international negotiations. These contributions were first presented at two workshops at IIASA in June 2003 and June 2004, where they were intensely discussed and revised by workshop participants. Their classification into formal models of, for, and in international negotiations grew out of these discussions (the editors were reluctant to propose their general characterization as "formal models from A to

Z"). Even though not all approaches and tools could be taken into account in a limited number of pages, we feel that this book represents the state of the art of formal models in international negotiations and gives some advice in the form of lessons for theory and practice.

The book is part of a series being produced by the PIN Program of IIASA, whose publications are listed in the front of this volume. We wish to express our gratitude for the collective participation of the steering committee members, all of whom provided a constructive atmosphere and helpful comments at many stages of the project. We also appreciate the flexible and responsive participation of the authors of the individual chapters who have joined the project from a number of countries. Special thanks are owed to Dan Druckman who encouraged us to initiate this project. We are also grateful for the ongoing support for the project by IIASA's director, Leen Hordijk, who made PIN's productivity possible, and also to IIASA's publications team. Basic support for all our activities and careful and pleasant attention to all the details of our work have been provided by Tanja Huber, our project administrator. Combining the systems analysis of IIASA with the diplomatic atmosphere of Empress Maria Theresa's palatial hunting lodge in Laxenburg, where IIASA carries out its state-of-the-art work within baroque walls, has been an inspiration for our efforts.

Laxenburg,
February 2007 *The editors*

Foreword

Negotiation is the central activity of diplomats and foreign policy leaders. Conflicts are avoided and ended through the art and science of negotiation. Every day, issues in foreign policy are resolved through negotiation. Water rights, environmental concerns, trade agreements, the birth of new international organizations, efforts at peacekeeping, and indeed every aspect of foreign interaction involves negotiation. That the subject and practices of international bargaining have, until relatively recently, been mostly studied within the limits of descriptive, after-the-fact assessments of specific negotiations is disappointing. Such studies have been useful, of course, but our negotiation knowledge and skills are likely to be further enhanced by complementing such research with investigations into the fundamental principles, processes, and practices of negotiation. As this volume makes abundantly clear, much about how to improve negotiations can be learned from and through the judicious use of formal models.

In the natural course of any negotiation, parties with differing interests or different information about how best to advance common interests must find the means to persuade their counterparts of a way forward. Such persuasion often involves a measure of the artful use of rhetoric, but in the end every negotiation is about strategic interaction [3]. Participants select courses of action with an eye on what it is that their counterparts say they need and what are believed to be their actual needs, as well as being equally focused on what the participants themselves need or the principals on whose behalf they are acting need. Thus, understanding and shaping negotiation outcomes depends on strategic calculations of exactly the sort captured by game-theoretic models of decision making under uncertainty. In this important publication, the editors, Avenhaus and Zartman, have assembled an exceptional cast of contributors and have significantly advanced the understanding and prospective use of formal models—including, but not limited to, game theory models—as tools used in the development, understanding, and conclusion of negotiations.

This volume usefully distinguishes between models *of* negotiations, *for* negotiations, and *in* negotiations. The distinction among these three types of analytic/modeling exercises neatly captures the idea of examining a negotiation to draw out principles of understanding, using these principles to evaluate alternative courses of action, and the application of these principles to facilitate the resolution of specific negotiations. As someone who has been engaged in these three activities from both an academic and commercial formal modeling perspective for 25 years, I am particularly excited by the potential this volume has for unleashing greater entrepreneurship, especially among political scientists, economists, and psychologists engaged in the modeling of complex bargaining problems. That there is great demand for such services from government agencies and the private sector there can be no doubt. That the knowledge is there to improve upon the performance of country or problem specialists alone there is also no doubt [1], [2]. While economists have been quite active and successful in designing and advising on game-theoretic actions for practical use, other social scientists have tended to lag behind. I know of only a handful of companies, for instance, formed by or operated by political scientists who are using formal models to advise government and private sector clients on how to improve negotiated outcomes. I certainly hope that this volume will stimulate our colleagues and students to turn their skills to the translation of insights from formal models into practical tools for policymakers and corporate decision makers. This, I believe, is the most promising path to improving decision making, tying it closely to logic and evidence rather than continuing reliance only on opinion, wisdom, and insider information. The latter are unquestionably useful, but they are fundamentally non-reproducible and inadequate as one shifts from one negotiation setting to another. Wisdom cannot be taught, but the rigorous interpretation of the logical and empirical implications of formal models can be. Indeed, many of the chapters in this volume point the way to how to do so.

The spread of formal model insights into the real-world, real-time negotiation process requires, as this volume highlights, a deep appreciation of both the strengths of formal approaches and their limitations. Extensive form games, for instance, are a valuable and often intuitive way to describe the—so to speak—essence of decision making. They force the analyst and the negotiator to identify exactly what critical decisions must be made by all sides in a negotiation. Game trees compel the analyst and the user to think hard about what is endogenous and what is exogenous to the decision process and, therefore, what can be influenced or reshaped and what cannot. That truly is the essence of resolving disputes. At the same time, the analyst must not overstate the comprehensiveness of the problem encapsulated in a game tree, and the decision maker must not allow the analysis of the game to be taken on blind faith—really there is little danger of this—but neither should he or she allow personal opinion or personal beliefs to remain comfortably unchallenged when the model's implications differ with intuition.

The strength of formal models is in their ability to lay out the logic of a situation and to make transparent what is assumed or taken for granted, what expectations look like, what information is critical for working through a problem, and at what junctures different choices lead irretrievably down better or worse paths. But every formal model—like every set of general principles, however derived—is necessarily a simplified, skeletal representation of the true complexity of any situation. Those who negotiate without the aid of models often risk blinding themselves to the big picture as they focus on intricate—and often inessential—details. But likewise those who rely mechanically on a formal model blind themselves to information not captured adequately by the model—such as differences in the skills brought to the negotiation by different participants—and so risk throwing away useful information that should be utilized to enrich their interpretation of analytic implications. The present volume does a masterful job of reminding us of the trade-off between the big picture and the subtle details.

Young [4] produced an earlier volume on a closely related theme about 15 years ago, and we can take heart and encouragement from the realization that much progress has been made and continues to be made. Increasingly, governments turn to formal, analytic tools combined with quantitative analysis to provide guidance in making critical decisions, testing in advance alternative approaches to central foreign policy matters and using models as relatively objective, detached tools by which analysts and policymakers can test the veracity or efficacy of their own judgments. I say this as one who knows at first hand how the United States government has used just such tools to help inform aspects of its policies toward terrorism in general and al Qaeda in particular, as well as in its approach to negotiations with North Korea and numerous other societies and troubling situations around the world. Those who take seriously the core messages of this volume will stand at the forefront of further advancing the goal of improving foreign policy decision making and the science of government negotiations in the years to come.

February 2007 *Bruce Bueno de Mesquita*

References

1. Feder S (1995) Factions and policon: New ways to analyze politics. In: Westerfield HB (ed) Inside the CIA's private world: Declassified articles from the Agency's Internal Journal, 1955–1992. Yale University Press, New Haven, CT, USA
2. Feder S (2002) Forecasting for policymaking in the post-cold war period. Annual Review of Political Science 5:111–25
3. Sartori A (2005) Deterrence by diplomacy. Princeton University Press, Princeton, NJ, USA
4. Young HP (1991) Negotiation analysis. The University of Michigan Press, Ann Arbor, Michigan, USA

Contents

List of Contributors

Markus AMANN (amann@iiasa.ac.at) is currently head of IIASA's Atmospheric Pollution and Economic Development program. Under his leadership the program developed the RAINS integrated assessment model and applied it to the international negotiations on the reduction of air pollution under the Convention on Long-range Transboundary Air Pollution and the European Union. Dr. Amann graduated from the Technical University in Vienna, Austria and received his PhD in Economics from the University of Karlsruhe, Germany.

Caitlyn ANTRIM (caitlyn@alum.mit.edu) is Director of the Center for Leadership in Global Diplomacy. She served as a delegate to the United Nations Conference on the Law of the Sea, for which she developed a model of the technology and economics of deep seabed mining that was used to analyze alternative formulas for agreement. As a member of the secretariat of the Intergovernmental Negotiating Committee on Drought and Desertification she established the Computer Information Center for delegates to the negotiations. She has also directed the American Academy of Diplomacy's Project on Multilateral Negotiation and managed the United States contribution to IIASA's Processes of International Negotiation Project. Her research focus is on the processes of multilateral diplomacy, particularly the roles of substantive, procedural, and inspirational leaders.

Rudolf AVENHAUS (rudolf.avenhaus@unibw.de) is Professor of Statistics and Operations Research at the University of the Federal Armed Forces Munich, Germany. Prior to his academic appointment in 1980, he was a research assistant in the Mathematical and Physical Institutes at the Universities of Karlsruhe and Geneva; a research scholar at the Nuclear Research Center, Karlsruhe; and a lecturer at the University of Mannheim. From 1973 to 1975, and again in 1980, he worked at IIASA. Professor Avenhaus has written numerous scientific publications, as well as *Material Accountability* (1977), *Safeguards Systems Analysis* (1987), and *Compliance Quantified* (together with M. Canty, 1996). In 1989 and 1990 he was chairman of his faculty; in 1993 and 1994, vice president; and in 1994, acting president of his university.

Hélène BESTOUGEFF (hbest@magic.fr) is a computer scientist and a professor at the CNRS-Université, Paris, France. Her research fields are databases and information systems, computation linguistic, multiagent systems, and game theory. She has been associated with research at Stanford University, IBM, and Xerox Park. She has published a number of scientific papers and written three books.

Steven J. BRAMS (steven.brams@nyu.edu) is Professor of Politics at New York University and has been an author or contributor to 15 books and about 250 articles. He has applied game theory and social choice to voting and elections, bargaining and fairness, international relations, and the Bible and theology. Brams is a former president of the Peace Science Society (1990–1991) and the Public Choice Society (2004–2006). He is a Fellow of the American Association for the Advancement of Science and has been a Guggenheim Fellow (1986–1987) and a Visiting Scholar at the Russell Sage Foundation (1998–1999).

Bruce BUENO DE MESQUITA (bbd2@nyu.edu) is Silver Professor of Politics at New York University, a Senior Fellow at Stanford's Hoover Institution, and a partner in Mesquita & Roundell, a consultancy with offices at Rockefeller Center, New York, that uses game-theoretic models to advise government and private sector clients on complex, multiactor negotiations.

David CARMENT (david_carment@carleton.ca) is Professor of International Affairs at the Norman Paterson School of International Affairs, Carleton University, Ottawa, and Fellow of the Canadian Defence and Foreign Affairs Institute (CDFAI). He is also the principal investigator for the Country Indicators for Foreign Policy project. He has served as Director of the Center for Security and Defence Studies at Carleton and has won several awards in the field. His most recent work focuses on developing failed state risk assessment and early warning methodologies and in evaluating models of third-party intervention. Recent publications focusing on these topics include, *Peacekeeping Intelligence, Conflict Prevention: From Rhetoric to Reality, Using Force to Prevent Ethnic Violence: An Evaluation of Theory and Evidence and Conflict Prevention: Path to Peace or Grand Illusion?*

Daniel DRUCKMAN (dandruckman@yahoo.com) is a professor in the Department of Public and International Affairs at George Mason University and in the School of Political Science and International Studies at Queensland University, Brisbane, Australia. He has also held the Vernon M. and Minnie I. Lynch chair in Conflict Resolution at George Mason. He has published widely on many topics (including modeling) and is the recipient of the 2003 Lifetime Achievement Award from the International Association for Conflict Management.

Michael GABBAY (mgabbay@islinc.com) is the Lead Scientist in the Network Dynamics Group of Information Systems Laboratories, Inc., in Seattle, Washington. His research focuses on the development of mathematical models and computational simulations of social and political systems, including group decision making and the dynamics of social networks. He received a PhD in physics from the University of Chicago with a specialization in nonlinear dynamics. He has also conducted research in the areas of nonequilibrium pattern formation, coupled oscillator dynamics, and data analysis algorithms.

Serdar Ş. GÜNER (sguner@bilkent.edu.tr) obtained his PhD degree from the Graduate Institute of International Studies, the University of Geneva. After post-doctoral research at the Department of Government, the University of Texas at Austin, he joined the Department of International Relations, Bilkent University. His research interests lie in game-theoretic modeling in international relations, evolutionary processes, and international relations theory. Besides numerous book chapters, some of his work is published in journals such as *Theory and Decision, Conflict Management and Peace Science, Security Dialogue*, and *Studies in Conflict and Terrorism*. He currently works on evolutionary game models of international systems.

D. Marc KILGOUR (mkilgour@wlu.ca) is Professor of Mathematics at Wilfrid Laurier University, Waterloo, Ontario, Canada, Associate Director of the Laurier Center for Military Strategic and Disarmament Studies, and Adjunct Professor of Systems Design Engineering at the University of Waterloo. His research interests lie in the analysis of interacting decisions, which he sees as the intersection of mathematics, engineering, and social science. He has applied game theory and related methodologies to problems in international security and arms control, environmental management, negotiation and arbitration, voting, fair division, and coalition formation, and he has pioneered the development of computer systems for decision support in strategic conflict. His four books and 150 journal articles cross many academic disciplines.

Thomas KRIEGER (thomas.krieger@unibw.de) is a Research Assistant at the Institut für Technik Intelligenter Systeme (Institute for Technology of Intelligent Systems) located at the University of the Federal Armed Forces Munich, Germany. He received his diploma in Mathematics in 2000 from the Technical University Dresden, Germany and his doctorate in Mathematics in 2003 from the University of the Federal Armed Forces Munich. His research interests include modeling and simulation as well as game theory.

Akira OKADA (aokada@econ.hit-u.ac.jp) is a graduate of the School of Economics, Hitotsubashi University, Tokyo, Japan. He received his PhD in 1982 from Tokyo Institute of Technology, Japan and has taught game theory at the Tokyo Institute of Technology, Saitama University, and Kyoto University. Since April 2004 he has been affiliated with the Graduate School of Economics, Hitotsubashi University, as

Professor of Economics. His research field is game theory and its applications to economics. He now acts as editor of the *Japanese Economic Review*.

Barry O'NEILL (barry.oneill@yale.edu) is Professor of Political Science at the University of California, Los Angeles. His current research considers infrasocietal institutions that influence international affairs—for example, honor, prestige, apologies, and insults. A further research area studies formal approaches to bargaining and mediation, such as clarifying the concept of negotiating over rights and negotiating in good faith. He has taught at Northwestern University, York University and Yale University. His book *Honor, Symbols, and War* received the Woodrow Wilson Award of the American Political Science Association for the year 2000.

Matthias RAITH (Raith@ww.uni-magdeburg.de) trained as an economist at New York University and the University of Bielefeld, Germany, where he earned his PhD in 1992. Since 2000 he has held the Chair of Entrepreneurship at Otto-von-Guericke University, Magdeburg, Germany. In 2001–2002 he was organizer and speaker of an international research group on 'Procedural Approaches to Conflict Resolution' at the Center for Interdisciplinary Research (ZiF) in Bielefeld. His research is focused primarily on the development of analytical methods and tools for facilitating individual and group decision processes. His publications address strategic implications of policy design, models of multistage strategic bargaining, and cooperative procedures for fair division and negotiation.

Dane ROWLANDS (dane_rowlands@carleton.ca) received his PhD in economics from the University of Toronto. In addition to his primary research on development finance, he is involved in research and teaching on the economics of conflict and on conflict intervention and has contributed to many books on peacekeeping and development. He is co-editor of the 2005 and 2006 volumes of the *Canada Among Nations* series, and is currently Associate Professor and Associate Director at The Norman Paterson School of International Affairs, Carleton University, Ottawa, Canada.

Michel RUDNIANSKI (michel.rudnianski@wanadoo.fr), a game theorist, has worked for the French Ministry of Defense, Prime Minister Office, which led him to develop the Theory of Games of Deterrence, which he has published in a number of scientific papers. He is presently the Academic Headmaster at the International Institute of Management, a department of the Conservatoire National des Arts et Métiers, one of the oldest French universities in Paris.

M. Remzi SANVER (sanver@bilgi.edu.tr; sanver@superonline.com) is Professor of Economics at Istanbul Bilgi University. His research is in Social Choice Theory with a particular focus on aggregation models and the manipulability/implementation of social choice rules. He is a member of the Society for Social Choice and Welfare, the Game Theory Society, the Society for Economic Design, the Foundation for Economic Design and the Center for Economic Design. In 2005 he received the

Middle East Technical University, Mustafa Parlar Foundation, Scientific Research Promotion Award.

Rudolf SCHÜSSLER (Rudolf.Schuessler@uni-bayreuth.de) is currently Professor of Philosophy at the University of Bayreuth, Germany. He studied philosophy, sociology and economics and took his doctoral degree in sociology and habilitation in philosophy. He is presently working on applied game and decision theory, the history of moral decision making in scholasticism, and several subjects of practical ethics. He is also developing a module on negotiations and conflict resolution for the Philosophy and Economics course at the University of Bayreuth.

Andrzej Piotr WIERZBICKI (A.Wierzbicki@itl.waw.pl) holds a PhD in nonlinear dynamics in control and has worked in a number of scientific research roles, teaching and lecturing for over 45 years. From 1996 to 2004, he was Director General of the National Institute of Telecommunications in Poland and from 2004 he has been a research professor at the Japan Advanced Institute of Science and Technology, Nomi, Ishikawa. Current interests include vector optimization, multiple criteria and game theory approaches, parallelization of optimization algorithms, diverse aspects of negotiation and decision support, diverse issues of information society and knowledge civilization, rational theory of intuition, theories of knowledge creation, and management.

I. William ZARTMAN (zartman@jhu.edu) is Jacob Blaustein Professor of International Organizations and Conflict Resolution at the Nitze School of Advanced International Studies of Johns Hopkins University. He is the author of *The Practical Negotiator*, *The 50% Solution*, and *Ripe for Resolution*; the editor of *The Negotiation Process* and *Positive Sum*, among other books; and the editor of *Preventive Negotiation*, the most recent book in the Processes of International Negotiation (PIN) series. He is organizer of the Washington Interest in Negotiations (WIN) Group and was distinguished fellow at the United States Institute of Peace.

Introduction: Formal Models of, in, and for International Negotiations

Rudolf Avenhaus[1] and I. William Zartman[2]

[1] University of the Federal Armed Forces Munich, Neubiberg, Germany
[2] Nitze School of Advanced International Studies, Johns Hopkins University, Washington DC, USA

Toward the end of the 1980s a major effort to promote the analysis of international negotiations was launched in the United States (USA) and subsequently at the International Institute for Applied Systems Analysis (IIASA), resulting, among other things, in the creation of the Processes of International Negotiations (PIN) Program at IIASA in Laxenburg, near Vienna. In the first publications and books of that time (e.g., *International Negotiation*, edited by Kremenyuk [20] and *Negotiation Analysis*, edited by Young [51]) the question of the relevance of formal models and quantitative approaches in regard to negotiation was carefully analyzed by several authors representing the different scientific disciplines contributing to the area.

Discussion on this topic has continued despite considerable progress in all directions. It gained new momentum through a controversy engaged between those who advocate the use of formal models in political science in general and others who question it (*International Security* 23(4) and 24(2) [spring and fall 1999]). The discussion, which centered around the three criteria for scientific work—logic and precision, originality and creativity, and empirical validity—as well as its length (more than 100 printed pages) indicate that there the problems are indeed wide open.

This book deals only with that section of political science defined by PIN's mission. It seeks insights into the whole problem area, but looks in particular for answers to three questions:

- How can the abstract concepts and results of rational choice theorists be made more understandable and plausible to political and social scientists not trained to work with formal models?
- What can be done to encourage practitioners' use of quantitative approaches, not only simple but also mathematically more advanced ones, in the analyses of real-world negotiations problems?
- How can practitioners (e.g., politicians and diplomats) become interested, take into account, and apply formal models of their more important problems?

Several elements in the debate of 10 years ago may later be used as tests to ascertain if the opinions expressed then still hold today or, hopefully, can at least be modified.

1 Views of the 1990s

Raiffa [34] writes explicitly in his introductory essay in PIN's first book, *International Negotiation*:

> The literature is replete with retrospective studies of past negotiations, and some of this literature attempts to use systematic analysis and modeling to rationalize and understand behavior and strategy. Analyses might examine in what ways real behavior is similar to the idealized behavior of rational actors in simplified, analogous games. Furthermore, game-theoretical reasoning might influence the pattern of empirical research; it might motivate the investigator to pursue lines of inquiry that would not be obvious without the insights from the abstract theory. . . .
>
> Analysis *of* negotiations is descriptive and may draw insights from normative (or abstract) studies. Analysis *for* negotiations has a different flavor. It is designed to help disputants and interveners (facilitators, mediators, or arbitrators) to do their job better in specific real-world negotiating problems. . . . But there is another aspect of analysis *for* negotiations that is different from background modeling of the environment of a dispute. How can analysis be used to help a disputant or intervener choose a course of action before, during, and after a negotiation? Here the paradigm is more decision analytical than game theoretical. For example: What are the alternatives? What are the uncertainties (about physical parameters as well as interests, values, and actions of other disputants)? What actions can be deferred until information is gathered along the way? What are the multiple conflicting objectives? The analyst might wish to model the actions to be taken by others in order to assess probability distributions over these action alternatives. In this regard, the sophisticated analyst should utilize the impressive body of growing literature on behavioral decision theory.

Young [51] in *Negotiation Analysis* expresses a somewhat different aim: to show how recent work in game theory, decision theory, and experimental psychology lends insight into the negotiation process.

> The principal theoretical tool for analyzing negotiations is the theory of game.. . . It applies to any situation in which the outcome of one person's actions or decisions depends, in a definite way, on the actions or decisions of others. In this sense, every negotiation is a game.

Bueno de Mesquita and Lalman [9] pose the question: Why model international relations?

> After all, no theory, regardless of its logical validity, elegance, or intellectual appeal, can substitute for a hard look at the facts. No abstraction is likely ever to be a practical guide to behavior until it has been confronted by reality. And we will most assuredly not allow ourselves to be swept away in theorizing without frequent recourse to the test of history.

> We model because we believe that how we look at the facts must be shaped by the logic of our generalizations. We are deeply committed to the notion that the evidence cannot be both the source of hypotheses and the means of their falsification or corroboration. By approaching our analytic task from a modeling perspective we improve the prospect that our propositions follow from a logical, deductive structure and that the empirical assessments are derived independently from the theorizing. But why look with such abstraction and in so arcane a manner as to rely on mathematical constructs in our quest for human understanding? If the deductions do not follow logically, then the formal structure will facilitate discovery of this condition.

A critical, yet constructive view is given by O'Neill [32] in his monumental review article on game-theoretical models of peace and war—around 700 references—in the *Handbook of Game Theory* (edited by Aumann and Hart [1]).

> No studies have yet compared game models of bargaining with real international negotiations, probably because the earlier mathematical bargaining theory focused on axioms for the outcomes and not dynamics.

He concludes, however:

> How should we judge Blackett's point [formulated in 1961], that if game theory were practical, practical people would be using it?... In fact game theory clarifies international problems exactly because they are more complicated. Unlike card games, the rules of interaction are uncertain, the aims of the actors are debatable and even basic terms of discourse are obscure. What does it mean to "show resolve"? What constitutes "escalation"? What assumptions imply that cooperation will emerge from international anarchy? The contribution of game models is to sort out concepts and figure out what the game might be.

Finally, Hopmann and Druckman's [14] essay on arms control and arms reduction, also in *International Negotiation*, expresses the views of political scientists who recommend the use of formal models only under well-defined circumstances:

> Beginning with the simplest assumptions we can construct a basic model of two rational, symmetrical, unitary individuals negotiating about a simple issue that can be treated on a single dimension. Such a model relies heavily on the game-theoretical traditions of bargaining ... for which simple assumptions are appropriate. In this model the negotiators are assumed to be utility maximizers, who strive to achieve agreements that provide them with the greatest possible benefits compared with the utilities associated with no agreement (or, equivalently, their best alternative to a negotiated agreement) [and] who have clearly defined utilities and complete information about one another's utilities. Given this information, the bargaining problem may then be "solved" analytically to arrive at either a unique outcome or a range of outcomes within which the parties "ought" to reach agreement. It thus seeks

> to identify the limits within which rational actors ought to reach agreement and how they ought to behave within those limits. . .
>
> Yet the conditions assumed by such formal models of bargaining seldom pertain in the real world of international negotiations. Often more than two parties are involved, and seldom are those parties internally unified. The utilities of the actors are often complex and changing, so that each party is uncertain about its own utilities, and the preferences of the other party are usually unknown as a result of deliberate concealment or even deceit. Therefore, very few international problems are "solvable" in this analytical sense.
>
> Formal bargaining models have certainly introduced some useful concepts for negotiation theory. . . Yet attempts to derive deductively valid "solutions" to the bargaining problem have largely failed to produce satisfying results when applied to international negotiations. The outcomes of negotiations on issues such as arms control are generally determined through the process itself, rather than being derived analytically from some clearly defined set of initial conditions.
>
> It is largely for this reason that formal models of bargaining have seldom been applied directly to the analysis of arms control negotiations, except as heuristic devices that may be useful in clarifying certain parameters and conditions for negotiation. The game-theoretical perspective lacks many of the features required for a valid theory of international negotiations, but it does offer some fundamental axioms on which most work is based. Insofar as these simple assumptions do suggest useful hypotheses, they may be preferred to more complex theories because of their parsimony.

These views, formulated as responses to Raiffa's [34] remarks 13 years ago, should be kept in mind when we are posing similar questions today.

But first, a word of warning. In the sense of an application of an abstract theory, one might see formal models and methods as tools in the search for truth, but one should be cautious not to interpret this as in the natural sciences. According to Mutoo ([26], p. 341), "it cannot be over-emphasized that this theory . . . does not purport to generate falsifiable predictions à la Popper—and indeed it cannot do so, since some of its key ingredients (such as the rationality of the players—which is implicitly embodied in the equilibrium concepts employed) are neither observable nor controllable." There is a wide discussion on the subject of rationality, sometimes leading to extreme positions: Is there any irrational behavior at all, or do we just not know the real preferences? If we accepted the latter position this would mean that we would have to adapt preferences to the outcomes which we want to explain; in other words, we would be in danger of becoming tautological.

We do not intend to enter into this discussion, but refer to the literature (see, e.g., [21]). However, there is the important area of experimental game theory or gaming that deals with these problems. It also serves, but not primarily, the purpose of either confirming the truth (in the sense of describing reality) of a model or else rejecting it. Its more important aim, nevertheless, is to discover the utilities, and therefore the motivation, of specific individuals or groups (see, e.g., [43]).

2 An Analytical Framework for Negotiation Theory

Any analysis of negotiations, bilateral or multilateral, needs some key concepts that help to structure the considerations. Therefore, in the following we start by describing the so-called analytical framework for negotiation theory as proposed for the first time by the PIN Steering Committee members and published in *International Negotiation* (Dupont and Faure [12]), and developed and used extensively in PIN's subsequent work. Thereafter, we will compare its elements to those of extensive form games and return to our basic question as to why, despite all their common features, the latter have in practice not yet been used in concrete applications.

As we define it, negotiation is purposeful communication consisting of *strategies* developed and implemented by two or more *actors* to pursue or defend their interests. The entire pattern of interaction constitutes a *process* played against a *structure* of background factors that change slowly over the long term and produce an *outcome* comprising the results attained in a negotiation. Let us consider these elements in some more detail.

Parties to international negotiations are actors—states, business firms, international organizations, and other institutions—who are drawn into the process because they are concerned in one way or another with the positive or negative values represented by the issues put on the agenda. Such values may be related to business opportunities, the mitigation of risks, the development of a regime for trade, or for the use and distribution of benefits. The values manifested by the negotiated issues draw particular sets of participants to each negotiation.

Decision dilemmas confronting negotiating parties may require specific strategies. Parties need to envisage the paths by which they calculate their ability to achieve their preferred outcomes and attain their goals, including an evaluation of how many of the preferred goals can fit into feasible outcomes. Strategies begin as unilateral plans for action but ultimately must take into account the goals and plans for action of the other side(s), as well as the contextual constraints and opportunities of the negotiating situation. Strategies can be thought out either as chains of interlocking decisions or as efforts to control the process of negotiation (or both).

All issues on the agenda of a negotiation are likely to proceed through a process involving a similar pattern characterized by the concept's prenegotiation, agenda setting, formulation, negotiation on details and, finally, agreement. In principle, all these stages are present to some degree in every negotiation. In some cases, prenegotiation consists only of an easily attained agreement on what issues will be negotiated, whereas in other cases, agenda setting and the clarification of issues will require a considerable amount of work at the table and therefore continue for a long time. In some complex multilateral negotiations, the initiation and decision to start a negotiation—prenegotiation—has been a complex and protracted process that has been going on for several years. In other cases, one or more of these process stages may hardly be noticeable. Sometimes, parties bring a formula for a bargaining approach from earlier negotiations, that is also acceptable to all parties in the upcoming talks. On other occasions the development of a formula may represent the most critical and difficult phase of a negotiation. However, in spite of all these possible variations across individual cases, the general process model should in principle be applicable to all negotiations.

Negotiations so characterized occur within a context that enables and constrains the search for an agreement. Such structures refer to internal components including the number of parties, the available and relevant means or power, and the relationships among ends. Elements such as power (a)symmetries, bargaining zones (or negotiating fronts or zones of possible agreement [ZOPAs]), and security points (or reservation prices, or threat points, or best/worst alternative to a negotiated agreement [BATNAs or WATNAs]) are significant in shaping the possibilities and paths of interaction (there is still no agreement among analysts on consensual naming of important concepts). External structures relating to precedents, power resources, institutions, and norms, among others, also have important effects on negotiations.

A negotiation typically strives to produce an outcome between the parties in the form of an international agreement. Its function is to specify the commitments the parties have accepted in the negotiation, if by outcome we mean the end results of such international talks. However, a detailed treaty is not necessarily a comprehensive representation of the outcome. On a closer look, outcome is a highly problematic concept that may be interpreted in different ways, depending on the theoretical outlook. For example, an analyst who wants to evaluate how important a concluded negotiation has been may argue that an outcome assessment should not only take stock of the text but also consider how parties implement and comply with the obligations that they have accepted in it.

Negotiations may generate other forms of informal results that are not specific in the treaty. Part of this informal outcome may be captured with the help of the concept of regime building ([19], [51], [48]). According to a standard view, regimes do not only include formal rules and procedures but also norms and consensual knowledge. Norms are not formalized as international law but give directions as to what is right and what is wrong. Knowledge clarifies the issues on the agenda, particularly by identifying a causal relationship between key variables.

If the negotiated commitments in the form of treaty rules are weak, the overall outcome of the negotiation may still be important because of the accepted norms and the shared knowledge that it has generated. This perspective on the outcome of negotiations is particularly significant when the issues on the table are both politically sensitive and technically complex. Negotiation on such issues has usually been recurrent, in the sense that state commitments have been built up stepwise over a longer period of time covering not only years but decades. Earlier stages of the cycle of negotiations produce end results conditioning regime-building accomplishments in ensuing phases of the talks. Regime rules may become gradually strengthened from one negotiation event to the next. However, in complex and politically sensitive areas, the gradual reinforcement of norms and consensual knowledge may be more important an outcome in a long-term perspective than the stepwise hardening of regime rules.

At the beginning, one remark about terminology. Whereas, in line with political science in general and PIN in particular, we always talk about negotiation, game theorists use the word bargaining. Perhaps this is because of Nash [28] who published the first important paper on cooperative bargaining using this wording. Thus, in talking about game-theoretic models, we will maintain this tradition.

3 Extensive Form Games

Models are logical dynamic simplifications of reality, designed to lay bare the underlying mechanism governing interactions among their components. They need to pass the test of logic in the relationships that they posit, and then, so armed, they sally forth to confront the test of reality, to see if the world they purport to describe is real. There are many types of models, depending on the terms of analysis that they embody; one prominent type involves game theory. According to the foreword of the *Handbook of Game Theory* [1], game theory studies the behavior of decision makers whose decisions affect each other. One may distinguish two approaches to game theory: the noncooperative and the cooperative. A game is cooperative if commitments—agreements, promises, threats—are fully binding and enforceable, otherwise it is noncooperative.

Even though at first sight cooperative games seem to be at least as important for the study of international negotiations as noncooperative games, and even though the former will play their role in subsequent chapters of this book, for reasons noted below we will consider for a while only noncooperative games. For the sake of completeness, let us also quote the late Nobel Laureate, Selten [42]:

> The traditional distinction between noncooperative and cooperative game theory has lost its original significance. Today noncooperative game theory is also applied to problems of cooperation. One may say that the explanation of cooperation requires the use of noncooperative theory. In cooperative game theory cooperation is assumed rather than explained.

Moreover, the present discussion will concentrate on noncooperative games in extensive form, as opposed to games in normal or strategic form. The latter, defined by the number of players, by the sets of strategies, and by the payoffs to the players that they obtain if they (all together) use a well-defined strategy combination, are deceptively simple. The concept of a strategy comprises many different aspects, for example, sequencing, information, chance, and others. These aspects, so important for negotiations, are much better expressed in games in extensive form and perfectly match the analytical framework presented before.

A noncooperative game in extensive form is a graphical representation of the possible moves of all players from the beginning of the game until its end. It has the form of a tree—growing from the top to the bottom—where a set of branches starting at some point indicate a player's alternatives at that point. A precise mathematical definition of extensive form games, following Myerson [27], is given in *Table 1*. The rooted tree describes the whole negotiation process. Players are actors under a different name. The information state in the game is at least part of the structure. The set of all possible moves of the players is the set of strategies, and the payoffs at the terminal node are quantitative measures of outcomes. Thus, one may ask again: Why does there seem to be a problem of applying game-theoretical formalism to real-world negotiations?

Table 1. Extensive form game, strategy and Nash equilibrium (after Myerson 1997)

An n-person noncooperative extensive form game is a rooted tree—usually growing from the top to the bottom—together with labels at every decision point or node and decision alternative or branch, defined as follows:

1. Each nonterminal node has a player label that is taken from the set (0,1, . . .,n). Nodes that are assigned a player label 0 are called chance nodes. The set (1,2, . . .,n) represents the set of players in the game, and for each individual player i in this set, the nodes with the player label i are decision nodes that are controlled by that player.
2. Every alternative at a chance node has a label that specifies its probability. At each chance node, these chance probabilities of the alternatives are nonnegative numbers that sum to one.
3. Every decision point or node that is controlled by a player has a second label that specifies the information state that the player would have if the path of the play reached this node. When the path of the play reaches a node controlled by a player, the player knows only the information state of the current node. Thus, two nodes that belong to the same player should have the same information state only if the player would be unable to distinguish between the situations represented by these nodes when either occurs in the play of the game.
4. Each alternative or branch at a node that is controlled by a player has an alternative or move label. Furthermore, for any two nodes x and y that have the same player label and the same information label, there must be one alternative or move at both nodes that has the same move label.
5. Each terminal or outcome node has a payoff label for each player, such that for each player i, there is a payoff u_i for player i, measured on some utility scale.

A strategy is a function that maps information states into moves. For each player i, let S_i denote the set of possible information states for i in the game. For each information state s in S_i let D_s denote the set of moves that would be available to player i when he moved at a node with information state s. Then the set of strategies for player i in the extensive form game is the combination of all successive sets of alternatives or moves D_s for s in S_s referred to as their Cartesian product.

A Nash equilibrium is a set of strategies with the property that unilateral deviation of a player from his equilibrium strategy does not improve his payoff.

A pure *strategy* of a player in an extensive form game is any rule for determining a move at every possible information state in the game; its precise mathematical definition is also given in *Table 1*. A mixed strategy is a probability distribution on pure strategies.

Having defined a strategy in an extensive form game, we can formulate a solution of such a game, following the path-breaking work of Nobel Laureate, Nash [29], originally defined in in a normal form game: A *Nash equilibrium* is a set of strategies with the property that any unilateral deviation of a player from his equilibrium strategy does not improve his payoff. There may be more than one such equilibrium, although some authors use the term “solution” only when the equilibrium is indeed unique.

4 Offer–Counteroffer Bargaining

As an illustration, a common negotiation process may be captured in an offer–counteroffer bargaining model ([49], [40]). Two players alternate offers sequentially. One side makes an offer and the other side can accept or reject. If the offer is rejected, the other side can then make a counteroffer of its own, and the first side accepts or rejects that counteroffer. If the counteroffer is rejected, the bargaining continues with another offer by the first side, and so on. Each offer and response represents one round of bargaining.

What hinders the two players from bargaining for ever? The answer is that time is money, or more generally, some valuable good; therefore, both players prefer to come to an agreement sooner rather than later. In Stahl and Rubinstein’s models the players discount the final bargaining for each additional round of bargaining using discount factors d_1 and d_2 between zero and one.

To illustrate this concept, let us recall one of the stories from Faure’s collection [13]. A foreign visitor to the pyramids very naively believes an Egyptian boy who invites him to climb on the back of his camel in order to be photographed without having to pay anything and, once up, is taken hostage; the boy will not lower his camel and let the man go before he has paid EGP 2.00. Obviously, the discount factors for the boy on the ground and for the visitor on the back of the camel in the Egyptian sun are very different. What should the poor man do?

The path to an answer passes through a demonstration of the Stahl–Rubinstein model developed by Morrow [25]. Two players have to share 100 utility units. The time line of a sequential bargaining game with three rounds of alternating offers goes as described in *Table 2* (see also *Figure 1*):

In a game with three rounds, the analysis, which is performed by reasoning backwards from the terminal or outcome nodes, shows that in equilibrium player I offers in the first round the amount $a_1 = 100(1 - d_2(1 - d_1))$, and player II accepts this offer. Thus, there is no real bargaining as we experience it in the real world. The second player’s payoff is the smaller, the larger d_2, and the smaller d_1. Impatience has its price. In the game with an infinite number of moves in equilibrium, player I

Table 2. Offer–counteroffer bargaining game (after Morrow 1994)

Two players have to share 100 utility units. The time line of a sequential bargaining game with three rounds of alternating offers goes as follows: Each round of bargaining consists of two moves, an offer and a response. For ease of exposition, each offer is the amount player I receives in that offer, player II gets the remainder. It is also assumed that time has a monotonal or constant cost. It is agreed that, if one player is indifferent between accepting an offer or rejecting it (i.e., gets the same payoff in both cases), he will accept the offer. The negotiation is limited to three rounds, expressed in the following six steps which are represented graphically in *Figure 1*.

1. Player I offers a_1 somewhere between 0 and 100 (that is, $0 < a_1 < 100$).
2. Player II accepts or rejects this offer. If he accepts, the game ends, player I receives a_1, and player II receives $100 - a_1$.
3. If player II rejects the first offer, he also offers a_2 to player I between 0 and 100 (that is, $0 < a_2 < 100$).
4. Player I accepts or rejects this offer. If he accepts, the game ends, player I receives player II's offer reduced in value with the passage of time (that is, $d_1 a_2$), and player II receives the rest (that is, $d_2(100 - a_2)$).
5. If player I rejects, he offers a_3, again with $0 < a_3 < 100$.
6. Player II accepts or rejects this offer. If he accepts the game ends, player I receives $d_1^2 a_3$, player II receives $d_2^2(100 - a_3)$. If player II rejects, both players get nothing.

As a result, in equilibrium player I offers in the first round some fixed amount and player II accepts this offer. One can extend this game to an infinite number of moves, and the result is principally the same: player II immediately accepts in equilibrium the initial offer made by player I. Thus, there is no real bargaining.

immediately offers the amount $100(1 - d_2)/(1 - d_1 d_2)$ and player II accepts, getting the rest. Of course, we have here the same dependence on the impatience of the players.

In order to illustrate these results for the case of infinitely many rounds, let us assume $d_1 = .99$ and $d_2 = .95$. Then player I offers to take the amount 84 himself, whereas for $d_1 = .95$ and $d_2 = .99$ he offers to take the amount 16.8.

Now, what does this model tell us about the man on the back of the camel in the hot Egyptian sun? In estimating one negotiation round to last 10 minutes, it is reasonable to consider the bargain to be limited in time: three hours make up for 18 negotiation rounds, a large number for the man in the sun.

Furthermore, let us reasonably assume that the man's time is much more expensive than that of the boy. This means that we have in any case $d_1 > d_2$ and therefore, the boy's share $(1 - d_2)/(1 - d_1 d_2)$ is larger than that of the man, $d_2(1 - d_1)/(1 - d_1 d_2)$, perhaps even much larger. Assuming the man has at least EGP 4.00 in his pocket, and the boy asks for EGP 2.00, then our analysis tells us that the man is well advised to pay the EGP 2.00 immediately, which gets him down off the camel's back with a minimum danger of sunstroke.

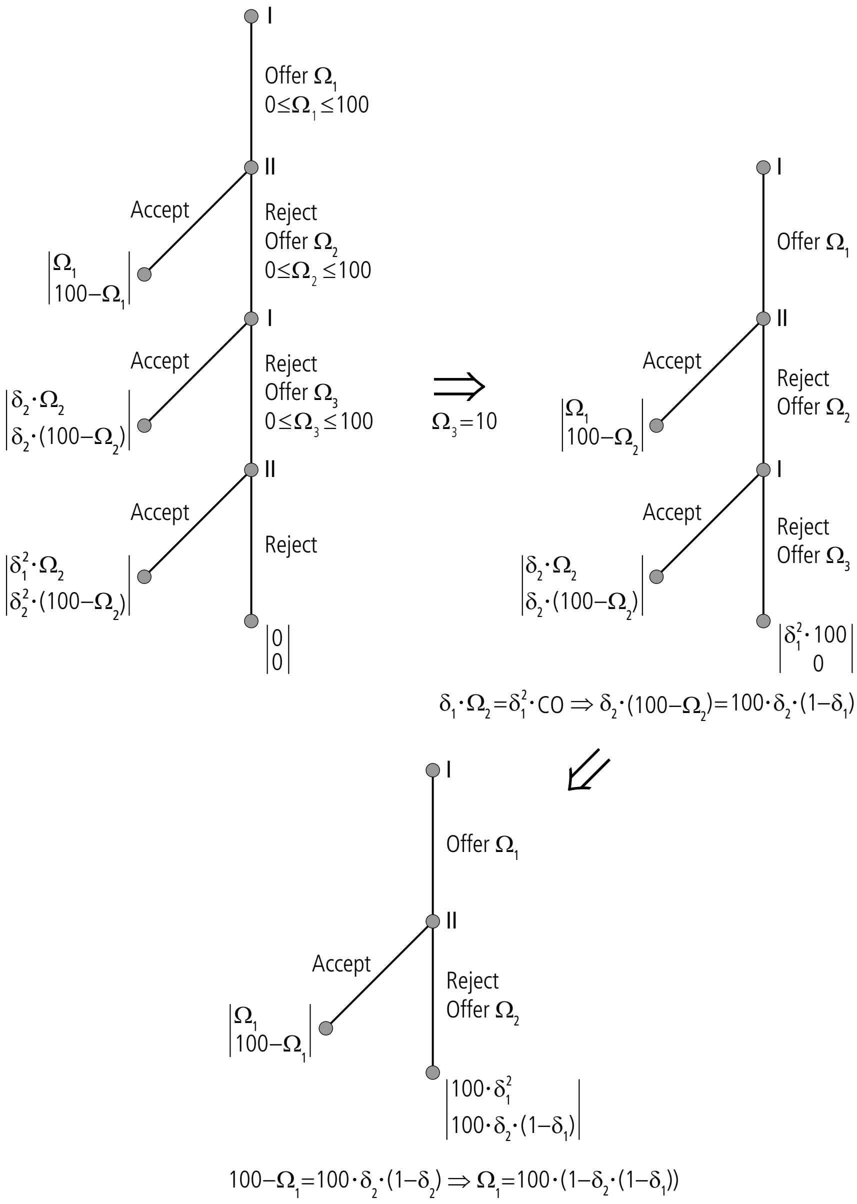

Fig. 1. Extensive form of the Stahl–Rubinstein bargaining model with three rounds and its solution by backward induction. The equilibrium is determined in such a way that the player who has to make a decision is indifferent

A more sophisticated model is Morrow's [23] treatment of crisis bargaining, where the reputation for resoluteness of different types of players is introduced. Whereas the bargaining of a defender with a challenger does not lead to war if the challenger knows of what type—resolute or not—the defender is, it may lead to war if the former has only imperfect information about the type of the latter. Other types of sophistication to capture key variables in reality will be discussed below.

5 Scope of the Book

Even if the initially quoted view of O'Neill [32] about the application of formal models in international negotiations over a decade ago may be too categoric today, the next part shows that the list of real applications in the sense of models for and in international negotiations is not too long. What, therefore, must and can be done in order to close the gap between formal modeling and general negotiation theory on one hand, and between the former art and practice on the other, as asked in the opening question?

For this purpose, three categories of formal models used in the area of international negotiation will be discussed. This classification emerges from the literature and will be sketched subsequently; it also arises from the discussions during and after the workshops that were part of the PIN project resulting in this book, and it follows and expands on Raiffa's categories.

Formal models *of* international negotiations are either abstract mathematical theories, like Nash's bargaining approach [28] or Stahl and Rubinstein's offer–counteroffer bargaining theory described above, which offer solutions to bargaining problems but do not really describe negotiation processes; rather, they give advice on how to agree immediately. They may also be descriptive models of negotiations of the past or the present, which may be simplified guides to a social scientist or a historian to understanding the behavior of statesmen and diplomats.

Formal models *in* international negotiations are heuristic and dynamic by their very nature. They describe international negotiations with the help of external data showing the implications of contextual or structural elements of use in the negotiations, but they do not model the actual negotiations themselves, as in the first category. They provide useful knowledge about possible mechanisms or situational consequences to inform negotiators about alternatives and outcomes and to spur creativity in finding solutions.

Formal models *for* international negotiations are used to combine the preferences of the parties into optimal outcomes; possible values and outcomes are prioritized and quantified and then analyzed—often with the use of computer programs—to find where joint best payoffs may lie. The method is termed negotiation analysis, based on the term decision analysis, but it actually provides a substitute for negotiation, bypassing the wasteful political process.

5.1 Previous work

Let us start our review of previous work with the first category of formal models defined above. We present formal models *of* negotiations of the Cuban Missile Crisis of 1962 as a first example. Brams [5] and Brams and Kilgour [7] have shown that this crisis may be characterized as a Chicken Dilemma Game (CDG). Like many applications of the first category, the crisis is described by a simple normal form game where—in this case—the two strategies of the United States are blockade of Cuba and air strike, and those of the Soviet Union (USSR) are missile withdrawal and missile maintenance, see *Figure 2*.

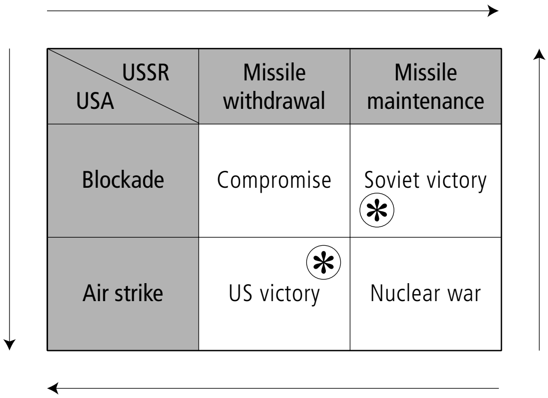

Fig. 2. Normal form of the Chicken-type Game between the United States and the Soviet Union describing the Cuban Missile Crisis in 1962. Arrows indicate preference directions; asterisks equilibria

The following ranking of the four outcomes for the United States (“>” means “better”): United States (US) victory > compromise and Soviet victory > nuclear war and for the Soviet Union: Soviet victory > compromise and US victory > nuclear war produce two pure strategy combinations (Blockade, Missile maintenance) and (Air strike, Missile withdrawal) as two pairs of equilibrium strategies, which means that a CDG model does indeed portray the situation.

In a CDG situation, the parties agree that deadlock is the worst outcome for both of them, but there are two dominant solutions. Each party’s efforts to move unilaterally to the best solution for itself take it to its own dominant solution or equilibrium strategy, in what is called the method of preference directions. The degree of “worstness” of the deadlock, which cannot be portrayed in a simple ordinal presentation such as that indicated here, determines whether the parties will move to the alternative joint outcome compromise involving blockade and withdrawal—as they did in this instance.

A second example of the first category of applications involves a different game configuration, illustrated by an arms control negotiation. Two states negotiate whether or not to continue with some kind of armament (see *Figure 3*).

I \ II	Stop	Continue
Stop	Status quo	II's military advantage
Continue	I's military advantage	* Armament of both

Fig. 3. Normal form of the Prisoner's Dilemma-type Game between states I and II describing their arms race. Arrows indicate preference directions; the asterisk indicates the only equilibrium

Both states have two strategies: to stop their armament or to continue. If both stop, then the status quo is maintained, if both arm, they have to carry the burden of increased armament costs without gaining a military advantage. If one state arms and the other does not, then the arming state has military advantage, which outweighs the costs. If both states rank the outcomes as follows: armament of both $>$ military disadvantage, and military advantage $>$ status quo, then each will seek armament (military advantage), and the method of preference directions shows that the resulting armament of both states represents a pair of equilibrium strategies, even though the maintenance of the status quo could make both of them better-off. The situation described in the model is called a Prisoner's Dilemma Game (PDG).

In these two examples the details of the negotiation process, in particular, their dynamics, are not modeled; the model portrays the situation, usually as a matrix or normal form, which can then be used to illustrate the choices (strategies) and consequences (outcomes) facing the negotiators (players). Let us repeat that such models of negotiation are useful in guiding the parties to think about the implications of their situation and how to get out of it or exploit it, but they do not portray movement (process) nor do they prescribe how to act outside the box. The application of game-theory models to portray negotiation situations is referred to as *proto-game theory*, using the concepts but no or only a few formal derivations [32].

A more elaborate model of an international negotiation is that of the Cuban Missile Crisis given by Wagner [50]. As a first step, Wagner specified the sequence of choices that were made, the alternatives among which each party had to choose at each opportunity, and the information each had, when choosing, about the choices made by the other. The crisis was begun by Khrushchev's decision to deploy missiles in Cuba. Wagner thus assumed that the game was initiated by a simple choice by Khrushchev between two alternatives, namely to place missiles in Cuba or not. Following Khrushchev's selection of the option of installing the missiles, it was Kennedy's turn to choose among a variety of distinguishable options. Wagner took three: Kennedy could accept Soviet missiles in Cuba, he could attempt to remove them by military force, or he could attempt to induce the Soviets to remove them. A graphical representation of this scheme is given in *Figure 4*.

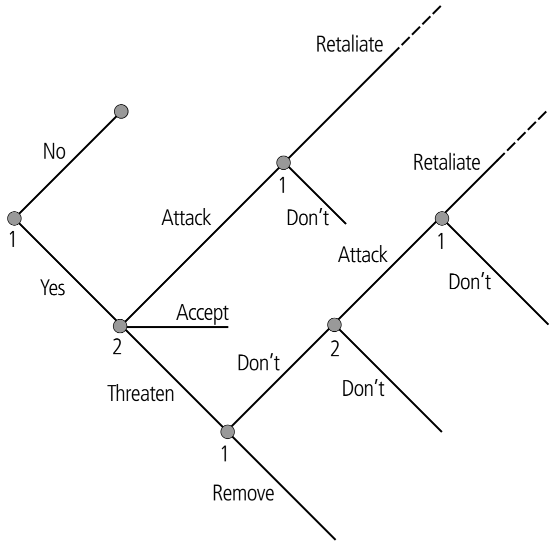

Fig. 4. Schematic representation of the extensive form game describing the Cuban Missile Crisis. Khrushchev is labeled player 1 and Kennedy player 2. Yes means placements of missiles in Cuba and No means no placement (after Wagner [1989])

The only way to induce Khrushchev to remove the missiles was to threaten to take some action that he would think was worse than removing them. Whatever action Kennedy might threaten to take, Khrushchev would have an opportunity to respond to Kennedy's response, and so forth. Thus, Wagner modeled this crisis as a game in extensive form. The game actually played ended with Khrushchev's decision to remove the missiles in response to Kennedy's threat. Since it seems extremely unlikely that Khrushchev preferred this outcome to the one that would have resulted from an initial decision not to place the missiles in Cuba, he cannot have anticipated this. Thus it seems likely that his expectations of Kennedy's behavior were altered by what happened during the crisis. One of the main problems in explaining Khrushchev's behavior is to understand how this could have occurred.

Instead of going through the long discussion to the payoffs to both players, the incorporation of incomplete information, the analysis of the threats, and of what both players learned in the course of the crisis about the other side's intentions, let Harrison Wagner himself tell us what was learned from the analysis:

> In order for bargainers to learn, bargaining must be costly since otherwise they may be suspected of bluffing. The main source of uncertainty in the bargaining that took place during the missile crisis was Kennedy's willingness to take military action if the missiles were not removed. Thus the risk of conflict entailed by the blockade seems most plausibly understood, not as something that the Americans used to put pressure on the Soviets so that they would back down, but as something that allowed the Soviets to put pressure on the Americans so that Kennedy's threat of subsequent military action would seem more credible. And since the Soviets had good reason to want to know how seriously to take Kennedy's threat, this was something they were not averse to doing. Seen in this light, the missile crisis was not a competition in risk taking, but a strange sort of cooperation in the exchange of information.

Models *for* negotiation represent the second category in our classification, showing the implications of contextual or structural elements of use in the negotiations, but not modeling the actual negotiations themselves, as the first category did. As noted, these models are used to combine the preferences of the parties into optimal outcomes; possible values and outcomes are prioritized and quantified and then analyzed—often with the use of computer programs—to find where joint best payoffs may lie. Their most prominent advocate is Raiffa ([33], [35]), who has used multi-attribute value analysis and decision analysis in his consultative work to help find solutions more efficiently.

A first and very important area of applications can be the international negotiations on climate change aiming at agreements on the global reduction of greenhouse gases. At the United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro in 1992 the United Nations Framework Convention on Climate Change was agreed to, according to which developed countries would stabilize their emissions to 1990 levels by the year 2000 without any legal obligation. This voluntary approach was not successful; therefore, the main purpose of the Kyoto Conference in 1997 was to provide legally binding targets of emission reduction and dates. The Quantified Emission Limitation and Reduction Objectives agreed upon at this Conference established emission reductions, reductions of growth rates, and, in addition, allows joint implementation of the trading of emissions.

Since then, several formal studies have been published which analyze the problem of the appropriate allocation of emission permits to different states. Akira Okada [30] presented a cooperative market model of trading in CO_2 emission permits. The underlying theoretical model has been developed by Shapley [44] and others. The study does not describe actual negotiations but rather evaluates several alternative allocation rules of initial emission permits for the United States, Russia, and Japan using numerical analyses of permit trading.

Beeker [4] uses noncooperative sequential bargaining models to describe negotiations between two states such that one state offers payments to another state in the interest of joint implementation: the idea is that, as it is much more expensive to reduce further emissions in a developed country than in a developing country, from a global point of view it is much more effective for the former to help the latter. Beeker develops three models, differing according to the information available to the two states; he does not consider, however, specific pairs of states, nor does he use any kind of numerical data.

Kaitala and Pohjola [18] present a dynamic model for international negotiations on the global climate change problem. Their basic assumption is that countries differ in their vulnerability to global warming and that two coalitions may form. One coalition may include countries that do not suffer from global warming, or whose damages are minor, and the other coalition may comprise countries that will suffer. The greenhouse gas problem is then modeled as an economic infinite-horizon differential game. Countries negotiate an agreement among Pareto-efficient programs by allowing for transfer payments that take into account the existing international asymmetric effects of global warming. Transfer payment programs are designed in such a way that it is possible at any stage of the agreement to punish violations against cooperation and to discourage the other player from selfishly polluting the atmosphere. Numerical examples are given, but concrete negotiations are not described.

Furthermore, a broad and important area of applications deals with negotiations on the use of common-pool resources (CPR) in general. Among the many concrete cases that are dominant and are discussed in the published literature are distribution problems of water resources on the one hand and of fishing grounds on the other. Ostrom *et al.* [31] have extensively studied the field as a whole; they examined CPR problems from three perspectives, namely theoretical, laboratory experiments, and field settings. Various models of noncooperative game theory are used as a formal language for applying their Institutional Analysis and Development (IAD) framework to CPR problems, and they are compared to experimental findings. Even though, for example, fishing grounds in coastal waters bordering different states are discussed, the concrete negotiations considered are not international in the sense that governments are involved; the approach presented could, however, also be used for this problem.

International negotiations on the allocation of water resources are studied for example by Richards and Singh [37] with the help of a generalized Nash bargaining model. They consider two states with two domestic groups each (city dwellers and irrigated-land farmers), which represent different interests, and they study the interactions of domestic and international negotiations. As a concrete case they have in mind the Jordan river and the negotiations between Israel and Jordan, although they do not present numerical data. Another example of the use of a model to show the implications of various proposals was the model used in the United Nations Conference on the Law of the Sea (UNCLOS), analyzed by Antrim in this publication.

A further area of international negotiations for which formal analyses have been performed of the kind discussed so far, is that of arms races and escalation, arms control, and disarmament. Richardson ([38], [39]) was the first to describe arms races quantitatively and phenomenologically, using a system of coupled linear differential equations. His work was brought to a wider audience by the review article of

Rapoport [36]. The wealth of theoretical and empirical studies of arms races has been well documented in bibliographies [11] and literature reviews ([10], [22], [15], [16], [2]).

By its very nature the theory cannot explain the causes of armament spirals; it can only describe the mechanisms by which they proceed. Therefore, in the 1970s, normative Richardson-type analyses were performed depicting national decision makers as maximizing their citizens' social welfare subject to a resource constraint and a dynamic stock adjustment constraint. Brito [8] presented an optimal control analysis in which each of two rival states optimally allocate resources between civilian consumption and arms expenditure. Simaan and Cruz ([45], [46], [47]) extended this analysis to a differential game perspective in which competitive and interactive aspects of the rivals' optimization problem are more fully taken into account. An overview of this type of—mathematically demanding—analysis is given by Sandler and Hartley [41].

From the other side, emphasizing the conflictual character of arms races, various game-theoretical models have been developed in the last 20 years that, at least in principle, provide insight into the reasons why states behave in ways that can seem paradoxical; studies by James Morrow ([23], [24]), by Zagare and Kilgour ([52], [53]), and by Avenhaus, Beetz, and Kilgour [3] are representative of this kind of work. The authors largely use game-theoretical models with incomplete information, and they illustrate their findings with real-world examples, but they do not analyze major international negotiations as such. At least two further studies from recent years should be given as examples, namely, the analysis of the 1978 Camp David Negotiations [6], and that of the negotiations of the fate of the Soviet nuclear weapons deployed in the Ukraine until 1990, published in 1996 by Jehiel *et al.* [17].

Finally, we defined formal models *in* international negotiations. These models provide useful inputs of external data into negotiations, showing the effects of related processes and the implications of proposed solutions. Trade negotiations use such models to calculate the effects of tariffs on commerce at different levels. Another example in the climate and pollution area is IIASA's Regional Air Pollution INformation and Simulation (RAINS) model presented by Amann in this publication.

5.2 Organization of the book

The contributions in this publication are grouped into four sections. As an overture, the first part contains general evaluations of negotiation models from different points of view—philosophy of science, mathematics, and political science.

O'Neill, a political scientist at the University of California at Los Angeles, expands his earlier evaluation of the use of game theory models in the analysis of conflict and conflict resolution. Carment and Rowlands, international relations specialists at Carleton University, Ottawa, undertake a critical evaluation of formal modeling of three-party negotiations or mediation, finding an accumulation but not an integration of synthesis of knowledge. Wierzbicki, of the Polish National Institute of

Telecommunications, undertakes a critical evaluation from inside the game of theoretic models and of the rational choice approach on which they are based, as an application of the broader theory in the scientific sense. Druckman, a social psychologist at George Mason University, Virginia, discusses four modeling approaches: negotiation support systems and stochastic modeling being process-oriented, and decision analysis and game theory being outcome-oriented.

The second section contains studies *of* modeling of international negotiations, only one of which uses game theory. Gabbay, a social psychologist at Information Systems Laboratories in San Diego, California, constructs a mathematical model of small group decision making as a nonlinear process, showing implications for related processes in negotiation. Avenhaus and Krieger, mathematicians at the Federal Armed Forces University, Munich, model international negotiations in both extensive and normal forms and show with the help of two examples that if the preferences cannot be expressed as scalar ones the number of equilibria is large. Rudnianski and Bestougeff, decision specialists at the University of Reims, France, develop models of games of deterrence as a guide for negotiators and mediators. Güner, an international relations specialist at Bilkent University, Ankara, attempts to capture the reality of three-party negotiations through game-theoretic modeling, showing the responses that a rational actor will make in a hypothetical situation.

The third section is devoted to models *in* international negotiations, logical constructs of the subject matter showing the implications of various plans for dealing with it. Amann, a Systems Researcher at the International Institute of Applied Systems Analysis (IIASA), discusses the RAINS model and its use in informing and structuring negotiations on transboundary air pollution in Europe. Antrim, an engineer from the Massachusetts Institute of Technology (MIT), presents the MIT model of the implications of various proposals for the exploitation of undersea minerals that played an important role in the United Nations negotiations for a new Law of the Sea convention.

In the fourth section, models *for* international negotiations, four presentations show how specific formal models help find solutions to help or obviate negotiation. Okada, an economist with Hitotsubashi University, Tokyo, Japan, presents a noncooperative game for the analysis of carbon dioxide (CO_2) emissions negotiations under the Kyoto Protocol, showing the conditions under which an equilibrium among the parties' preferences exists. Kilgour, a mathematician from Wilfrid Laurier University, Ontario, Canada, presents a graph form of a game-theoretic analysis for use as a tool for negotiators in conflict resolution. Brams, political scientist at New York University, and Kilgour and Sanver, economists at Istanbul Bilgi University, specifically provide a game-theoretic solution for fair division that removes the need for strategy, which is a key element in negotiation. Schüssler, a philosopher at the University of Bayreuth, Germany, refines fair division ideas to analyze possibilities and eventualities at Camp David. Finally, Raith, an economist at the Otto von Guericke University, Magdeburg, Germany, also examines the fair division idea to inform conflict resolution procedures.

The study ends with an analysis of lessons for theory and for practice, as we seek to advance both the field of formal modeling and its application to international negotiation theory and practice.

For the benefit of formal modeling and its application to international negotiations in forthcoming years, based on our experience and our conviction, efforts in the following directions should be intensified:

- Formal modelers should make more of an effort to render their abstract concepts and results more understandable and plausible to scientists not trained to work with these models.
- Practically oriented modelers who develop such formal theories for real-world problems should make more of an effort to interact with those practitioners who are supposedly the customers for these theories.
- Practitioners in governments and international organizations should make more of an effort to understand and to use these models, eventually proposing modifications, extensions, or, probably most frequently, simplifications so that they can be applied successfully.

These are the self-assigned tasks that we have taken on in this work. The chapters present critical examples of models of, in, and for negotiation in order to test their applicability to understanding and practice. Not all of these models are based on game theory; other types simplify reality in other ways. But all purport to be based on an abstract design containing dynamic relations that are useful for negotiation. This volume does not cover all possible approaches; it is not an encyclopedia of applied modeling. The chapters are illustrative of the many approaches currently in use that seek to elucidate the high points in a broad range of activities. Optimally, both the mountains and the valleys will serve as a guide to bridging the gap between theory and practice and as a spur to further bridge-building efforts.

References

1. Aumann RJ, Hart S (eds 1994) Handbook of game theory, Volumes 1 and 2. North-Holland Publishers, Amsterdam, The Netherlands
2. Avenhaus R, Fichtner J (1984) A Review of Richardson Models. In: Avenhaus R, Huber R (eds) Quantitative assessment in arms control. Plenum Press, New York, USA, London, UK, 143–178
3. Avenhaus R, Beetz J, Kilgour DM (2005), Quantitative Models for armament escalation and negotiations. In: Faure GO, Zartman IW (eds) Escalation and negotiation. Cambridge University Press, Cambridge, UK
4. Beeker D (2000), Kompensationsregelungen im Klimaschutz und Opportunismus. PHD Dissertation, University of Siegen, Germany
5. Brams SJ (1985) Rational politics. Congressional Quarterly Inc., Washington DC, USA
6. Brams SJ, Taylor AD (1999) The win–win solution. W.W. Norton and Company, New York, USA
7. Brams SJ, Kilgour DM (1988) Game theory and national security. Basil Blackwell, New York, USA

8. Brito D (1972) A dynamic model of an armament race. International Economic Review 13(2):359–375
9. Bueno de Mesquita B, Lalman D (1992) War and reason – domestic imperatives. Yale University Press, New Haven, CT, USA
10. Busch PA (1970) Mathematical models of arms races. In: Russett BM (ed) What price vigilance. Yale University Press, New Haven, CT, USA
11. Cioffi-Revilla CA (1979) Mathematical models in international relations: a bibliography. Institute for Research in Social Science, Technical Paper No. 4
12. Dupont C, Faure GO (1991) The negotiation process. In: Kremenyuk V (ed) International negotiation – Analysis, Approaches, Issues. Jossey-Bass Publishers, San Francisco, CA, USA, 40–57
13. Faure GO (ed 2003) How people negotiate – resolving disputes in different cultures. Kluwer Academic Publishers, Dordrecht, The Netherlands, Boston, CT, USA, London, UK
14. Hopmann P with the collaboration of Druckman D (1991) Arms control and arms reduction: view I. In: Kremenyuk V (ed) International negotiation – Analysis, Approaches, Issues. Jossey-Bass Publishers, San Francisco, CA, USA, 269–288
15. Intriligator MD, Brito DL (1976) Formal models of arms races. Journal of Peace Science 2:77–88
16. Intriligator MD (1982) Research on conflict theory. Journal of Conflict Resolution 26, 307–328
17. Jehiel P, Moldovanu B, Stacchetti E (1996), How (not) to sell nuclear weapons. The American Economic Review 86(4):814–829
18. Kaitala V, Pohjola M (1995) Sustainable international agreements on greenhouse warming – A game theory Study. In: Carlo Carraro C, Filar JA (eds) Control and game-theoretic models of the environment. Birkhäuser, Basel, Switzerland, 67–87
19. Krasner S (ed 1983) International regimes. Cornell University Press, Ithaca, NY, USA
20. Kremenyuk V (ed 1991) International negotiation – analysis, approaches, issues. Jossey-Bass Publishers, San Francisco, CA, USA
21. Leinfellner W, Köhler E (eds 1998) Game theory, experience, rationality. Kluwer Academic Publishers, Dordrecht, The Netherlands
22. Mol KD, Luebbert GM (1980) Arms race and military expenditure models. Journal of Conflict Resolution 24:153–185
23. Morrow JD (1989a) Capabilities, uncertainty and resolve: a limited information model of crisis bargaining. American Journal of Political Science 33:941–972
24. Morrow JD (1989b) Bargaining in repeated crisis: a limited information model. In: Ordeshook P (ed) Models of strategic choice in politics. University of Michigan Press, Ann Arbor, MI, USA, 207–228
25. Morrow, James D. (1994) Game theory for political scientists. Princeton University Press, Princeton, NJ, USA
26. Mutoo, Abhinay (1999) Bargaining theory with applications. Cambridge University Press, Cambridge, UK
27. Myerson, RB (1997) Game theory – Analysis of conflict. Harvard University Press, Cambridge, MA, USA. Third Printing
28. Nash JF (1950) The bargaining problem. Econometrica 18:155–162
29. Nash, JF (1951) Non-cooperative games. Annals of Mathematics 54:286–295
30. Okada A (1999) A cooperative game analysis of CO_2 emission permits trading: evaluating initial allocation rules. Discussion Paper No. 495, Institute of Economic Research, Kyoto University, Kyoto, Japan

31. Ostrom E, Gardner R, James Walker J (1994) Rules, games, and common-pool resources. The University of Michigan Press, Ann Arbor, MI, USA
32. O'Neill B (1994) Game theory models on peace and war. In: Aumann RJ, Hart S (eds) Handbook of game theory. Vol. 2, North-Holland Publishers, Amsterdam, The Netherlands, 995–1053
33. Raiffa H (1982) The art and science of negotiation. Harvard University Press, Cambridge, MA, USA
34. Raiffa H (1991) Contributions of applied systems analysis. In: Kremenyuk V (ed) International negotiation. Jossey-Bass Publishers, San Francisco CA, USA
35. Raiffa, H, with John Richardson J and Metcalfe D (2002) Negotiation analysis: the science and art of collaborative decision making. Harvard University Press, Cambridge, MA, USA
36. Rapoport A (1957) Lewis F. Richardson's mathematical theory of war. Journal of Conflict Resolution 1:249
37. Richards A, Singh N (1997) Two Level negotiations in bargaining over water. In: Parthasarathy T, Dutty B, Potters JAM, Raghavan TES, Ray D, Sen A (eds) Game theoretical applications to economics and operations research. Kluwer Academic Publishers, Dordrecht, The Netherlands, Boston, MA, USA, London, UK
38. Richardson LF (1939) Generalized foreign politics. British Journal of Psychology, Monograph Supplement No. 23
39. Richardson LF (1960) Arms and insecurity. Boxwood, Pittsburgh, PA, USA
40. Rubinstein A (1981) Perfect equilibrium in a bargaining model. Econometrica 50:97–110
41. Sandler T, Hartley K (1995) The economics of defense. Cambridge University Press, Cambridge, UK
42. Selten R (1987) Models of strategic rationality. Kluwer Academic Publishers, Dordrecht, The Netherlands, viii
43. Selten R (1998) Game Theory, Experience, Rationality. In: Leinfellner W, Köhler E (eds) Game theory, experience rationality. Kluwer Academic Publishers, Dordrecht, The Netherlands, 9–34
44. Shapley LS (1967) On Balanced sets and cores. Naval Research Logistics Quarterly 14:453–460
45. Simaan M, Cruz J (1973) A multi-stage game formulation of the arms race and its relationship to Richardson's model. Modeling and Simulation 4:149–153
46. Simaan M and Cruz J (1975a) Formulation of Richardson's model of arms race from a differential game viewpoint. The Review of Economic Studies 42(1):67–77
47. Simaan M, Cruz J (1975b) Nash equilibrium strategies for the problem of armament races and control. Management Science 22:96–105
48. Spector BI, Zartman IW (eds 2003) Getting it done: post-agreement negotiations and international regimes, United States Institute of Peace (USIP) Press, Washington DC, USA
49. Stahl I (1972) Bargaining theory. Economic Research Institute, Stockholm, Sweden
50. Wagner, RH (1989), Uncertainty, rational learning, and bargaining in the Cuban missile crisis. In: Ordeshook PC (ed) Models of strategic choice in politics. The University of Michigan Press, Ann Arbor, MI, USA, 177–206
51. Young, HP (ed 1991), Negotiation analysis. The University of Michigan Press, Ann Arbor, MI, USA
52. Zagare FC, Kilgour DM (1993) Asymmetric deterrence. International Studies Quarterly 37:1–27
53. Zagare, FC, Kilgour DM (2000) Perfect deterrence. Cambridge University Press, Cambridge, UK

Part I

General Evaluations

Before embarking upon examples of the uses of formal modeling of, for, and in negotiation, some general critiques and evaluations should set the stage for subsequent discussion. With 50 years' experience in mathematics and decision theory, Wierzbicki strews the path with caveats about specificity, rationality, prescription, and complexity. Formal models find their best use in simplification, training, and description, and in developing both a practice and an understanding of intuition. However, as O'Neill shows, formal models, most specifically those derived from game theory, have overcome many of the early criticisms of their method and have begun to pick at the pieces of reality rather than being limited to oversimplified two-actor situations. Models have moved beyond an affirmation of known generalities and explored their conditions and consequences. But they are still captives of fixed assumptions and initial inputs, in contrast to creative transformations through the negotiation process. Game-theoretic and some other types of models focus on outcomes rather than process, as Druckman demonstrates, with an interaction between the two aspects of negotiation and the methodologies analyzing them still remaining to be explored. As Carment and Rowlands find, formal modeling methodologies have tended to remain within their own discourse, building a body of analysis based on previously verified findings and discarding disproven arguments, but falling short on integration with alternative methodologies or a synthesis of findings at different levels of analysis. Other types of modeling, as well as less formal methods, are concerned with process, the meat of the negotiation itself and the prelude to outcomes. While models *of* negotiation attempt to replicate process to explain outcomes, models *for* negotiation enter the process only by providing inputs for it, and models *in* negotiation circumvent the process by proposing optimal outcomes. As noted, an interaction among the three parts of negotiation remains the frontier for exploration.

Game Models of Peace and War: Some Recent Themes*

Barry O'Neill

Department of Political Science, University of California, Los Angeles, USA

In 1994 I surveyed game models of peace and war for the *Handbook of Game Theory*: this review summarizes some of the developments since that time. When the Cold War ended, concern over nuclear war yielded to broader issues, such as ethnic conflict, humanitarian intervention, nuclear proliferation, sanctions as tools of compellance, the establishment of stable democracies in new nations, globalization and trade, and the formation of supranational bodies. Current game theory applications on war and peace responded to these political changes, so that the variety of topics the field now treats has made it especially interesting. More than before, the literature can be seen as analyses of different kinds of international bargaining.

Game models have become more sophisticated, not just in their mathematics but in the political theory behind them and in their use of real data. In regard to the latter, their empirical testing has moved beyond economics or the study of legislatures and voting. Without the natural scales of money or votes, respectively, international relations game theory has had to develop new techniques for its empirical tests. These changes have contributed to a new role for game models within the larger field: before, they tended to formalize existing hypotheses in international relations, but now they are generating new ones.

The survey does not try to be complete, but follows some major themes, with selected references. Interesting work that will not be discussed includes the causes of civil conflict and the difficulty of negotiating its end (e.g., [79], [17]), economic sanctions (e.g., [43], [72], [73]), international treaties in general ([16], [21], [30]), and the viability of large states [1].

1 The Interaction of Domestic Politics with International Politics

A decade ago many critics claimed that games required that the players be nation states, and so they ignored domestic politics. Back then it was true that most papers

* The author would like to thank Kathy Bawn, Jeff Lewis, Barbara Koremenos, Ken Schultz, and Alex Wendt for their suggestions, and the Russell Sage Foundation for its support.

portrayed nation states as the actors, but that was the modeler's choice—it was not required by the method—and it probably reflected Cold War thinking. The citizenry of the Soviet Union and the United States (USA) generally stood behind their own governments' goal of deterring an attack from the other side, so modelers had no need to have several separate players within a state. Recently, the issues have been, for instance, whether the North Atlantic Treaty Organization (NATO) intervenes in Kosovo, or Britain joins the United States in invading Iraq, or whether United States (US) voters will support a prolonged foreign intervention, which brings in domestic actors, and newer studies have included them.

The notion that game models can span the domestic and international levels is not new. Eisner [10] presented a game on the Vietnam War in which the US population, the US government, and the Viet Cong were separate players. Intriligator and Brito [26] devised a coalitional game whose players were the governments, militaries, and populaces of the United States and the Soviet Union, and where tacit alliances could be formed, even between the two military establishments based on their common interest in high defense spending. However, current interest in modeling domestic politics is due especially to two authors: Putnam [60] writing on the idea of "two-level games" and Fearon ([12], [13]) on "audience costs." Putnam's article will not be discussed here because it was not formal (although the idea has been formalized by Iida [24], [25]). However, Fearon had an explicit model of a crisis as a war of attrition, with incomplete information. Each leader must decide in continuous time whether to attack, escalate, or back down. Backing down means that the leader pays an "audience cost," in the form of lower popularity or even expulsion from office, and the higher the escalation has gone, the greater the cost paid for backing down. It is through these audience costs, which are the public's conditional reactions against its own leader's action, that Fearon connects the international and domestic spheres. Setting up the costs is a leader's strategic move, brought about by declarations such as "this will not stand" or "our national honor and prestige are at stake." Fearon's model indicates that, given a crisis has started, its outcome is determined not so much by relative military strengths as by the two leaders' relative abilities to put these potential costs in place. They are a form of "tying one's hands" or "burning one's bridges," but since a leader pays them only in the case of backing down, they are different from costly signals of resolve, where the costs apply unconditionally. Fearon [15] compares hands-tying and costly signaling as alternative ways to establish credibility. He sets up a simple game in which either tactic can be used and shows that hands-tying is more efficient in a well-defined sense. He also shows that any equilibrium that survives the "intuitive criterion" will have a side tie its hands if and only if it is prepared to fight; thus, in particular, there will be no bluffing or equivocating. He observes that this result seems to go against real behavior, where leaders often make their threats equivocal but still carry them out. The United States has sometimes proclaimed a policy of "strategic ambiguity"—that if a "rogue state" uses weapons of mass destruction against it, it may or may not respond with nuclear weapons. This behavior must be explained by introducing further dynamics, such as a danger that a threat that is overly definite will force the other side to challenge it. Fearon argues that very little international bluffing occurs.

Some authors, including Wright [83], had seen democracies as being at a disadvantage in a crisis compared with dictatorships because they cannot enforce internal unity, but Fearon argues that they do better just because their leaders are more accountable to their citizenry and so can generate audience costs better. (The notion that democracies do better in bargaining in general because of domestic constraints has come to be called the "Schelling conjecture" [74].) Which form of government fares better is an empirical question, as is the previously mentioned prediction that military strength is less relevant once a crisis starts. Partell and Palmer [55] investigated these issues using coded databases on past crises and found Fearon's conclusions were generally supported.

In the 1994 model the only strategic actors were the leaders—for some reason the domestic constituencies had no foresight. If they found that a leader had bluffed and backed down, they turned him out of office even if there were no benefit in doing so, and in fact his action might have been in the country's interests. Although a country benefits from having a populace known to be ready to do that, actually doing it would be irrational and violate subgame perfection. Another challenge to a claimed advantage of democracies would be that audience costs are just as important for dictators, perhaps more important, since the latter may be less secure in office and face worse penalties if ejected.

Smith [72] addressed the irrationality issue by positing an electorate that takes a leader's unfulfilled promise as a sign of incompetence, and therefore dismisses him or her at the next election. President Bush, Sr., used strong rhetoric in 1990 but ultimately failed to eliminate Saddam Hussein, and this contributed to his defeat. Schultz ([65], [66], [67]) also addressed the question of a democratic advantage in several papers and a book. He postulated a two-party system where at the start of the crisis each party declares its position on the use of force. From these positions, the rival learns the attitude of the democracy's citizenry and uses this information to decide whether to push ahead or back off. For Schultz, the mechanism that generates credibility is not that leaders are afraid of losing their jobs, but that the multiple voices of a democracy demonstrate its resolve. His models have testable consequences for the relative success not only of democracies and dictatorships, but also when the hawkish party or the dovish party is in power, and he tests these and other predictions using crisis datasets. These papers bridge the international/domestic gap more explicitly than Fearon's earlier ones.

In Kurizaki's model [32] a state can make a threat in public or in secret. Past approaches have taken the attitude that it must be public to generate audience costs and thus be credible, but he points out a disadvantage, namely, that the recipients of threats may have such costs too, which make them reluctant to yield. He identifies ranges of parameters where the secret threat is better. Four game theory- or statistics-oriented authors, Bueno de Mesquita *et al.* [4], conducted a joint project (1999, 2000) on the interaction of domestic institutions and foreign policy. Their hypotheses relate to the democratic peace, the degree of caution democracies show in entering a war, the amount of resources they allocate to war, and other concepts.

If international relations theorists were to point to one important and applicable finding, it would be that democracies almost never fight each other. Wars have occurred between two nondemocracies or between a democracy and a nondemocracy, but there have been essentially none in modern times between democracies (using reasonable definitions of "war" and "democracy"). Called the "democratic peace," this idea was taken up by policymakers in the Clinton administration to justify enlarging NATO to include East European states and former Soviet republics, and was echoed in George W. Bush's rationale for invading Iraq. A literature has grown up to explain the finding, and several mechanisms have been suggested. One is that two democracies share norms of nonviolent conflict resolution so that they approach negotiations with an optimism and readiness to compromise; another is that they possess a broader distribution of power with more checks so it is harder for a single irresponsible person to start a war. Schultz's model generates another possible explanation, namely, that democracies are better at perceiving each other's true resolves and less often fall into war from a misconception that the other is bluffing.

The models of Sartori [63] and Guisinger and Smith [22] also involve leaders' concern about the credibility of their commitments, but are innovative in describing the mechanism underpinning credibility. The typical idea had followed Kreps and Wilson's analysis [31] of the chain-store problem, namely, that leaders have an internal degree of resolve embodied in their utility functions which others infer from their actions. One's reputation is thus others' opinion about one's goals and values. However, Sartori, as well as Guisinger and Smith, define reputation as a player's recent record of play. Reputation arises as part of the definition of the particular equilibrium they select in their repeated game, an equilibrium in which a player who is caught bluffing is punished for several periods. Guisinger and Smith ask: Does reputation for honesty reside with the country or with the leader? This question can be interpreted as one about which equilibrium is chosen. Credibility applying to the country means that a bluffing state will be punished by others' defections for a set number of periods, while reputation applying to the leader means that punishment lasts only until the populace, which is also a player, replaces the leader.

Tsebelis' theory [75] of veto players is another way to span domestic and international politics. He combines ideas from spatial representations of politics and coalitional-form game theory with some dynamic moves. Various actors, such as a country's president, assembly, and foreign negotiating partner, have ideal points in a policy space and seek an agreement that is preferable to the status quo. Some of them have the right to propose revisions to the status quo and others only to veto proposals. In the British system, for example, the cabinet proposes and parliament reacts, whereas in the United States, the legislature proposes and the president vetoes. Countries will have many veto players if several parties form the government. Tsebelis makes different predictions for the foreign relations exhibited by parliamentary and presidential systems in their international bargaining. Milner and Rosendorff [45], for example, developed these ideas in international political economy.

2 Statistical Tests of Game-Theoretical Models

Ten years ago a frequently repeated criticism of game models involved their empirical verifiability. The objection came in three versions: that the theories were untestable since beliefs and preferences could not be measured in real situations; that they were tautologous, saying simply that people pursue what they want and want what they pursue; or that the theories, in fact, had been tested and had failed. These ideas are mutually inconsistent, but some critics have used two of them together, or all three. In the last decade game models have generally been accompanied by reasonably sophisticated tests. Sometimes the evidence is presented as coded data or as a set or comparative case studies focused on the point of the model, and sometimes it is less systematic—the author will describe some historical events that illustrate the model—but the inclusion of specific, detailed evidence has become standard.

Attempts to test game models confront special problems. One problem involves how to measure the variables—how to assign utilities to leaders' goals or probabilities to their opinions. A second problem is that formal theories make precise predictions, while normal statistical methods are set up to test imprecise ones. In classical hypothesis testing, which is the dominant method in the social sciences, the hypothesis of interest, the one that the researcher believes, is imprecise, for example, one parameter is different from zero, or greater than another. The researcher does not specify what its value is. This is the complete opposite of the natural use of formal theory (e.g., predicting that a certain optimum value will be chosen), where the precise prediction is the one the research wants to maintain. The result is that, applied directly, classical tests reject the nonbelieved hypothesis in informal models, but reject the believed one in formal approaches. These two difficulties, scaling and statistical testing, are usually finessed at once by testing only the theory's comparative statics, not its precise predictions, for example, whether trade is associated with a decrease in war, not whether there was a specific degree of decrease. The model becomes a logical tool for generating hypotheses about associations.

A third difficulty is nonlinearity or discontinuity in a game's predictions. As the utility of a course of action increases, actors do not become gradually more likely to choose it but switch over abruptly as soon as it becomes optimal. Even if one added continuity by building in some kind of noise, it is easy to think of examples where the likely outcomes would not be monotonic with the parameters. Consider a country that gradually increases its military strength, other variables held equal, and consider the resulting likelihood of a war. When a country is especially weak it may be attacked; as its strength increases the danger that war may decrease; when it becomes a major power its strength may rise, but fall again when it becomes a hegemon. This complexity comes from the intricacy of intelligent agents—the different logical arguments they apply in the situations, but most statistical methods assume linearity or at least monotonicity.

A fourth difficulty is especially challenging: that subjects which are optimizing agents introduce a selection effect. A case can be in or out of the dataset depending on their decisions, and they will generate a biased sample. For example, a researcher hypothesizes democratic leaders who are cautious about foreign adventures, as they are more likely to be turned out of office if they are caught bluffing, and wishes to estimate the likelihood of a bluffer being turned out. That likelihood will be influenced by factors besides democratic government, and a leader may choose to bluff more often in just those situations that are favorable to survival. The observed relative frequency of being turned out will underestimate the strength of the effect.

A fifth difficulty has become known as the "zero probability problem." The typical game model implies that nonequilibria are never played at all, so a single violation renders the theory impossible. Since no theory claims to account for all relevant factors, this is too strict a test.

Smith [72] treated the question of whether nations hesitate to go to war against those with allies who are committed to helping them. Data indicate that only one alliance commitment in four is honored, but the selection effect noted above may be in play here, in that the denominator includes those cases where the aggressor went ahead and attacked, perhaps having reason to believe the commitment would not be honored. Smith offered a vigorous but not quite successful attack on the problem using an extant econometrics method, the Heckman procedure. Signorino ([68], [69]) and Smith [73] treated these issues with new methodologies. In reaction to the zero probability problem, Signorino [68] replaced the Nash equilibrium with the quantal response equilibrium (QRE) concept of McKelvey and Palfrey [44], which allows all moves with some probability. In a QRE, a player's choice has a random component beyond the parameters of the strategic situation, and Signorino interpreted this as being due to variables present in the utilities but unmeasured by the researcher. He developed ways to estimate the game payoffs from the data, along with procedures to test the models. A researcher identifies a set of observable variables relevant to a state's goals in the situation—democracy, degree of trade, gross national product (GNP), and so on—and represents the state's utility for each outcome as a linear function of these, plus an error term. For a game of complete information describing a challenge and response, he shows how to estimate the coefficients in each state's utility function from datasets of national traits and wars. Lewis and Schultz [40] extend the method to incomplete information games, using the idea that the two adversaries' knowledge about each other's utilities is based on the same set of variables as the researcher is considering. Their assumption that the researcher and the players have the same basis for estimating goals is interesting and sensible, but one drawback is that the estimation procedure is restricted to games with only one equilibrium.

Interest in testing has blossomed, but it is hard to say whether this is because of the game modelers' realization that it matters, or criticisms from the larger field. I believe a spur to much of this was Bueno de Mesquita's utility game theory models of war, starting around the 1980s (see, for example, [3]). A former participant in the Correlates of War project at the University of Michigan, he was imbued with the idea that variables must be operationalized so that theories can be tested statistically, and although his particular ways of doing that met criticism, he provoked others to try to improve on them.

3 Reconstructions of Extant Theory: Neorealism, Deterrence, and Arms Races

Reading informal studies on peace and war, one finds the same decision analysis framework as in game models: parties hold beliefs and choose courses of action to maximize their likely benefit. The accounts are more complex and detailed, but less precise and organized. The conclusions may not agree with formal treatments, especially when they undertreat the strategic element—putting oneself in the other's place to anticipate reactions to their moves. Game modelers have been like "logic police," correcting and refining extant informal theories. In an earlier book Powell [56] did this for deterrence and escalation, and *In the Shadow of Power* [57] he takes on neorealist theory. This is the notion that states set their international strategies by their concern for survival and pursue self-preservation in an anarchic world. In the past, game theory had dealt with neorealism piecewise, treating individual problems of deterrence, the security dilemma, or cooperation, but Powell aims to be more comprehensive. His overall conclusion is that many of neorealism's predictions do not follow from its premises; they have to be stated as conditional on the parameters of the situation. One neorealist issue concerns the resources of power: Are states maximizing their absolute gains or their gains relative to other states? The latter might be the goal if one were trying to keep ahead of rivals to avoid domination. Powell notes that a state has several responses to a threatening rival: to switch its resources from "butter" to "guns," to compromise with the rival, or to supplement its resources by acquiring allies. If it considers the strategy of allocating resources between economic growth and military expenditure, the goals of absolute and relative gains may not be in opposition, since increasing one's own wealth gives one more resources to defend oneself when needed. The best plan over time is a game solution that cannot be summarized in terms of absolute or relative gains.

Interest in deterrence and arms races has survived the end of the Cold War, as these situations continue to arise in international affairs, for example, the efforts to deter North Korea from developing nuclear weapons. Zagare and Kilgour's work [84] integrates the informal theory, focusing on issues of subgame perfection. Baliga and Sjöström [2] present an incomplete information model in which even a very small probability that a player is of the aggressive type triggers an arms race. Kydd ([33], [35]) constructs a game-theoretical model of arms racing, where states must judge whether the other is innately aggressive or is only being provoked to build its weapons by one's own arms and the dynamics of the security dilemma. Another of his models connects arms racing with the outbreak of war. The impetus for new weapons is not mutual fear of each other's weapons, but a desire to gain military advantage or force the rival into bankruptcy. It addresses the question of when arms control is mutually beneficial in such a situation.

Trust has been used to explain the relative economic and technological advancement of different societies, and Kydd applies the idea internationally ([34], [36], [38]). The second of these treats NATO enlargement. The proposal to add some former Soviet allies and territories was controversial, and positions on this seemed to cut across the political spectrum. Opponents pointed out that NATO was originally rationalized as a defense against the Soviet Union, so with that threat gone, the new Russia would see plans to expand NATO up to its borders as a threat. The pro argument was that it has always been more than a defense pact, that its Charter aimed at stabilizing internal relations among its members; then the worry was Germany, but now it is Eastern Europe. Expanding NATO reduces the likelihood of war in countries next to Russia's borders, so Russia should welcome it. Joining NATO incurs a cost for some countries, since it requires democratization, civilian control over the military, and the resolution of border disputes and of frictions with ethnic minorities. Kydd argues that joining is a costly signal that the new member expects to be a cooperative partner in the alliance. In his game the West either offers membership to a third party or does not, and the third party responds. Then, East and West play a bilateral trust game. Depending on the criteria for membership, enlargement might reassure Russia or threaten it.

4 Bargaining over the Start or Termination of a War

Following political changes the game analyses have tended to shift from nuclear war to war in general. The explicit emphasis on international bargaining has increased [58]. An example is Fearon's treatment [14] of how a war can break out between rational states. Since fighting involves a mutual loss of resources the two rivals will always have possible agreements that are mutually preferred, so why do they not settle on one of these beforehand? In his model, two states choose between war and a bargained outcome. Bargaining over the issue in dispute involves dividing a continuous commodity, but in a war one side or the other fully achieves its goal and each pays a cost of fighting. The probabilities of one side or the other winning the war are common knowledge, as are the costs. Given an assumption of risk aversion over the commodity, Fearon shows that there always exists a negotiated settlement preferable to fighting. Why then do they fight? His point is not that war is impossible, but that within the rationalist paradigm of goals, beliefs, and actions, one can list a minimal set of conditions that guarantees a peaceful settlement. This becomes a research method—if a war breaks out, one tries to pinpoint which of them was violated. The possible reasons for war are: (1) inability to credibly commit to a settlement; (2) inability to convincingly share what one knows with the other; and (3) indivisibility of the commodity under negotiation.

As expected, later authors have called for modifications. Slantchev ([70], see also [71]) argues that Fearon's assumption that war is a one-shot gamble is unrealistic, and he presents a model in which some equilibria have states going to war to inflict suffering on each other for the sake of bargaining advantage. Powell [59] describes general conditions in which states will be willing to fight, even with complete information. Fearon regards risk aversion as innately rational, but O'Neill [50] suggests that this is not always so. In many conflicts over sacred space, for instance, each party wants exclusive use of the space and an agreement to split it 50/50 would give each a lower expectation than a gamble for all or nothing. This implies in essence that they have risk-seeking utility functions. Further, in contexts where there is no quantifiable commodity, which is the usual case in international conflicts, the idea of risk aversion is not even coherent unless it is modified in certain ways and these changes no longer support the conclusion. Whether one fully accepts Fearon's theory or not, he helped organize the debate on the causes of war and has been cited broadly in the regular international relations literature.

Earlier papers tended to focus on bargaining to avoid war, but more recent research discusses bargaining during the war for the purpose of ending it. There is a good basis for this, since wars usually stop because the parties reach an agreement, not because one or the other can no longer fight on. The work goes back to Wittman [82] who argues informally that the relative military strength of the two combatants may affect the terms of a peace agreement but will not influence the ease with which they reach it. Conflicts of weak versus strong are no easier to end than those between equals, since if the weaker power becomes more conciliatory the other will become more demanding. Wagner [77] examines these ideas formally with a bargaining model in the spirit of Rubinstein. A state that refuses the other's offer of peace at a given stage must pay the cost of a continuation for some time and risk loss on the battlefield before it has another offer.

Werner [81] asks about the terms of a negotiated peace: To what extent do they reflect the military situation that has developed or the political aims that triggered the war originally? She notes some puzzling differences in settlement terms for different wars, for example, that in 1945 the United States required Japan's surrender to be almost unconditional, conceding only that it could keep its emperor, whereas it ended the 1991 Iraq war with essentially a return to the political status quo. The losers in both cases had been defeated militarily, so was the difference due to US war aims—that the United States felt satisfied with the Iraq outcome without occupying the country? Or did it have to do with the losers' relative bargaining positions at the end of the conflict? She sets up an alternating-offers model aimed at predicting the settlement. It is too difficult to measure original war aims and compare them with the final settlement, but indirect tests relating the length of war to the settlement indicate that both considerations, original war aims and the current military situation, have the effects expected from the model.

A growing literature involves international mediation. The concept of a mediator in Kydd [37] is a third party who can inform a disputant about the rival's resolve. A mediator whose goal is known as simply to get the conflict settled will not be credible; the mediator must be perceived as sharing a party's interests, as "biased" in a sense. O'Neill [52] looks at the opposite case, however, asking what a disinterested and powerless mediator can contribute. Game models show how the mediator can make an existing equilibrium focal, either by applying a known mediation procedure or by simply suggesting it, or he can introduce a new desirable equilibrium by a random device, or he can promote an agreement by filtering information between the parties, as in the models of Jarque *et al.* [27]).

International relations realists might regard mediation as part of power politics, not a transformation of parties' attitudes. Favretto [11] looks at the optimal degree of bias the third party should hold in favor of one of the actors, and comes up with a U-shaped curve. Middle degrees are worst since the mediator is not trusted by the disfavored party and not seen as motivated enough to make credible threats that it will enforce its way. She compares this mechanism with interview and archival data she gathered from officials in Bosnia and Slovenia.

5 Honor, Prestige, Face, Symbolism, Trust

My own work ([49], [53]) has applied game theory taking a view of international relations as a large-scale society. The premise is that national leaders deal with each other using rules of conflict similar to those they use within their own culture. For example, they worry about issues of honor, face, and prestige, etc., mediated through various devices that we are familiar with, such as symbolic messages, promises, and commitments, insults, apologies, challenges, and the invocation of norms. Honor is defined as an internal attitude of an actor, a desire to defend his family, territory, and group. It also has a social component in that, by definition, the actor wants others to see him as honorable. He adopts certain strategies to generate that belief, but others make their assessments knowing that the player may be pretending for the sake of others' perceptions, and this double-think situation leads to the use of game-theoretical methods. The approach can explain international violence that is otherwise seen as irrational. Social psychologists and some political scientists have paid attention to honor: Cohen and Nisbett [8] and Liberman [41] have cited the greater influence of honor in the American South as a reason for the international attitudes of some Southern political leaders. The socially perceived element of honor gives a theoretical basis for a domestic audience penalizing a leader for backing down. Differences across cultures in their rules make predictions about their relative abilities to establish credibility in crisis situations.

Face can also be defined within a game framework as a measure of the degree to which others defer to a party in direct interactions [49]. Unlike honor it is not about the party's internal type or others' opinions but about how the group has chosen to coordinate its behavior at the current equilibrium. *Prestige* differs from honor and face in that it involves second-order beliefs about someone's quality: a party has prestige for a quality if others believe that others believe that the party possesses it. O'Neill [53] investigates the prestige associated with nuclear weapons and shows some of its odd workings, for example, that giving positive evidence of one's quality can induce the audience to lower their estimate of that quality. This goes against the usual results, namely, that if one can show one's quality one had better do it or the audience will not believe it. The idea has consequences for why nuclear weapons are especially attractive as bearers of prestige and what the world community can do about it.

The work uses the recent idea of "global games" ([7]; [46]), this approach to incomplete information may have wider potential in international relations research. They are so-called because all parties have the same payoff function, but none is sure what it is; that is, all face the same situation but with different evidence. An example would be a proposed treaty or an economic sanctions regime that the players strongly want but that costs each of them a small amount to join. The regime must be supported by a certain threshold number of states to be effective—the benefit is a step function. Each party has different information about where that threshold is. If everyone were certain about the threshold's value there would be many equilibria, but with some degree of uncertainty as in the global game, the equilibrium will be unique, telling each player to join or not depending on the value of their signals. The assumptions are in contrast to typical incomplete information models where utilities' functions are different, and each one knows its own but is uncertain about the others'. (Two authors who have criticized the traditional assumption outside the global games approach and/or built models to the contrary are Iida, [23] and Koremenos [30].)

6 Terrorism

An important approach to terrorism was that of Sandler and Lapan [62] and others, who studied the substitution effect—the readiness of terrorists to respond to countermeasures by redirecting their efforts, with consequences that might be more or less harmful. The idea continues to be supported by statistical evidence (see Enders and Sandler [9] and their references).

Some papers have been prompted by the attacks of September 2001. Insurance companies were reticent to write policies on buildings liable to attack, since they had no basis for estimating a distribution function on the possible damage. Some industry consultants noted that this was not the usual problem of estimating a risk, like floods or fires, but one of calculating the expectation of a game: one between terrorists and defenders. In one model similar to the traditional Colonel Blotto game, the defender and attacker allocate their resources optimally against each other, typically with the defender moving first. The applied mathematics developed in the 1950s to calculate the effectiveness of defenses against intercontinental bombers and missiles gave an outcome. Geography has made the United States homeland generally safe from threats, the two exceptions being intercontinental missiles and terrorism, and it is interesting that the same kind of mathematics was used for both.

The consultants did not claim to have reached estimates reliable enough for real decisions, and I am skeptical that this will ever be possible. There is a popular attitude that game theory is like an omniscient computer that can generate precise advice, but for that purpose one would have to know the rules of the game better, and especially to understand the adversary's goals, which are a missing ingredient here. A better role of game theory in this area should be to force us to think of the adversary's goals. Analyses in the US press had much to say about al-Qaeda's possible plans, but there were few good treatments of purposes. Game analyses prompt thought about this. (A related idea voiced by critics is that game theory does not work with an irrational opponent, for example, one who uses the tactic of suicide bombing. This turns on which of the many meanings of "irrational" one is tapping. The relevant one here requires that terrorists perceive their behavior as instrumental toward some goal, which is certainly the case.)

Relevant to the Israeli–Palestinian conflict, Bueno de Mesquita [5] asks when a government should compromise with an internal revolutionary group. The official

US answer is never, since this gives perverse incentives to the terrorists, but he argues that compromising sometimes makes sense. Groups engaging in political violence are not homogeneous—they include elements that are more moderate or more extreme. When the government offers a concession in exchange for an end to violence, the moderate cells may accept it while the more extreme ones may refuse, and this split generates a positive and a negative consequence. Positively, the moderate group may restrain the extremists or provide intelligence that helps the government beat them. On the other hand, if the government cannot eliminate the extremist faction even with the moderates' help, the terrorist campaign will move into more violent hands. Bueno de Mesquita shows the factors that determine when a government compromise will help or hurt. A skeptic would probably cite a list of cases where government concessions led to worse violence, but he points out the selection bias here, that our mental dataset forgets just those compromises when the terrorist movement disappears, for example, Canada's Front de Libération du Québec. In a related paper [6], he also treats a government choosing the harshness of its antiterrorist measures, balancing directly preventing attacks against triggering greater recruitment. He offers an explanation for the puzzling observation that terrorism is correlated with lack of economic opportunities, yet most terrorist operatives are not from the poor sectors.

Kydd and Walter [39] discuss the use of violent acts as a tactic to scuttle peace negotiations. We are familiar with the cycle in the Middle East: Israeli–Palestinian relations improve and raise expectations, then extremists on one side or the other engage in a bombing, assassination, or other provocative acts, so that the moderates suspend negotiations. The moderates' reaction is the puzzle here, since they ought to see what the other side's hardliners are doing. However, they seem to play along, either voluntarily or not. One could attribute this to their emotional reaction to the attack, but Kydd and Walter suggest a strategic rationale. The moderates are uncertain about each other's good faith. If one signs a treaty granting territory in exchange for peace, will the adversary's leadership be satisfied or will it pocket the concession and then make greater demands? In their model, the Israeli government sees a relevant piece of evidence as whether a Palestinian leader will crack down on Hamas and the Islamic Jihad. It is uncertain whether continuing violence means that he is sincere in his promise but cannot control these elements, or is insincere and deliberately chooses not to control them. If the extremists commit violence with success, this increases likelihood in the eyes of Israel that he is insincere, and the government may decide to reject a compromise. In essence the extremists conducted attacks just when peace was near to force Arafat to signal a fact about himself that he would rather have concealed, but that Israel wanted in the open—whether his interest in an agreement was strong. Kydd and Walter note that before 1996, the attacks did not disrupt the peace process, but Palestinian elections in that year gave Arafat a strong mandate, 88 percent of the vote. Strategically timed violence would become more successful in blocking peace, according to their theory, since Arafat would be seen as more likely to have control. Their formal model presents a counterintuitive conclusion, that if Arafat is seen as powerless over home affairs the extremist attacks will be less damaging to the peace process, and they can cite evidence that this is true.

The notion that honor is relevant is strong in many of the societies that have generated terrorist violence. From the viewpoint of other societies their actions have seemed irrational and malevolent, but from an honor viewpoint they seem to follow the old script regarding how to respond to attacks on the honor of one's group. O'Neill [51] looks at special problems of mediating disputes over honor, where one's concession is costly even if no agreement is reached, since it demonstrates one's willingness to compromise over honor.

A related topic regarding terrorism, a possibility for future modeling, was suggested by Sagan (personal communication). Until now the largest nuclear powers had a fairly reliable system to keep control of their weapons during a crisis, partly because of technical devices such as "permissive action links" (PALs), partly because of their ability to afford good security measures, and partly because of their traditions of civilian control over the military. These factors may be less significant in countries that have just acquired nuclear weapons or may obtain them soon, such as India, Iran, North Korea, or Pakistan. Combining this lack of security with the consideration that their weapons may be more vulnerable to an attack than the US and Soviet missiles, which were based in silos or submarines, the newer nations' response in a crisis would be to disperse their nuclear weapons. Dispersed weapons may be safer from an external attack, but they are more vulnerable to capture by violent organizations within the society. Sagan terms this the "vulnerability paradox."

7 The Structure of International Governmental Organizations

With the end of the Cold War, the United Nations (UN) Security Council has seen fewer vetoes and been able to conduct more interventions. A scenario has repeated itself—the United States seeks approval from the Security Council for some security-related action, with option of either intervening with the help of certain allies or of going it alone. It prefers a Security Council endorsement to reduce the expense of the intervention and to increase support at home. The permanent members of the Security Council may have their own motives in their choice—they see it as a basis of their international power and do not want to see it bypassed on important events. Using the structure of ideal points, political space, and veto power, Voeten [76] offers a simple analysis of the situation, well-grounded in political examples.

One theme has been the fair allocation of power in an international voting body, like the Security Council or the European Union. Some authors have applied voting power indices such as the Shapley–Shubik or Banzhaf values. In the European Union, voting is usually by weighted majority, ranging from three votes for Malta up to a ceiling of 29 for France, Germany, Italy, and the United Kingdom. The rule for most topics is a double majority, that the approving states must reach 72 percent of the total population and 50 percent of the total number. Many papers have calculated countries' powers according to various plans, and there are now Web sites where one can input voting weights and get back exact Shapley–Shubik or Banzhaf values.

O'Neill and Peleg [54] discuss double majority voting for international organizations whose aim is to balance equal status for each country with greater rights for larger contributors. They show that the method is free of certain flaws that afflict other ways of doing this, for example, requiring that each member get a certain minimum voting weight or setting aside a proportion of the voting weight to divide equally. The latter systems seem attractive on their face, but they can sometimes increase inequality in power, using any sensible measure of that notion.

A continuing debate is the membership and voting rules of the UN Security Council. Kerby and Göbeler [29] explore the implications of various changes, using power indices. The small and non-Western countries in the UN have sought to increase the number of nonpermanent members beyond the current ten, but in fact this would likely decrease their total power as the large veto players would have a bigger pool to dip into to find the required majority [48]. The power of small countries on the Security Council depends very much on the majority quota being large enough, so the issue becomes one of setting it higher, but not so high as to paralyze the Council.

Several authors have joined in a debate over the validity of power indices for the European Union (e.g., Garrett and Tsebelis, 2001 [20]; Felsenthal and Machover, 2001 [19]). Garrett and Tsebelis object that the method assumes symmetry among the legislators, whereas they almost always have special relationships, either alliances or common attitudes, toward the issues. This is a problem that can be fixed with some extra modeling. Shapley offered a solution in the 1970s by producing an elegant theory depicting voters' attitudes as vectors in Euclidean space and proposals as points. The degree of preference was then the projection of the points on the vector. He was unsuccessful in getting it published in the major political science journal, but since then it has seeped into the literature little by little from applications by others (e.g., [61]). A related example deals with the UN Security Council [48], where the political space is constructed from countries' previous voting patterns in the UN General Assembly.

Maggi and Morelli [42] give an important treatment of the choice of voting rules in international organizations. NATO and certain others require unanimity, whereas many UN agencies or the International Monetary Fund have majority or weighted majority rules. The special aspect of votes in international organizations is that the results are unenforceable, and the losing members will abide by them only for the prospect of future cooperation. The authors use a repeated game to identify the range of parameters (the discount factor and correlation between parties' interests) where one or another voting rule will be chosen.

8 Critiques of International Relations Game Theory

A decade ago opponents of game modeling had used various arguments against it [47], and it is interesting to see how these arguments have changed. Recent criticism has expanded from the intellectual content to issues of academic politics. In 2000 some nonmathematically oriented scholars, communicating over the Internet, formed the organization Perestroika. Their premise was that political science had been hegemonized by quantitative and formal methods and that the main journals had confused technical prowess for professional competence and were excluding articles using qualitative or historical methods. Departmental committees were using publication in these journals as a criterion for hiring and promotion, so doctoral students were being forced to use game theory in their theses. Fascinated by theorems and numbers, political scientists were avoiding taking any stand based on ethics so their research had become irrelevant to real problems. The Perestroika movement included different political science fields and aimed at setting these differences aside and working for the inclusion of qualitative methods. Its name referred to the "restructuring" that broke Communist control in the Soviet empire, but most supporters did not call for excluding mathematical methods entirely, only for restoring a balance. It can point to some victories: members installed as officers in the national political science organization and somewhat more qualitative articles in the main journals.

Walt [78], a prominent Perestroikan, published a critique of international relations game theory that reflected the group's attitudes. He raised these criticisms:

- When simple games were used for illustration they were understandable, but now the work has become unnecessarily mathematical.
- The area's conclusions were already known from the work of regular scholars.
- It ignores empirical testing.
- It has promoted a "cult of irrelevance," where its practitioners have nothing to say about real international issues.

The last claim is puzzling. On the face of it, treating the causes of war would not seem to be glorifying irrelevance. It seems to hinge on the other three claims, in that a field is irrelevant if it contributes no understandable, new, or well-founded conclusions. The charge of no empirical testing had some validity a decade ago, but the present survey disproves it. The special difficulties encountered in testing game models, in particular, the selection bias, scaling problems, and the inapplicability of current statistical methods, may be why it took extra time to reach this point. Walt's claim that the conclusions just repeat known facts rests on an incomplete appreciation of what these conclusions are. His account of costly signaling, for example, misses the central idea. Regarding the originality of costly signaling, he is right that it had already appeared informally—one can find it in the work of Jervis (e.g., 1989) and even in Thucydides, as Schelling noted [64]. However, it is one thing to enunciate a mechanism in broad terms and another to deride it rigorously, determine the conditions under which it does or does not hold, and draw out its consequences. If the costly signaling idea has been stated before, it has also been misstated and even denied. Most of the concepts in the regular literature have a long history but continue to be discussed and developed. Walt acknowledges these considerations but still ends up calling the field "old wine in new bottles." His critique also ignores the base rate: relative to the whole field of peace and international security, the number of game modelers is tiny and their results so far have been out of all proportion to their numbers. Walt can be commended for going beyond generalities and discussing specific models, but his points generally do not inform or help modelers.

A related intellectual movement in the social sciences is constructivism, which views social relations as having less to do with objective facts as "constructed" by the actors. Wendt ([80], see also Fearon and Wendt [18]) critiqued game theory international relations from a constructivist viewpoint. Essentially his points are:

- That in the minds of its users, game theory often shifts from a method to an ontology. The idea that actors choose among courses in pursuit of interests slips over to a mindset that this is the whole of human interaction.
- That game theory, in general, takes preferences as given. Since change in preferences is an important source of social progress, the approach reinforces the status quo.
- That, more broadly, since actors' beliefs and desires are determined by their social relationships, an approach that starts with the individual and then works outward to society will miss the important causal factors.

Wendt does not address specific models but compared to Walt, his criticisms are more directed to the intellectually distinctive parts of applied game theory. His first two claims seem valid to a degree. An example would be the game-theoretical approach to mediation, which might focus on conveying information between the parties, in contrast to a major movement among practitioners called "transformative mediation." The remedy for the tendency of one's viewpoint to narrow is greater interaction with other approaches to one's subject. His third criticism, that the theory is reductionist, is hard to evaluate without considering an alternative, and right now there is no theory of human interaction that is as precise and that does not start at the individual level.

9 Conclusions

The field provides an interesting example of intellectual change within a discipline. What is behind recent developments such as the expansion of issues treated, the emphasis on domestic politics, and the focus on testing? One could weave a story that simply connected the domestic shift to the internal logic of the field. Credibility had been the ongoing problem with deterrence theory, and introducing domestic factors was the logical solution arrived at—a consequence of the internal debate. One could look for explanations in the sociology of the profession, that certain individuals with connections to others introduced new ways. I believe that alarm over nuclear war during the 1980s drew people into the field, and those who were most interested stayed to contribute their own research and to train graduate students. The winnowing has produced high quality research with more sustained efforts on serious problems.

References

1. Alesina A, Spolaore E (2003) The size of nations. MIT Press, Cambridge, MA, USA

2. Baliga S, Sjöström T (2004) Arms races and negotiations. Review of Economic Studies 71:351–369
3. Bueno de Mesquita B, Lalman D (1992) War and reason. Review of Economic Studies 71:351–369
4. Bueno de Mesquita B, Morrow J, Siverson R, Smith A (1999) An institutional theory of the democratic peace. American Political Science Review 93:791–808
5. Bueno de Mesquita E (2005a) Conciliation, counterterrorism, and patterns of terrorist violence. International Organization 59:145–76
6. Bueno de Mesquita E (2005b) The quality of terror. American Journal of Political Science 49:515–530
7. Carlsson H, van Damme E (1993) Global games and equilibrium selection. Econometrica 61:989–1018
8. Cohen D, Nisbett R (1996) Culture of honor: the psychology of violence in the south. Westview, Boulder, CO, USA
9. Enders W, Sandler T (2002) What do we know about the substitution effect in transnational terrorism? In: Andrew S, Ilardi G (eds) Research on terrorism: trends, achievements, failures. Routledge, New York, USA
10. Eisner R (1968) War and peace: a new view of the game. Department of Economics, Northwestern University, Evanston, IL, USA
11. Favretto K (2004) Does impartiality matter? Bias and coercion in major power mediation. Working paper, Department of Political Science, University of California, Los Angeles, USA
12. Fearon J (1994a) Domestic political audiences and the escalation of international disputes. American Political Science Review 88:577–92
13. Fearon J (1994b) Signaling versus the balance of power and interest: an empirical test of a crisis bargaining model. Journal of Conflict Resolution 38:236–69
14. Fearon J (1995) Rationalist explanations for war. International Organization 49:379–414
15. Fearon J (1997) Signaling foreign policy interests: tying hands versus sinking costs. Journal of Conflict Resolution 41:68–90
16. Fearon J (1998) Bargaining, enforcement and international cooperation. International Organization 52:269–305
17. Fearon J, Laitin D (1996) Explaining interethnic cooperation. American Political Science Review 90:715–35
18. Fearon J, Wendt A (2002) Rational choice and constructivism: a skeptical view. In: Carlsnaes W, Risse T, Simmons B (eds) Handbook of international relations. Sage, Thousand Oaks, CA, USA
19. Felsenthal D, Machover M (2001) Myths and meaning of voting power: comments on a symposium. Journal of Theoretical Politics 18:31–97
20. Garrett G, Tsebelis G (2001) Even more reasons to resist the temptation of power indices in the EU. Journal of Theoretical Politics 13:99–105
21. Gilligan M (2004) Is there a broader-deeper trade-off in international multilateral agreements? International Organization 58:459-484
22. Guisinger A, Smith A (2002) Honest threats: the interaction of reputation and political institutions in international crises. Journal of Conflict Resolution 46:175–200
23. Iida K (1993a) Analytic uncertainty and international cooperation: theory and application to international economic policy coordination. International Studies Quarterly 37:431–57
24. Iida K (1993b) When and how do domestic constraints matter? Two-level games with uncertainty. Journal of Conflict Resolution 37:403–26
25. Iida K (1994) Are negotiators hawkish of moderate-negotiator selection in two-level games. Sociological Theory and Methods 9:3–20

26. Intriligator M, Brito D (1984) A game theory model of the military industrial complex. Department of Economics, University of California, Los Angeles, USA
27. Jarque X, Ponsati C, Sakovics J (2001) Mediation: incomplete information bargaining with filtered communication. Working paper, Departament d'Economia i Ha, Universitat Autonoma de Barcelona, Spain
28. Jervis R (1989) The meaning of the nuclear revolution. Cornell University Press, Ithaca, NY, USA
29. Kerby W, Göbeler F (1996) The distribution of voting power in the United Nations. In: Huber RK, Avenhaus R (eds) Models for security policy in the post-cold war era. Nomos, Baden-Baden, Germany, 221–30
30. Koremenos B (2005) Contracting around international uncertainty. American Political Science Review 99:549–565
31. Kreps D, Wilson R (1982) Reputation and imperfect information. Journal of Economic Theory 27:253–279
32. Kurizaki S (2004) Efficient secrecy, public versus private threats in international diplomacy. Working paper, Department of Political Science, University of California, Los Angeles, USA
33. Kydd A (1997) Game theory and the spiral model. World Politics 49:371–400
34. Kydd A (2000a) Trust, reassurance and cooperation. International Organization 54:325–57
35. Kydd A (2000b) Arms races and arms control: modeling the hawk perspective. World Politics 44:228–44
36. Kydd A (2001) Trust building, trust breaking: the dilemma of NATO enlargement. International Organization 55:808–28
37. Kydd A (2003) Which side are you on? Bias, credibility and mediation. American Journal of Political Science 47:597–611
38. Kydd A (2005) Trust and mistrust in international relations. Princeton University Press, Princeton, NJ, USA
39. Kydd A, Walter B (2002) Sabotaging the peace: the politics of extremist violence. International Organization 56:263–96
40. Lewis J, Schultz K (2003) Revealing preferences: empirical estimation of a crisis bargaining game with incomplete information. Political Analysis 11:345–67
41. Liberman P (2002) Honor, arms and war. Paper presented at the Annual Meeting of the American Political Science Association, Boston, MA, USA, September
42. Maggi G, Morelli M (2004) Self-enforcing voting in international organizations. Working paper, Department of Economics, Princeton University, Princeton, NJ, USA
43. Martin L (1992) Coercive cooperation: explaining multilateral economic sanctions. Princeton University Press, Princeton, NJ, USA
44. McKelvey R, Palfrey T (1995) Quantal response equilibria for normal form games. Games and Economic Behavior 10:6–38
45. Milner H, Rosendorff P (1997) Interests, institutions, and information. Princeton University Press, Princeton, NJ, USA
46. Morris S, Shin HS (2003) Global games: theory and applications. In: Dewatripont M, Hansen L, Turnovsky S, (eds) Advances in economics and econometrics. Cambridge University Press, Cambridge, UK
47. O'Neill B (1994) Game theory models of peace and war. In: Aumann R, Hart S (eds) Handbook of game theory. Elsevier Science Publishers, New York, USA
48. O'Neill B (1996) Power and satisfaction in the United Nations Security Council. Journal of Conflict Resolution 40:219–37

49. O'Neill B (1999) Honor symbols and war. University of Michigan Press, Ann Arbor, MI, USA
50. O'Neill B (2001) Risk aversion in international relations theory. International Studies Quarterly 45:617–40
51. O'Neill B (2003) Mediating disputes over honor: lessons from the era of dueling. Journal of Institutional and Theoretical Economics 159:1–19
52. O'Neill, B (2004) What can a disinterested, powerless mediator do? American Political Science Association, Chicago, IL, USA
53. O'Neill B (2005) Nuclear weapons and national prestige. Manuscript, University of California, Los Angeles, USA
54. O'Neill B, Peleg B (2000) Reconciling power and equality in international organization: a voting method from Rabbi Krochmal of Kremsier. Jewish Political Studies Review 12:3–14
55. Partell P, Palmer G (1999) Audience costs and interstate crises: an empirical assessment of Fearon's model of dispute outcomes. International Studies Quarterly 43:389–405
56. Powell R (1990) Nuclear deterrence theory. Cambridge University Press, Cambridge, UK
57. Powell R (1999) In the shadow of power. Princeton University Press, Princeton, NJ, USA
58. Powell R (2002) Bargaining theory and international conflict. Annual Review of Political Science 5:1–30
59. Powell R (2004) The inefficient use of power: costly conflict with complete information. American Political Science Review 98:231–41
60. Putnam R (1988) Diplomacy and domestic politics: the logic of two-level games. International Organization 42:427–60
61. Rapoport A, Golan E (1985) Assessment of political power in the Israeli Knesset. American Political Science Review 79:673–92
62. Sandler T, Lapan H (1988) The calculus of dissent: an analysis of terrorists' choice of targets. Synthese 76:245–61
63. Sartori A (2002) The might of the pen: a reputational theory of communication in international disputes. International Organization 56:121–49
64. Schelling T (1984) Confidence in crisis. International Security 8:55–66
65. Schultz K (1998) Domestic opposition and signaling in international crises. American Political Science Review. 92:829–44
66. Schultz K (2001a) Democracy and coercive diplomacy. New York, USA, Cambridge, MA, USA
67. Schultz K (2001b) Looking at audience costs. Journal of Conflict Resolution 45:32–60
68. Signorino C (1999) Strategic interaction and the statistical analysis of international conflict. American Political Science Review 93:279–97
69. Signorino C (2003) Structure and uncertainty in discrete choice models. Political Analysis 11:316–44
70. Slantchev B (2003a) The power to hurt: costly conflict with completely informed states. American Political Science Review 97:123–33
71. Slantchev B (2003b) The principle of convergence in wartime negotiations. American Political Science Review 97:621–32
72. Smith A (1998) International crises and domestic politics. American Political Science Review 92:623–38
73. Smith A (1999) Testing theories of strategic choice: the example of crisis escalation. American Journal of Political Science 43:1254–83
74. Tarar A (2001) International bargaining with two-sided domestic constraints. Journal of Conflict Resolution 45:320–40

75. Tsebelis G (2002) Veto players: how political institutions work. Princeton University Press, Princeton, NJ, USA
76. Voeten E (2001) Outside options and the logic of security council action. American Political Science Review 95:845–58
77. Wagner H (2000) Bargaining and war. American Journal of Political Science 44:469–84
78. Walt S (1999) Rigor or rigor mortis? Rational choice and security studies. International Security 23:5–48
79. Walter B (2006) Information, uncertainty and the decision to secede. International Organization
80. Wendt A (1999) The social theory of international relations. Cambridge University Press, Cambridge, UK
81. Werner S (1998) Negotiating the terms of settlement: war aims and the bargaining. Journal of Conflict Resolution 42:321–43
82. Wittman D (1979) How a war ends: a rational model approach. Journal of Conflict Resolution 21:741–61
83. Wright Q (1942) A study of war. University of Chicago Press, Chicago, IL, USA
84. Zagare F, Kilgour M (2000) Perfect deterrence. Cambridge University Press, Cambridge, UK

Formal Models of Intervention: A Stocktaking and Analysis of the Implications for Policy*

David Carment and Dane Rowlands

Norman Paterson School of International Affairs, Carleton University, Ottawa, Canada

> International-relations theory can identify and frame important questions, but Pentagon and State Department officials will probably always be more interested in detailed case studies, prepared by area-studies experts. Theorizing about the causes of war might occasionally generate clean, law-like propositions that appeal to policymakers. But more typically, the discipline generates broad patterns that can be applied to particular cases only with a great deal of caution.
>
> "We have to recognize that there are limits to the predictive powers of political science," says Robert J. Art, "That's not an excuse to be sloppy. It's just to say that we don't have unified grand theories of many phenomena, especially not something as complex as war. None of us can predict the consequences of what will happen in the Middle East. Maybe this is why policymakers don't pay much attention to academics." [23].

1 Introduction

This chapter is a stocktaking exercise to evaluate formal models of third-party intervention in order to determine what can be done to improve the possibilities that formal techniques of evaluation can have a broader policy-relevant audience and impact. We do this in three stages. First, we consider a list of claims in the literature with respect to the way in which game theory evaluates the impact of third parties on conflict processes. O'Neill's chapter (in this publication) provides an excellent overview of recent contributions in the general literature on conflict; our goal is to complement that analysis by providing a profile of the current intervention literature in order to identify gaps, controversies, and areas of consensus in "intervention and mediation theory" as a research program.

* Portions of this chapter appear in an article entitled Formal Models and Intervention: Success as a Research Program and Policy Relevance, *Whitehead Journal of Diplomacy and International Relations*, Winter/Spring 2007, Vol VIII, No. 1.

Second, we examine formal modeling of mediation and intervention in terms of its effectiveness as a research program with a common set of research questions, appropriate levels and units of analysis, and empirical content. Here we consider the methodological divisions and similarities in the literature and the strengths and weaknesses of formal modeling. More specifically, we will consider what, if any, efforts have been made within formal modeling to confront key objectives within the broader research program: accumulation (building on previous findings and modifying or discarding arguments for which empirical support is lacking); integration (drawing on alternative methodologies that provide similar findings in a different context), and synthesis (using a multilevel of analysis approach).

We find that formal modeling has succeeded to a limited extent in addressing accumulation, but it has not successfully addressed integration or synthesis. We argue that if formal modeling is to be rendered accessible to policymakers and non-experts, more effort should be made to address questions of integration and synthesis. Examples in support of such an approach are provided. Pushing this idea even further, it might be suggested that formal modeling could serve as an answer to the problems identified by the editors of this publication because it offers a more explicit, concise, and dynamic account of the negotiation and intervention process.

Of course, some critics object to formal modeling on epistemological grounds. Urry [47], for example, argues that the existence of nonlinearity and complexity in social systems can lead to an absence of proportionality between cause and effect, and divergences between the individual and aggregate-level behavioral responses—the "fallacy of composition." Consequently, Urry argues, system complexity inhibits or prevents the creation of predictive models in international relations because of the inherent difficulty of capturing all potential explanatory factors and the potential gravity of omitted variables. In addition, complications such as imperfect information add further complexity to any formal modeling and predictive endeavors [10].[2] Clearly, there is a deep epistemological divide between those who reject formal approaches to theory building and those who consider it an essential dimension of research (as O'Neill also discusses).

We contend that even if one rejects the value of formal models as a means of investigating causal relationships under an idealized set of well-specified assumptions, they can at least be used to identify plausible behavioral regularities that can be tested against the evidence in order to guide policy.

In the third part of this chapter we draw out the implications of analysis for policy-relevant research. We suggest that by addressing the aforementioned issues of accumulation, integration, and synthesis, formal modeling can have a substantial

[2] These arguments are not universally accepted. In some disciplines, such as economics, it is standard practice to derive insights into complex social relationships by appealing to stylized but formal representations of what are believed to be their key facets. In some cases complexity can be represented through formal models. For example, Campbell and Mayer-Kress [10] examine how political behavior can be modeled using the sorts of nonlinear dynamic equations that represent chaotic behavior. In other cases it may be necessary to rely on large sample properties of aggregate behavior to overcome the random or idiosyncratic actions of individuals.

policy-relevant impact. This view is consistent with the approaches espoused by Druckman and Serdar (in this publication) in which the viability of different models as a way of improving decision making and analytical capacity is systematically compared. Furthermore, this approach is also consistent with recent work on negotiation and mediation. Kleiboer [29], for example, argues that modeling produces an incentive for more theory-driven empirical research, which, by her assessment, is more informative for the researcher and the policymaker. We conclude with an assessment of the implications for policy and directions for further research.

2 Findings from Recent Research

We begin with an attempt to understand the meaning of third-party intervention and why its policy relevance is important. It should be obvious that intervention is not just about mediating civil wars ([52], [33], [51]). In practice, the primary responsibility of a third party is usually to prevent destructive conflict through a variety of noncoercive political channels and actions. This view is consistent with the principles of third-party assisted negotiation and deterrence wherein the key strategic goal is the active participation of a third party to de-escalate conflict before it becomes violent or to prevent the recurrence of such violence [14]. Such an approach may prove ineffective after hostilities break out and violence is widespread. In reality, there is little agreement within the discipline on the kinds of bargaining strategies necessary for the termination of conflicts.

Pruitt's [38] *The Tactics of Third-Party Intervention* discusses the processes by which third parties can contribute to conflict resolution with intransigent parties. He discusses the Oslo talks between Israel and the Palestine Liberation Organization; the London talks that established Zimbabwe; intermediaries, chains, and track-two diplomacy; the Northern Ireland peace process; the importance of secrecy; and chain shortening and final negotiations. Pruitt contends that third parties need to use heavier tactics in protracted conflicts via a series of interlinked steps of "chain-shortening": serious conflicts are often moved to the verge of settlement by chains of intermediaries building on track-two diplomacy. The chains tend to get shorter as intermediaries drop out. Secrecy is usually essential for resolving these severe conflicts. He concludes that long-standing ethnic conflicts can frustrate the best efforts of even the most powerful interlocutors, but if third parties remain patient and on the lookout for the ripe moments, these tactics can then be highly effective.

Pruitt's research highlights several interrelated problems in the literature. First, there is a significant conceptual and theoretical problem of identifying the independent effects of individual strategies.[3] Some see the process as one of bargaining and negotiating with belligerents in interlinked graduated strategies; while others

[3] For the purposes of this investigation, third parties, whether they are states or multilateral actors, are those that have an interest in creating a stable environment in which peace can be nurtured and in developing a durable framework for a lasting negotiated settlement. These strategies can be augmented by coercive efforts that are essentially concerned with the recurrence, cessation, and prevention of violence.

see passive and forceful measures being applied concurrently. These differences of view are in part due to conceptual confusion regarding the term intervention. Furthermore, mediation, third-party intervention, and even negotiation, are often used interchangeably and the term intervention itself has different meanings with respect to technique, type of intervener, strategy, and outcomes. Intervention does not refer simply to the physical presence of a "managing agent" intent on using coercion to dissuade belligerents from using force to solve their differences ([16]:358). Nor has intervention been confined to involvement by states or organizations through military means. Third-party intervention encompasses a broad range of techniques, although it is hard to find agreement on what these might be ([3], [39]). For example, Fisher [21] draws a line between those strategies that are clearly pacific (such as conventional peacekeeping, track-two diplomacy, and consultation) and those that are not (such as peace enforcement).[4] He argues that the choice of third-party strategy is dependent on the nature of the strategies with which it must interact, and therefore there is no single best approach.[5]

Second, as Wall and Druckman [48] point out, dispute severity has a strong effect on an intervener's choice of techniques.[6] Most of the literature regards third parties as exogenous to the conflict, maintaining considerable influence over the conflict and remaining outside of it. In reality, there is a dynamic process of interaction between belligerents and intervener such that intervention techniques are endogenous to the conflict process but are often not treated as such. Zartman and Touval [53] make this point in their research on conflict triangulation. Although, in theory, third parties can induce negotiated settlements through the creation of hurting stalemates, they must be careful not to get caught up in the interests of the conflicting parties. In some instances, biased interventions may be unavoidable but they can also be productive in cases where one party might be eliminated (or at least have its power sufficiently reduced). A third party can reduce the chances of a more powerful party overwhelming its opponent. Through such pressure tactics, third parties can speed up the movement toward a settlement through the imposition of deadlines as well as integrated and cross-cutting strategies.

Third, there is interdependence and contingency among different strategies applied simultaneously or consecutively, and this interdependence is difficult to sort out conceptually and empirically. Dixon [16] makes this point in his research of conflict management principles based on different types of disputes. These principles range from public appeals, communication, observation, and physical intervention to pure mediation, humanitarian aid, and adjudication.

In brief, most of the literature sees intervention as a continuous, independent "spectrum of techniques" rather than as an integrated and interdependent set of

[4] A third-party coalition intervenes against a protagonist on behalf of either a group or a state center in order to suppress or support the internal ethnic challenge.

[5] At crisis onset, several traits distinguish the strategy of the third-party intervener. There are three possible strategies available: (a) forceful intervention, (b) efforts at mediation coupled with a low intensity conventional peacekeeping mission, and (c) withdrawal.

[6] The authors focus on mediation in peacekeeping missions and the role of dispute severity, time pressure, and the peacekeeper's rank.

strategies. In this view, different third-party techniques are set in motion at different points within a conflict. At one end of the "interventionist spectrum" is pure mediation, the facilitation of a negotiated settlement through persuasion, control of information and identification of alternatives by a party who is perceived to be impartial.[7] Key elements in pacific forms of third-party intervention, such as mediation, are the nature and level of consent, and the level of coercion required to reach a settlement.

Fourth, quantitative findings are not consistent and are often not supported in the context of single or multiple qualitative studies (also indicated in O'Neill's chapter). In this regard, two studies stand out in stark contrast to one another. Doyle and Sambanis' research on multilateral United Nations (UN) operations argues that UN missions have made a positive difference to the preponderance of civil wars since World War II. They examine 124 post–World War II civil wars, and in most cases multilateral UN peacekeeping operations made a positive difference [17]. Conversely, Regan and Abouhar [40] conducted a study to assess the effectiveness of third-party military or economic intervention in managing civil conflicts and curtailing their duration. Their findings revealed that, in general, most external amelioration efforts were unsuccessful in reducing the expected length of a domestic conflict, although interventions supporting one protagonist are associated with shorter conflicts compared to neutral interventions. Furthermore, it was noted that unilateral interventions tend to lengthen the expected duration of a conflict.

In summary, one of the most basic impediments to integrating formal modeling into policymaking is simply the absence of consensus regarding what the primary object of analysis should be. One of the advantages of formal modeling, the need for precise variable definitions, thus also limits its acceptability and applicability across contending and divergent research agendas.

3 Formal Modeling as a Research Program

Conflict and conflict management are inherently complex phenomena, and modeling them for both precision and nuance is difficult. For this reason, deriving policy implications from formal models is difficult. There are several reasons. First, policy recommendations that flow from formal studies can be inconsistent or contradictory, leaving the policymaker to choose alternative forms of analysis that provide a more consistent perspective. To be fair, however, some of these inconsistencies can be traced to differences in the evaluation technique and not to flaws in the methodology itself. For example, some studies are concerned with explaining intervention outcomes. Others focus on the relative effectiveness of different types of actors, while still other focus specifically on procedure (O'Neill in this publication).

Other problems reside within the logic of the model itself. In this vein, Bueno de Mesquita ([8], [9]) argues that a theory of conflict must first be deduced and must

[7] The term impartiality implies that the third party is acting in the interests of all of the parties. Where pure mediation may imply the absence of bias toward the interests of the parties, it does not mean that the mediator is neutral or indifferent to the outcome. Presumably, mediators have an interest in seeing that a violent conflict ends as quickly as possible.

be logically consistent internally. Deduction begins with value-based assumptions about what are the important areas to study (O'Neill). Generally, but not always, this occurs through consensus among researchers working within a common paradigm. This approach is consistent with "sophisticated methodological falsification" used to test propositions of deductively derived theories [27]. The internal validity of a theory resides in whether or not its conclusions can be arrived at without faulty logic and whether the properties of the model, elucidated in a case, are clear. If a deduction follows logically from a set of assumptions, then that deduction is necessarily valid under the precise conditions assumed in the theory. Falsification requires more than observation; it requires a clear analytical critique of the logic and concepts used in the model. However, it takes an empirical investigation to establish the external validity and the real relevance of a deduction.

Second, there is a need to separate out *explaining* the process of strategic interactions from *understanding* the decisions that policymakers face at any specific point in time. Kilgour, O'Neill, Druckman, and Güner (all in this publication) highlight this fundamental point. For example, Kilgour's taxonomy highlights several ways in which game theory can be used as a decision support tool. His taxonomy also indicates that there are several ways in which modeling can contribute to understanding the negotiation process and negotiator choices. Along these lines we are proposing that the modeling and explanation dimensions be refined before developing an approach that would help guide policymakers on specific choices that they might have in a given context.

Third, there is a need for accessibility. In our review of the formal modeling literature, including some of the chapters in this publication, there are few efforts to render formal modeling accessible to policymakers. It should be noted, however, that our assessment of formal approaches is not exhaustive but indicative. Our appraisal covers only third-party techniques and issues related to third-party intervention, including mediation and negotiation. For a more discursive analysis of conflict models, O'Neill's chapter is a good starting point.

To address these three concerns we consider the efforts that have been made within formal modeling to confront key objectives within the broader research program of conflict analysis: *accumulation: building on previous findings and modifying or discarding arguments for which empirical support is lacking; integration: drawing on alternative methodologies that provide similar findings in a different context and synthesis: using a multilevel of analysis approach.*[8] The underlying assumption here is that if formal modeling is able to identify fundamental generalities about the nature of strategic interactions through an accumulation of supported results to provide consistency in findings in comparison with other methodologies examining the same phenomenon, and if it is to apply these findings to a range of levels of analysis to both individuals and groups, then a stronger case can be made for pursuing the study of conflict through formal modeling techniques and, more importantly, their

[8] The concept of integration is covered more thoroughly in Druckman [19] and Brecher [6].

application in the policy realm.[9] If, alternatively, the evidence does not suggest that accumulation, integration, and synthesis are improving, then policymakers are unlikely to be supportive by using formal models in their work, and efforts to render them accessible are likely to fall on deaf ears.

3.1 Accumulation

Turning now to the first category of accumulation. There is a higher degree of success in this area than in integration and synthesis with respect to three key areas: intervention as a bargaining strategy, rationality, and bias.[10] Each assumption is considered in turn.

Intervention as a dynamic process

First, there appears to be agreement that intervention is a kind of dynamic bargaining strategy in which the characteristics of the intervener and its choice of strategy are treated endogenously.[11] The decision by a third party to intervene imposes costs on the conflicting parties as well as on themselves and, depending on the conditions, third parties do have an incentive to escalate. As Schelling [44] points out, *escalation* is the coercive side of negotiating a peace plan in which the fear of even greater cost imposition motivates actors to make concessions at the bargaining table.[12]

[9] O'Neill's chapter provides a good assessment of how formal models have been applied across levels of analysis and to different phenomena. He is encouraged by a higher degree of integration and synthesis today than there was 10 years ago.

[10] Some of these elements—for example, bargaining, are covered in O'Neill although his chapter covers a much broader range of convergence—for example, domestic politics, etc., which we have not included.

[11] Static decision making assumes either that behavior does not change because it is in equilibrium or that the decision process can be collapsed into one grand decision. For example, the expected utility theory of Bueno De Mesquita [8] treats nations as a unitary rational actor; that is, a utility function can be constructed that is consistent with a nation's foreign policy acts, such as alliance formation and wars, where it is assumed that political leadership of a nation will share a common basis of values. Existing rational models, like The War Trap (1985) model are static in the sense that they do not consider the possible responses to each possible move. Part of the reason for this assumption of the static approach is the belief that the choice of strategy follows from an actor's evaluation of the possible strategies. This approach cannot explain how, for example, crises develop over time or how to manage them over time in order to reduce the probability of war.

[12] Many of the earliest attempts to address the question of escalation between states placed the bargaining process at center stage. According to Schelling, one of the potential effects of escalation is to convince an opponent to back down by exploiting his fear that future escalation will lead to disaster. According to Harvey [25] and Powell [36], since both actors are engaged in demonstrating their superior ability to tolerate these risks, escalation is conceptualized as a game of competitive risk taking. Although others have recognized that the rate at which states escalate (impose costs) can have an important effect on the bargaining process [34], none has specified, theoretically, the conditions under which third-party actors select different levels of escalation to achieve their ends.

Schelling was the first to note that deterrence situations are akin to non-zero-sum games such as Prisoner's Dilemma Game (PDG) or the Chicken Dilemma Game (CDG). Unfortunately, in these games the least-best outcomes may arise as a consequence of either the pure Nash equilibrium strategies (PDG) or lack of coordination in games where there are multiple Nash equilibria (such as CDG), or a mixed strategy equilibrium. The more these games are repeated in an uncoordinated and iterative setting, the higher the probability of a disastrous outcome eventually occurring. Schelling's first contribution was a reorientation of game theory in a manner that would make the games more realistic and potentially more policy-relevant by allowing for the introduction of elements of commitment and resolve in strategic interaction. His primary aim was to highlight one of the key problems of nuclear deterrence: the tradeoff between the magnitude of the threat involving global annihilation and its credibility. However there is clearly a use for his approach in understanding and explaining third-party intervention strategies.

For example, Brams and Kilgour's early research [5] presents a model of crisis bargaining based on CDG. Their purpose is to isolate the optimal threat to cope with observed or potential provocation in a two-person conflict game. They ask under what conditions the threatened level of retaliation should be less than proportionate. The answer depends on the payoff structure; if the purpose is to deter aggression at potentially minimum costs, one must tailor the threatened punishment to the level of provocation or aggression—punishment to fit the crime. The sequence is such that deterrence is demonstrated by a single retaliatory countermove. The precise responses and game outcomes, of course, depend crucially on the assumptions that define the payoff structure and the information sets available to different players. These features are critical in all deterrence and escalation games (see below).

Like Schelling's, the Brams and Kilgour study has relevance to intervention theory, especially its bargaining and commitment components. For example, Carment and Rowlands [13] construct an intervention game with full information in which a dominant combatant and the intervener play sequentially. As in Brams and Kilgour, the payoff matrix ultimately determines the extent to which either side is prepared to escalate in order to acquire benefits at the margin, while full information implies that both sides know how far their opponent is prepared to go to achieve or avoid certain outcomes. Powell's study of war [37] is certainly helpful in understanding why. He shows that recent formal work on conflict management issues draws very heavily on Rubinstein's [42] seminal analysis of the bargaining problem and the research that flowed from it. More importantly he suggests that there is now what might be called a standard or canonical model in which war arises as a consequence of a bargaining failure. Powell's essay reviews the standard model and current efforts to extend it to the areas of (a) multilateral bargaining, which is at the heart of old issues such as balancing and bandwagoning as well as newer ones such as the role of third-party mediation; (b) the effects of domestic politics on international outcomes; (c) efforts to explicitly model intrawar bargaining; and (d) dynamic commitment problems. What is striking is the degree to which Powell is able to claim and demonstrate accumulation within the modeling of conflict, with certain model structures and underlying intuitions being commonly repeated.

Similarly, Cetinyan's [15] formal model of intervention and ethnic grievances is indicative. He shows that relatively weak ethnic groups mobilize and rebel against their governments just as frequently (or infrequently) as strong ones. However, such seemingly irrational behavior is consistent with a rationalist approach to ethnic separatism. A bargaining model that treats all the relevant actors as strategic players suggests that power disparities between an ethnic minority and the state—including those based on a group's access to third-party intervention—should affect how the state treats the group but not the likelihood that the group rebels against the state. Greater mistreatment by the state should not be correlated with greater external intervention on a group's behalf. New empirical support for the model is drawn from the Minorities at Risk data set, and the discussion has implications for the field of international relations beyond ethnic conflict.

Rationality

Rationality is a second assumption where there is a broad consensus. A good starting point here is Fearon [20], who provides a very narrow but rigorous attempt to identify a typology of rational conflict. Fearon focuses exclusively on rational conflict in its strictest, almost hyperrationalist, sense. The focus on rationality is central to formal modeling and clearly mimics its use in more traditional economic frameworks. The reason for this focus is simple: formal models need to be able to identify fairly general behavioral rules, typically derived from compelling first principles, that will be able to specify precisely how agents will act in a wide variety of different situations. Rationality provides a compelling and yet relatively simple way of identifying such behavioral rules. Clearly, it would be far more difficult, though perhaps not impossible, to build formal models on the basis of inconsistent or irrational behavior. Nonetheless, it is useful to state very clearly that for the most part formal modeling means an explicit or implicit acceptance of the assumption of rational behavior in a general sense.[13]

The other useful element that emerges from Fearon [20] is the assertion that there are only two purely rational explanations of conflict. The first explanation is one of private information with an incentive to misrepresent. Effectively, it is a situation in which at least one side in a conflict does not know the willingness or capacity of its opponent to engage in war. In order to deter aggression the opponent has a clear incentive to exaggerate its ability or willingness to fight. Therefore, any signals will be ignored or discounted, leading to the possibility of "collectively irrational" assessments of combat willingness and ability, that is, the sum of the players' own expected gains from fighting exceed that which would be available for sharing. The only credible signal, then, becomes the actual fighting itself. While Fearon provides

[13] This assumption is questioned by Wierzbicki's chapter in this publication. It is still a question of debate as to whether behavioralist approaches can be reconciled with rationalist approaches by using notions such as bounded rationality, prospect theory, and other mechanisms that in some ways dilute the notion of pure rationality. Empirically consistent behavior, however, can certainly serve as a foundation for formal modeling even if it violates, or appears to violate, principles of rationality.

only a heuristic discussion of this explanation, plus some empirical examples, other authors do present formal models of this process. Some deterrence models (e.g., [50], [30]) and models of escalation [11] use extensive form games to illustrate how a player may probe an opponent through various stages of combat until sufficient information is revealed for a resolution of the conflict to occur on mutually acceptable terms. Brito and Intriligator [7] present a more complex representation of a similar problem of uncertainty (asymmetric information in this case), where the first stage involves selecting a strategy for arming (aggression or deterrence) and a second stage involves the possible use of a challenge that could lead to war. The result is that if a country does not know its opponent's true propensity for fighting, its optimum strategy may be to react to challenges in a probabilistic manner—sometimes acquiescing and sometimes resisting—in order to deter bluffing by a weaker opponent. As a consequence of this mixed strategy, war may occur in an otherwise rational framework. Note that there are important differences about how time is used in these models to allow for the signaling and absorption of information, and whether or not war is modeled as a binary process of war or no war [7], or as a continuum of different levels of combat intensity or type ([30], [11]).

The second rationalist explanation of war in Fearon [20] deals with problems of commitment. Two more formal approaches are provided in this case: preemptive war with offensive advantages and preventive war as a commitment problem. The former is the traditional "gunslinger" problem that has simple interpretations in a PDG framework. As in Brito and Intriligator [7], you are either at war or you are not or, more prosaically, in gunfighter's parlance, you are either quick or you are dead. There is no temporal point at which one side can back down once all information for calculating the final outcome is known. As Schelling [44] and others have noted, this type of model seems very applicable to nuclear confrontation. Stability and prevention in this sort of situation comes either from confidence building and commitment mechanisms or from eliminating the first-mover advantage by the presence of a credible retaliatory strike.

The question of conflict arising as a consequence of military technology is also explored in models such as those of Anderton [2]. This type of model approaches conflict from the bottom up, using battlefield-level equations of force attrition (Lanchester equations) as the foundation for determining combat advantage. War occurs when one side can unambiguously win in an offensive action as a consequence of its strength and the technology of combat (advantages to defense or offense) at that time. While a useful addition to our understanding, the model is not intended to explain the emergence of conflict from a holistic perspective. Specifically, the model violates at least two of Fearon's rationality strictures. First, since in the full-information version of the model the outcome is predetermined and known to both players, why does the eventual loser not capitulate or bargain to avoid war? Second, the simplistic interpretation of the model would suggest that a country goes to war because it can win in combat, without paying attention to the costs and gains of war.

The second of Fearon's formal presentations of the commitment problem is more interesting in its use of a dynamic multiperiod framework. In this model the most powerful country initiates war in the first period in order to prevent an up-and-coming rival from dictating less preferred outcomes in future periods. This story is a modification of the declining hegemon argument for war, where a hegemon in its sunset lashes out at rivals in a doomed attempt to maintain its status. Fearon provides the important insight from rationality that conflict emerges because the challenger cannot credibly precommit to not challenging the current hegemon in the future. While the outcome of current fighting is unknown to both players and may lead to the initiator's defeat, Fearon shows that the initiator's expected outcome for fighting now can exceed the certain bad outcome to which it would have to acquiesce in the future. The rising challenger faces a classic problem of time inconsistency; it would like to be able to assuage its rival's fears by precommitting to a nonaggressive policy in the future but has no credible mechanism for such a commitment; hence, Fearon's emphasis on the importance of commitment.

Fearon manages to provide two rational explanations of war using the strictest definition of rationality. Many of the less formal approaches, while plausible, fail to come to terms with the inherent irrationality of war, namely, that war is destructive and hence Pareto-inferior. There is always a preferred outcome to war; it is just that the two (or more) sides cannot always get to it or abide by the required terms. However, Fearon's typology is, in fact, incomplete. Garfinkel and Skaperdas [22], for example, provide a formal model that is arguably distinct from Fearon's models, though it is closely related to the preventive war with commitment problems. Using a two-period model, Garfinkel and Skaperdas portray two countries that must divert resources toward their military establishments in order to secure their share of resources. War becomes preferable for both sides and Pareto-superior in expected utility terms because victory by either side settles the security dilemma and eliminates (or significantly diminishes) the need for subsequent investment in wasteful military preparation. The gain from war now is the reduction in military spending in the future. This approach can be thought of as cashing in on an extreme "peace dividend," though recent history has taught us that such dividends are often illusory or short-lived. Interestingly, Garfinkel and Skaperdas note that the shadow of the future is often thought to improve the chances of cooperative agreements (at least, in iterative PDGs) by facilitating future punishment for current transgressions. By contrast, in their model the future contributes to the emergence of war by allowing the winner the opportunity to enjoy the fruits of victory.

Other formal models effectively accept conflict as an inherent and unavoidable element of human affairs, implicitly accepting that commitment problems prevent more peaceful means of settling distributional disputes and seeking instead to explain only the intensity of conflict. For instance, models by Hirshleifer [26] are prime examples of this approach to formal conflict analysis. While Hirshleifer is careful to build on rationalist foundations of utility maximization, there is no explicit consideration of why opponents are unable to overcome coordination problems to reach Pareto-superior outcomes. Associated models of intervention that build on Hirshleifer's approach ([41], [14]) are thus able to examine the details of marginal reactions to intervener behavior but cannot easily deal with the corner solutions of complete peace or complete victory.

Other modeling approaches offer reasonable portraits of conflict that dilute to varying degrees Fearon's rationality constraints. For example, Bueno de Mesquita [9] provides a complex model of bilateral country relationships with probabilities associated with the outcomes of certain policy choices regarding confrontation, alliance, and other key foreign policy choices. Where conflict ultimately seems to emerge in a risk-neutral world, however, is in the meta-level inconsistencies between expected outcome assessments. While restrictions impose individual rationality and consistency in these calculations, there does not appear to be any need for different players to have mutually consistent assessments. This asymmetric information problem may be quite reasonable, and, as shown above, Fearon and others have demonstrated why it may come about. As with Hirshleifer, however, the model is not explicit about why these informational problems are not communicated away. A second cause of war easily demonstrated in Bueno de Mesquita's model [8] is risk-loving behavior by one or more actors, though this assumption about risk preferences is one that Fearon [20] explicitly avoids.

Moving even further from Fearon are models such as those of Gurr [24] where traditional rationalist assumptions are replaced by those arising from behavioral psychology. Explicitly, Gurr models people as reacting aggressively to perceived injustices, specifically the presence of various types of relative deprivation. Behavioralist approaches have strong empirical resonance, cannot be dismissed out of hand, and can indeed be modeled formally (see Gabbay, this publication). But Gurr is explicitly focusing on rebellion, and thus one might legitimately question the extent to which these impulsive reactions can sustain longer-term combat, especially in cases where violence yields no apparent reward or has any effect on the original stimulus. In addition there are (surmountable) problems of moving from the individual to the aggregate level in such models.

The effects of bias

A third area where there is emerging consensus focuses on the importance of bias and moral hazard. Consider Kydd's [31] assessment of biased mediators. Consistent with work by Carment and Rowlands [13], Kydd argues that mediators are often thought to be more effective if they are unbiased or have no preferences over the issue in dispute. His article presents a game-theoretic model of mediation drawing on the theory of "cheap talk" that highlights a contrary logic. Conflict arises in bargaining games because of uncertainty about the resolve of the parties. A mediator can reduce the likelihood of conflict by providing information on this score. For a mediator to be effective, however, the parties must believe that the mediator is telling the truth, especially if the mediator counsels one side to make a concession because their opponent has high resolve and will fight. An unbiased mediator who is simply interested in minimizing the probability of conflict will have a strong incentive to make such statements even if they are not true; hence the parties will not find the mediator credible. Only mediators who are effectively "on your side" will be believed if they counsel restraint. The intuition behind Kydd's result is simple and persuasive, and effectively the mediator is acting to replace steps in a game of escalation under imperfect information. As in traditional public goods problems in economics, Pareto-inefficient solutions emerge because of the difficulties of preference revelation, the basis of Fearon's first type of "rational war." A biased mediator can credibly solve the problem of incentives to misrepresent, at least in terms of one player. On the other hand, a more straightforward analysis of bias ([46], [53]) shows that biased mediators are often necessary to produce an agreement, provided they can deliver the side toward which they are biased, based on their own interest in an agreement that ends a conflict that is costly to them.

Using a theory of mediation and peacekeeping Smith and Stam [45] point to the sources of recent successes in the Middle East and the reasons for the more general pattern of failed third-party mediation. They contend that deductive theorizing is a superior means of advancing rigorous explanations of conflict management. In this light, third-party intervention and mediation are explored in the context of a random walk model of warfare and war termination [45]. In considering how third parties can hasten the end of conflict, it is shown that while mediators can use side payments or threats to intervene directly, unlike in Kydd they cannot help nations resolve informational differences. The model's equilibria demonstrate that conflict continues until beliefs converge sufficiently for both sides to agree that the costs of fighting exceed any likely gains and return to ripeness theory. Thus, at issue is whether the mediator can end such wars by speeding up the convergence via nonviolent presentation of information.[14] It is concluded that deductive reasoning allows for the parsing out of those mechanisms through which third parties influence conflict.

It could be argued, therefore, that accumulation in the formal models of intervention is approaching, but has not yet achieved critical mass, as demonstrated by three elements apparent in the previous discussion. First, common concepts such as rationality and attendant methodologies such as game theory have provided a sufficiently robust foundation to support a growing volume of papers that are mutually coherent in the sense of using recognizably similar assumptions, methods, and definitions. Second, this mutual coherence means that different contributions to the literature can be positioned relative to one another so that differences between them may be traced to specific variations in their underlying assumptions. Within this matrix of linked but distinct contributions it becomes possible to extend theoretical insights and close gaps simply by manipulating specific modeling elements. Third, and perhaps most importantly, the accumulated models have generated testable hypotheses that have begun to be examined empirically, and with reference to one another. While empirical testing remains fraught with all the attendant difficulties of poor data, ill-suited proxies, and problematic identification, at least there appear to be approaches and models worthy of refining and clear indications of remaining deficiencies. The relatively recent emergence of more comprehensive and nuanced data sets with wide

[14] A specific example is used as an illustration. Looking at the mediator's motivations, it is seen that the mediator must be a "genuine honest broker"—always providing honest information—for the belligerents to believe the mediator. The kinds of mechanisms that foster success for mediation are apparent in United States President Carter's mediation efforts between Israel and Egypt.

acceptance is symptomatic of the process of empirical refinement. These three elements have not matured to the point where the literature can transcend its associated academic community, but the potential for robust empirical verification will make such analyses increasingly useful, and probably essential, for policymakers.

3.2 Integration

Our second category is integration; namely, efforts to draw on findings from different methodologies and present them in one package. Within the literature we have reviewed thus far, findings on effectiveness and techniques appear to be quite diverse and inconclusive, and there appear to be only a few efforts at integration. O'Neill's review of the literature in this publication supports this view. Exceptions include Druckman's chapter and Druckman's recent work on integration [19]. Two other exceptions are notable; for example, O'Brien's [35] pattern classification using algorithm-fuzzy analysis of statistical evidence (FASE). O'Brien claims to be able to provide accurate forecasts not just on the occurrence but the intensity level of country-specific instabilities over five years with about 80 percent overall accuracy. His contribution is notable because he suggests that one way to demonstrate progress in a field of scientific inquiry is to show that factors believed to explain some phenomenon can also be used effectively to predict both its occurrence and its nonoccurrence. His study draws on the state-strength literature to identify relevant country macrostructure factors that can contribute to different kinds and levels of intensity of conflict and country instabilities.

Another exception is Wohlander's [49] very instructive research agenda based on the premise that a theory that produces a general, causal explanation of third-party intervention and specifies the precise conditions under which it does and does not occur, is both viable and policy-relevant. He shows that, overall, the theory he develops predicts approximately two-thirds of cases correctly when subjected to rigorous empirical tests.

His research develops a theory of intervention by laying out a story about how strategic third parties and disputants make interdependent decisions in the context of an ongoing militarized dispute and then formalizing this story into a simple game-theoretic model. In addition, the theory produces theoretically interesting, empirically supported insights about the relationships between the resources of the actors involved in a militarized dispute and the likelihood that intervention occurs. His research concludes with an application of the theory to the debate in the international relations literature over whether balancing or bandwagoning is the more common form of intervention. The application shows that the theory produces a more powerful explanation for the occurrence of balancing and bandwagoning than the existing literature offers, and suggests that the debate is misspecified.

Part of the problem in achieving full integration is that in contrast to our example of economics, formal conflict theorists do not yet have a sufficiently developed consensus about how to model conflict (not just intervention) or identify when one approach is more applicable to another (see O'Neill). This problem mirrors in some way the problems that are manifested in the intervention literature in general. For example, Rowlands and Carment [41] and Carment and Rowlands [14] model intervention formally by extending the approach of Hirshleifer [26], but they also draw on case studies in doing so (see [41], [14]). Whether the formal results and policy implications transcend this specific type of application remains an open research question.

In the absence of a comprehensive set of conflict (and thus conflict intervention) models, the only alternative is to use formal approaches to intervention that are sufficiently robust that they transcend any underlying conflict model. One possible candidate is deterrence theory, the formal analysis of which has wide (though not universal) acceptance; this has sufficient rigor in structure to be generalizable and sufficient flexibility in interpretation to be tailored for specific application.[15] More importantly, deterrence theory has been broadly applied using a variety of different methodologies, inductively and deductively, using assumptions of rationality and nonrationality. Rational deterrence already has a proven ability to permeate government institutions, having been the foundation of Cold War security policy. It also has empirical content; Carment and Harvey [12] have used deterrence theory to examine intervention outcomes in Bosnia and Kosovo. By accepting conflict as a "given," rational deterrence approaches to intervention can avoid the problem of specifying the nature or initial causes of the war and focus on how to deter continued undesirable behavior.

While promising, certain caveats need to be acknowledged before embracing rational deterrence as the only or best approach to intervention analysis. First, failures of deterrence-based intervention have been frequent despite its apparent acceptance within the policy community. Whether these represent teething problems in recalibrating the theory to fit intrastate conflict conditions or more fundamental defects that preclude its universal application is difficult to say. Certainly, it is plausible that deterrence would be more difficult when dealing with irregular forces lacking a clear political or military hierarchy and operating outside the control of a clearly recognizable political structure. Second, it is not apparent that rational deterrence is the most efficient basis for organizing intervention, especially if the "cause" of the war is informational asymmetry. Third, even if it is the most efficient approach, past practice suggests that interveners may not have the inclination to apply sufficient effort to make deterrence effective. Finally, while deterrence theory may provide a short cut to modeling conflict intervention, it still requires an understanding of what motivates the different combatants in a conflict, which brings us back to the initial and fundamental problem regarding the nature of the conflict.

3.3 Synthesis

Turning now to our third category of *synthesis* or the inclusion of findings at different levels of analysis, it could be argued that intervention and mediation theory

[15] See O'Neill in this publication. Achen and Snidal [1] and Lebow and Stein [32] present, respectively, the pro and con sides of the debate over rational deterrence theory. While raising many useful points, Lebow and Stein demonstrate in their argument an incomplete understanding of the theory.

should lend itself to synthesis, especially if there are efforts to draw on the deterrence literature because so much has been written on the topic from a variety of methodological perspectives. Here too, there appear to be few efforts to take findings from one level of analysis (e.g., the individual or group) and apply it at higher levels of aggregation. An exception to this would be mediation and negotiation research wherein insights and research on small group interactions—specifically questions of bias and impartiality—lend themselves to questions of third-party effectiveness at the state level [14].

There is, of course, a limitation to synthesis, as rationalist explanations assume interveners are capable of making decisions on a conflict according to coherent, well-ordered preferences. For formal modeling purposes, interveners and belligerents are generally treated as rational actors for analytical convenience despite the presence of complex coalitions and domestic political economy processes. Even if we accept the abstraction of third parties and belligerents as rational unitary actors, we still must satisfactorily specify the objectives of their decision makers.

Arrow's theorem suggests that although states unified under a multilateral coalition may act as if they are unitary decision makers, they may also act incoherently in the sense of not revealing a complete set of transitive preferences. It may be impossible to argue that any collection of persons or states is acting as if they were pursuing an identifiable goal. Bueno de Mesquita [9] has suggested that we cannot truly understand any international behavior or process unless we specify the role of decision makers in the process. The value of any theory is not just that it solves problems but that it offers new ways of conceptualizing. Rationality is a normal human trait that tends to relate ends and means and a way of measuring these in a proportionate manner. The difficulty is in estimating the values that policymakers assign to particular goals or objectives and in estimating their willingness to bear the potential risks and costs of a particular action.

Maoz [34] offers some valuable lessons on synthesis, as well as on accumulation and integration. Maoz first develops a game-theoretic model with modified versions of (a) conflict initiation, (b) conflict management, and (c) negotiation as viewed first from the perspective of a single actor and then from the perspective of both actors. This kind of approach crosscuts levels of analysis and draws on findings from disparate research on management and conflict analysis. Maoz uses this model to address three questions: (a) What is the relationship between the preferences of individual decision makers and aggregate outcomes which are individual decision makers and groups observed at the international level? (b) What is the relationship between choice and consequence in determining and assessing foreign policy outcomes? (c) Is there a link between micro and macro decision making and choices as evolutionary patterns develop over time? In responding to these questions, Maoz argues that micro and macro decision-making behavior cannot be treated as discrete and independent variables if one wants to explain change in outcomes over time. Thus, he sets about attempting to synthesize micro and macro models in order to explain changes over time. In bringing the individual and group approaches together, Maoz argues that there are three factors that shape decision making: situational variables, personality traits, and organizational roles.[16] Through a formal mathematical approach and critical testing (based on three criteria—simplicity, fertility, beauty), Maoz arrives at a model that is more than and different from the sum of its components. According to Maoz, the model leads to propositions that are both surprising (which he values) and theoretically testable, while being capable of explaining situations where good results would not be expected. The model is capable of explaining both ad hoc and strategic interactions. For example, the strategic approach places an emphasis on long-range goals in a stable international environment, while the ad hoc approach suggests that consistency is the exception rather than the rule. While this complex model still requires additional testing and corroboration, it points to the potential for models to be rich enough to handle a variety of detailed phenomena at the micro level while providing some confidence that important macro relationships have been taken into account.

In summarizing this section we offer three suggestions for improving accumulation, integration, and synthesis. First, with respect to accumulation, a key goal would be to build on *bargaining models* in order to create a dynamic theory that would emphasize the interplay between interveners and belligerents using bargaining strategies. This approach would capture the essence of decisions that each side faces. Unlike traditional intervener perspectives, which assume that decisions are dictated by the choice of a grand strategy at the beginning, the rationales for individual decisions within the conflict would be examined and optimal moves could be determined at each turning point. The second step is to model each actor's evaluation of its future benefits from its actions. Unlike static models, this approach assumes that actors evaluate the streams of benefits and costs that flow from a decision rather than just the immediate gain. A third step would be to model decisions as time-dependent calculations. This approach assumes that the decision to be made is when to act rather than just how to act.

Finally, it is important to understand the sequence of choices made by both interveners and belligerents. Brams and Hessel [4] postulate that sequential games are a useful basis from which to analyze a player's choices that are nonmyopic (farsighted choices that anticipate the entire sequence of moves and countermoves that players will make) and in equilibria. Their analysis of the threat strategies of states suggests that both players (the threatened and threatener) will play the same game over and over again with the threatener being the player who can hold out longer at a mutually disadvantageous outcome.

[16] While Maoz wants to synthesize these two distinct types of decision making, he is also concerned to bring both these processes (which can be considered as one-way processes) into contact and synthesis with interactive processes as derived largely from game theory. Much of his work in his book is devoted to testing all models in order to demonstrate their weaknesses when standing alone and their strength in being synthesized.

4 Implications for Policy

> There are two major reasons why policymakers pay greater attention to case studies than empirical models. First they are generally older, having completed their primary education well before the behavioral revolution and the government doesn't provide much incentive to stay current in your field. Case studies are generally easy to understand and appreciate regardless of your educational level or methodological training. Second, the tension between qualitative and quantitative analysis in the government is in times and places much more acrimonious than it is in many political science departments, but this applies more at the level of the government analyst than the policymaker.... "I don't think there's anything unique about international behavior that makes it less predictable than, say, economic behavior," says Philip A. Schrodt, a professor of political science at the University of Kansas. "If anything, an economic system is far more complicated than an international system. And yet we just constantly engage in economic forecasting." [23].

Any model of mediation and intervention that expects to be policy-relevant must do three things. First, it must specify which elements of intervention are the most effective in order to assist policymakers in designing more effective policies. In order for a theory to be politically useful it must have a solid body of empirical evidence to back its propositions. We have argued that efforts to provide empirical support for formal models are still in the nascent stages but are improving.

Second, it must be able to help decision makers think through or analyze problems in a manner that is superior to that which they would have used without it. In this case, intervention and mediation theory serves as a set of analytical tools, and policy relevance stems directly from observing the behavior of interveners and belligerents each with its own logic and behavioral properties. In this respect there is room for optimism.[17]

Third, mediation and intervention modeling must be able to help identify systematic deviations from optimal decision making and the identification of certain correcting principles.[18]

In each of these areas there has been some progress.[19] Political science and econometric ideas permeate Washington and other Western nation centers of decision making. The Pentagon, State Department, and Central Intelligence Agency are

[17] For example, many of the chapters in this publication, including those by Kilgour, Druckman, Güner, and Brams *et al.* strive to achieve these objectives.

[18] See, for example, Brams *et al.* in this publication. The authors identify an approach that encourages negotiators to engage in compromise and avoid strategizing.

[19] For example, military planners have long relied on formal numerical models of conflict at the battlefield level, and computer packages such as the Dupuy Institute's Tactical Numerical Deterministic Model have gained some notoriety both as a means of simulating past or potential conflicts and as a demonstration of how careful empirical detail and constant recalibration can improve the accuracy of formal mathematical models.

filled with people who understand political science theories and the findings derived from them. Some use these theories inherently to evaluate proposals and model outcomes, though they do not use them explicitly nor do they rely solely on them for devising policy options. Theories and models without clear policy implications or actionable forecasts are just not useful to them on a day-to-day basis.

The connection between formal modeling and policy is not a simple one. It is useful first to consider analogous situations where models have come to underpin policy analysis and formulation. Good examples come from economics, the social science discipline where formal modeling is the most widely accepted. Policy discussions on economic matters are routinely informed by the analysis derived from formal models. However, four observations are relevant in terms of the lessons that may prove useful for conflict resolution modeling and policymaking.

First, the degree of consensus within the discipline of economics regarding the basic models and their assumptions is overwhelming. Second, the models that tend to underpin most economic policy discussions have generally been tested and refined through empirical investigation. Hence the influence of formal modeling is often indirect and filtered through a more complete "scientific" analytical structure that includes some degree of empirical verification. While differences of opinion are common, these tend to be over either relatively minor implications of the models, or their empirical foundations. Consequently, it is often relatively easy to isolate the points of contention and identify empirical approaches to settling the dispute.

Third, the substantial consensus in the discipline is perpetuated through the essentially uniform courses taken by economists in their university training. Finally, most government economic policymakers who consume formal analyses are themselves trained in the discipline. Even if they cannot themselves produce the formal analyses, they will possess sufficient familiarity with the assumptions, techniques, concepts, and terminology, not to mention disciplinary biases, to make the models accessible and more compelling.

While this portrait is no doubt idealized, it is arguable that all of these elements are present only in a much weaker form, if at all, when it comes to formal models of conflict analysis. Each of these points is examined in turn.

First, formal modeling of conflict intervention cannot easily be translated into policy terms because of the aforementioned barrier caused by the lack of a developed consensus. If there were stronger agreement on how to model conflict, or how to model intervention, policymakers would have greater confidence in the consequent policy recommendations. The natural suspicion that should arise when one is confronted with a novel and unconfirmed theory is reinforced by the fact that modeling economics and modeling political behavior are inherently different. In contrast to Glenn's quote above [23], we might suspect that fewer actors and a simpler (less-regulated) system will lead to more challenging and less-predictable behavior. While the law of large numbers may allow us to ignore aberrant individual behavior in a disaggregated market context, no such convenience exists in international relations. The presence of fewer actors encourages strategic behavior where actions can be changed abruptly and dramatically in response to the choices of others. Markets are typically constrained by well specified and enforced rules, something the international system decidedly is not.

Addressing the difficulties of when and how to obtain a specific outcome through further research will contribute to the efforts to reconstruct the principles that underpin successful third-party intervention. The absence of a clear consensus about the theories and models of conflict and intervention relates immediately to the absence of an empirical consensus. Indeed, the two are clearly related in a scientific sense: empirical investigation should be weeding out those models that fail the test of evidence. While considerable progress has been made on the empirical examination of both conflict and intervention, consensus remains elusive, and many of the points of contention and testing remain at the most fundamental levels of analysis.

Second, the research endeavor has not proceeded to the stage of refinement and qualification. Consequently, the margin for error for any associated policy suggestions remains daunting, with potential implications for thousands of lives and millions of dollars. There may be some understandable reluctance to rely on a single, formal model that may produce counterintuitive results when dealing with such emotive consequences. In simplest terms, formal models of intervention largely remain untried, untested, and potentially untrue.

The third barrier is simply one of the larger challenges for formal modeling arising from the lack of consensus. In economics, university courses are largely standardized, and formal modeling is pervasive. Indeed, economics seems to suffer from the opposite problem to the study of conflict, with common-sense discursive analysis being viewed with extreme skepticism or dismissed entirely in the absence of corroborating formal models. The battle between competing ideas, methodologies, and normative standards is far more intense in international relations than in economics. As a result there are conflicting schools of thought that are unwilling to yield ground to, or acknowledge the legitimacy of, their competitors. This absence of convergence is apparent in course structures, doctoral thesis expectations, and even journal refereeing. As long as there are large sections of the academic establishment that are incapable of understanding, let alone producing, formal models with mathematical representations, then there will be tremendous difficulty in forging a consensus on how they might be incorporated into policymaking.[20]

Finally, the breadth of training present within the policymaking community itself will generate the same sort of resistance to applying formal models of intervention. Even when policymakers and modelers are drawn from the same discipline, it is usually the case that the models have been tested empirically and translated into more accessible language prior to their emergence in any policy discussions. Disciplinary uniformity and complementarity undoubtedly expedite this process. First, policymakers can engage the theory and theory builders directly, providing direction in terms of how the model may need to be modified, refined, or repackaged in order to be useful in policymaking. Second, the extent of these modifications may be minimized by the presence of a common analytical and terminological framework for discussion. Third, the affinity of policymakers to modeling will be stronger if they

[20] See for example, Schüssler in this publication.

have formal training that is in common with the modeler. Finally, there will be a natural bureaucratic reluctance to adopt novel techniques when their adoption may lead to failure and where there is a good chance that an external evaluator will disagree with the theory underlying the technique. It is harder to blame a single bureaucrat for a policy failure when he or she is following the prescriptions that emerge from a model with widespread currency in the relevant policy and academic community. This desirable convergence is much less likely in conflict intervention than, say, fiscal policy.

5 Further Research

A predictive capacity based on dynamic theories of intervention and careful empirical work can provide policy-relevant forewarning with which interveners may be armed. This chapter has highlighted some of the theoretical and empirical challenges that quickly emerge in identifying the consequences of intervention strategies. Addressing such challenges is crucial, however, as current policy initiatives continue to race ahead of clear and precise strategic analysis.

As a next step, we would suggest greater incorporation of findings from different methodologies and greater efforts at synthesis. Current models need to be placed within a typology in order to identify areas of distinction and areas of similarity. Fearon [20] and Powell [37] begin this process. Once a typology has been created, classes of models can be developed and refined in a cumulative fashion while gaps can be identified by theoretical innovation and by the presence of empirical anomalies.

Current and future models then need to be subjected to more systematic empirical testing, starting with case studies (see [8], [9], [43], [28], [14]) for a justification of case studies in model building). For readers, especially policymakers, who may be unfamiliar with formal modeling, a case study can provide an accessible and practical way of acquiring the insights of a model and its insights into causality. Simultaneously, irregularities between a model and a case study, while insufficient to refute propositions, can identify potential directions for model modification or refinement. By the same token, consistency between the model and the cases does not provide irrefutable support for the model. For this we would need the confidence of larger sample studies. Druckman [18] demonstrates how case study and large N study approaches can be synthesized.[21]

The work of Maoz [34] illustrates another useful direction for research in terms of addressing all three of our overarching criteria: accumulation, integration, and synthesis. By explicitly marrying micro and macro levels of analysis a model can provide richness without sacrificing generality, and vice versa. Several questions are

[21] Four case-based research approaches to analysis of data on international negotiation are discussed: the single, analytical case study, the temporal or time-series case study, the focused comparison of a small number of similar cases, and aggregate comparisons of a large number of different cases.

raised by his study. First, what do we mean by rationality when we are looking at dynamic processes? Static assumptions of rationality may be fundamentally different from rationality involving two or more actors in a dynamic process. Second, as O'Neill has shown, dynamic models should include the element of uncertainty in order to better estimate reactions to anticipated behavior. Third, there may be general problems with rational choice models that do not incorporate some form of game-theoretical approach in describing dynamic processes of international conflict.

The question of uncertainty points to the last research direction that needs to be addressed before formal models can be incorporated into policy. Information plays an important role in game theory and in real life. Analyzing models for robustness, particularly with respect to variant assumptions on information, is critical for good policymaking. The recent furor over intelligence failures in both the 9/11 attacks and the 2003 invasion of Iraq demonstrate the centrality of information sets in determining behavior. Testing inferences for sensitivity to information both at the formal and empirical levels are critical, as it permits decision makers to examine risks with a better sense of probabilities and induces more appropriate levels of confidence in the models that are underpinning policy choices. Ultimately, policymakers will adopt formal models when there is confidence in them. This integration will not occur until the academic community has that same requisite confidence that is born of sufficient accumulation, integration, and synthesis that are the hallmark of good scientific research.

References

1. Achen C, Snidal S (1988) Rational deterrence theory and comparative case studies. World Politics 41(1):143–169
2. Anderton R (1992) Towards a mathematical theory of offensive/defensive balance. International Studies Quarterly 36(1):75–100
3. Bercovitch J (1996) Resolving international conflicts: the theory and practice of mediation. Lynne Rienner Publishers, Boulder, CO, USA
4. Brams SJ, Hessel MP (1984) Threat power in sequential games. International Studies Quarterly I:23–44
5. Brams SJ, Kilgour DM (1987) Winding down if preemption or escalation occurs: a game-theoretic analysis. Journal of Conflict Resolution 31(4):547–72
6. Brecher M (1993) Crises in world politics: theory and reality. Pergamon, New York, USA
7. Brito D, Intriligator M (1985), Conflict, war, and redistribution. American Political Science Review 79(4):943–957
8. Bueno de Mesquita B (1980) An expected utility theory of international conflict. American Political Science Review 74:917–931
9. Bueno de Mesquita B (1985) The war trap revisited: a revised expected utilities model. American Political Science Review 79(1):156–173
10. Campbell D, Mayer-Kress G (1997) Chaos and politics. In: Grebogi C, Yorke J (eds) The impact of chaos on science and society. 18–63, United Nations University Press, Tokyo, Japan
11. Carlson L (1995) A theory of escalation and international conflict. Journal of Conflict Resolution 39(3):511–534

12. Carment D, Harvey F (2000) Using force to prevent ethnic violence. Prager Press, Westpoint, CT, USA
13. Carment D, Rowlands D (1998) Three's company: evaluating third party intervention in intrastate conflict. Journal of Conflict Resolution 42(5):572–599
14. Carment D, Rowlands D (2001) The role of bias in third party intervention: theory and evidence. Belfer Center for Science and International Affairs, John F. Kennedy School of Government, Harvard University, Cambridge, MA, USA. Work in progress (2001–08), October
15. Cetinyan R (2002) Ethnic bargaining in the shadow of third-party intervention. International Organization 56(3):645–677
16. Dixon WJ (1996) Third-party techniques for preventing conflict escalation and promoting peaceful settlements. International Organization 50(4):653–681
17. Doyle M, Sambanis N (2000) International peacekeeping: a theoretical and quantitative analysis. American Political Science Review 94(4):779–801
18. Druckman D (1986) Four cases of conflict management: lessons learned. In: Bendahmane DB, McDonald JW (eds) Perspectives on negotiation. 263–288, Center for the Study of Foreign Affairs, Foreign Service Institute, Washington DC, USA
19. Druckman D (2003) Toward integrated knowledge. In: Cheldelin S, Druckman D, Fast L (eds) Chapter 18 in conflict: from analysis to intervention. Continuum, London, UK
20. Fearon J (1995) Rationalist explanations for war, International Organization 49(3):379–414
21. Fisher RJ (1996) Pacific, impartial third-party intervention in international conflict: a review and an analysis. In: Vasquez JA, Johnson JT, Jaffe S, Linda Stamato L (eds) Beyond confrontation: learning conflict resolution in the post-cold war era. 39–59, University of Michigan Press, Ann Arbor, MI, USA
22. Garfinkel M, Skaperdas S (2000) Conflict without misperceptions or incomplete information: how the future matters. Journal of Conflict Resolution 44(6):793–807
23. Glenn D (2002) Calculus of the battlefield: do game theory and number crunching – The new math of international relations – shed light on the conflict with Iraq? The Chronicle of Higher Education, 8 November. Available from http://chronicle.com/free/v49/i11/11a01401.htm See also interview with Jack Levy at http://chronicle.com/colloquylive/2002/11/war/
24. Gurr TR (1970) Why men rebel. Princeton University Presa, Princeton, NJ, USA
25. Harvey F (1998) Deterrence failure and ethnic conflict: the case of Bosnia. In: Carment D, James P (eds) Peace in the midst of wars: preventing and managing international ethnic conflicts. 230–264, University of South Carolina Press, Columbia, SC, USA
26. Hirshleifer J (2000) The macrotechnology of conflict. Journal of Conflict Resolution 44(6):773–792
27. James P (1993) Neorealism as a scientific research enterprise: toward elaborated structural realism. International Political Science Review 14:123–148
28. King G, Keohane RO, Verba S (1994) Designing social inquiry—Scientific inference in qualitative research. Princeton University Press, New Jersey, NJ, USA
29. Kleiboer M (1996) Understanding success and failure in international mediation. Journal of Conflict Resolution 40:360–389
30. Kraig M (1999) Nuclear deterrence in the developing world: a game theoretic treatment. Journal of Peace Research 36(2)141–167
31. Kydd A (2003) Which side are you on? Bias, credibility, and mediation. American Journal of Political Science 47(4):597–611
32. Lebow R, Stein JG (1988) Rational deterrence theory: I think, therefore I deter. World Politics 41(1)208–224

33. Licklider R (ed 1993) Stopping the killing. New York University Press, New York, USA
34. Maoz Z (1990) National choices and international processes. Cambridge University Press, Cambridge, UK
35. O'Brien SP (2002) Anticipating the good, the bad, and the ugly—An early warning approach to conflict and instability analysis. The Journal of Conflict Resolution 46(6):791–811
36. Powell R (1989) Nuclear deterrence and the strategy of limited retaliation. American Political Science Review 83(2):503–519
37. Powell R (2002) Bargaining theory and international conflict. Annual Review of Political Science 5:1–30
38. Pruitt D (2000) The tactics of third-party intervention. Orbis 44(2):245–54
39. Regan PM (1996) Third party interventions into intra-state conflicts: identifying the conditions of successful intervention attempts. Journal of Conflict Resolution 40(1):336–359
40. Regan P, Abouhar R (2002) Interventions and civil conflicts. World Affairs 165(1):42–54
41. Rowlands D, Carment D (1998) Moral hazard and conflict intervention. In: Wolfson M (ed) The Political Economy of War and Peace. 267–285, Kluwer Academic Press, Boston, MA, USA
42. Rubinstein A (1982) Perfect equilibrium in a bargaining model. Econometrica 50:97–109
43. Russett BM (1974) Power and community in world politics. W H. Freeman, New York, USA
44. Schelling TC (1960) Strategy of conflict. Harvard University Press, Cambridge, MA, USA
45. Smith A, Stam A (2003) Mediation and peacekeeping in a random walk model of civil and interstate war. International Studies Review 5(4):115–135
46. Touval S (1975) Biased intermediaries: historical and theoretical considerations. Jerusalem Journal of International Relations I 1:51–69.
47. Urry J (2003) Global complexity. Polity, London, UK
48. Wall JA, Druckman D (2003) Mediation in peacekeeping missions. The Journal of Conflict Resolution 47(5):693–705
49. Wohlander S (2003) A theory of third-party intervention in disputes in international politics. Dissertation Abstracts International, A: The Humanities and Social Sciences 62(7):2565-A.
50. Zagare F (1992) NATO, Rational escalation and flexible response. Journal of Peace Research 29(4):435–454
51. Zartman IW (1994) Two's company and more's a crowd. In: Zartman IW (ed) International Multilateral Negotiation, Jossey-Bass, San Francisco, CA, USA
52. Zartman IW (2005) Cowardly lions: missed opportunities to prevent state collapse and deadly conflict. Lynne Rienner, Boulder, CO, USA
53. Zartman IW, Touval S (2006) International mediation in the post cold-war era, In: Crocker C, Hampson F, Aall P (eds) Turbulent peace. 2nd edition, United States Institute for Peace (USIP) Press, Washington DC, USA

Rationality of Choice versus Rationality of Knowledge

Andrzej P. Wierzbicki

National Institute of Telecommunications, Warsaw, Poland

1 Game Decision Situations

Game decision situations are characterized by independent decision makers with diverging preferences making independent decisions. While these situations are apparently described by noncooperative game theory, practical applications of this theory are limited by several factors:

1. Generally, but particularly in strongly competitive situations, decision makers are not inclined to reveal their preferences, especially in the basic form of a utility or value function. This makes it extremely hard, and often impossible, to compute game equilibria that could be used in negotiations, for example, for estimating BATNA (best alternative to negotiated agreement) as a status quo point. Game equilibria can be computed only under extremely simplifying assumptions (e.g., that all players minimize the costs of environmental protection) and thus can serve only as rough approximations.
2. In more complex situations, characterized by nonlinear or multicriteria models, equilibria are essentially nonunique (a nonlinear equation usually has multiple solutions). The analysis of multiple equilibria, their stability, etc., is for various reasons[1] not very far advanced in game theory that concentrated on arguments for choosing a unique equilibrium as a rational outcome of a game. In addition, multiple equilibria often lead to *conflict escalation processes*. If one player chooses an equilibrium that seems rational to him, and another responds with a choice of a quite different equilibrium, the outcome might be a deep disequilibrium, harmful to both sides—and to the players' cooperative agreement. For the relation between escalation and negotiation, see Zartman and Faure [27].
3. Conflict escalation processes are often observed in practice (see below for an example related to telecommunications market liberalization). Thus, political scientists are right in arguing that analyzing conflict escalation might be more important than analyzing game equilibria.

[1] Also for ideological reasons: if there are multiple market equilibria, what is the value of the argument that it is best to leave the choice of the equilibrium to the market?

Telecommunications market liberalization is a subject of international negotiations; for example, the author participated in Poland's preaccession negotiations with the European Union concerning telecommunications services and infrastructure. The liberalization was aimed at introducing new operators to markets for telecommunications services. According to the classical game and market theory, this could be achieved by removing regulatory obstacles and opening up the market. However, specific features of telecommunications markets make such simplifying conclusions impractical.

There is a paradox in the objective of market liberalization resulting from the fact that telecommunications services usually involve connecting people, including customers of different operators. Interconnection agreements and contracts are necessary. If an incumbent operator does not agree on and adhere to such a contract, the opposite side—a new operator—would be unable to provide adequate service to his customers. Thus, simply opening a telecommunications service market to new operators does not, in practice, result in its liberalization. Incumbent operators use every legal maneuver available not to agree on interconnection contracts or to prolong negotiations, even if required by law to agree. New entrants begin legal battles to harm incumbent operators. The incumbent operators, even if forced to sign an interconnection agreement, use various technical tricks to hurt new entrants; they might publish interconnection requirements that are up to 2,000 pages in length. Although a cooperative agreement could theoretically increase gains for both sides—new customers usually attract an increased demand for services—such an equilibrium is not necessarily reached and a conflict escalation process is typical. Therefore, paradoxically, telecommunications market liberalization cannot be achieved by deregulation, but requires stronger regulatory measures to enforce interconnection agreements.

The above example shows how actual human behavior often differs from the assumed rational behavior of a decision maker. On the other hand, one can wonder whether "rational behavior" is used correctly in rational choice theory. There is nothing irrational in the behavior of the incumbent operator who just wants to maximize his short-term profits and also fights to maximize his market share while keeping short-term profits reasonably large at the same time. However, investigating such situations would lead to an analysis of multicriteria games. While such an analysis is possible (see Wierzbicki [22] for example) the equilibria of multicriteria games are essentially nonunique (because Pareto sets seldom consist of single points) and conflict escalation processes can easily arise.

When trying to apply formal models to analyze noncooperative game situations as a starting point for preparing negotiations, a specialist—even if well aware of diverse mathematical methods and possibilities—is forced to arrive at the following conclusions:

a) The paradigmatic insistence of rational choice and game theory on the normative or prescriptive character of its conclusions *implies its impracticability*. A typical senior decision maker is affronted when told that he or she behaved irrationally and responds with the accusation that the theory has no bearing on his or her decisions and, as a consequence, must be false.

b) The concept of noncooperative equilibrium can be practically used *only for very simplified models of game situations*. We can give a rough approximation of the status quo or BATNA by using formal game models. However, knowledge of basic concepts and of the limitations of game theory can be helpful when trying to understand actual negotiating situations.
c) The best results when preparing for negotiations can be obtained by constructing and using *simulation games* for this purpose. Such games, however, must be complex to be realistic and usually include a number of criteria simulated for every player. The computation of a unique equilibrium (for example, by assuming weighting coefficients stated for every player) is not only usually difficult but often worthless—since there might be many such equilibria. Much better results for assessing a status quo outcome are obtained by simply playing a repetitive or dynamic simulation game and discussing whether its outcomes approach an equilibrium, have a disequilibrium character, or even illustrate a conflict escalation process.
d) One is forced to conclude that noncooperative game theory, though useful as a language specifying various concepts and describing certain behavior, is not useful as an actual tool for preparing for negotiations. Historically, not enough emphasis has been given to a possibly *descriptive character or variants of game theory or to its applications in simulated gaming*.

2 Group Decision Situations

Group decision situations are characterized by independent decision makers that either decide or are forced (by the nature of decision situations, such as when subdividing a joint budget allocation) to make joint decisions. The resolution of such situations is the essence of negotiation processes. These situations are described by cooperative game theory, with its concepts of cooperative game equilibria (Nash [15], Raiffa [19], [20], the latter axiomatic version given by Kalai and Smorodinsky [9], etc.)[2] and of coalition games. However, this theory is for several reasons too simple to be really useful in supporting negotiation processes:

1) To be useful in negotiations the theory should put more emphasis on describing (or prescribing) equilibration processes—diverse mechanisms resulting in cooperative equilibria—than on describing (or prescribing) equilibria.

A qualitative, soft theory of negotiations in political science emphasizes procedural aspects of negotiations, describing them as a complex process. Many essential concepts concerning types of negotiation procedures have been developed, for example, the single-text procedure (see [7]). Not enough research on formal models was devoted to describing these procedures in mathematical model language (i.e.,

[2] In fact, this solution was suggested by Raiffa in several of his earlier publications. See, for example, Raiffa [19], later characterized axiomatically by Kalai and Smorodinsky [9]; thus we shall call it a Raiffa–Kalai–Smorodinsky cooperative equilibrium or solution.

analyzing them as equilibration processes), although this direction of research was indicated earlier by Raiffa [20].

2) In most practical group decision situations, agreements must be reached concerning multiple objectives or criteria without specifying utility or value functions for each participant. While one cooperative equilibrium concept (Raiffa–Kalai–Smorodinsky solution) can be generalized for multiple objectives, not enough emphasis in game theory has been given to this problem.

Model-based decision support developed a methodology to overcome this obstacle. A mathematical model of the corresponding decision situation was formed including the objectives or criteria of all the decision makers concerned, and an interactive multiobjective analysis (including vector optimization that results in Pareto optimality of proposed solutions) of the model is then performed. During such interactive analysis, the preferences of decision makers are specified, for example, in the form of their aspiration levels (or also reservation levels, generally reference levels) for their respective objectives. After a specification of such levels, a resulting version of the scalarized vector optimization problem is solved and corresponding solutions are presented to the decision makers for further discussion. This can be interpreted as an interactive process of seeking an acceptable cooperative equilibrium of a multicriteria game. In fact, the initial solution proposed for such interactive processes is a so-called neutral solution that is a generalization of Raiffa–Kalai–Smorodinsky cooperative equilibrium (which corresponds precisely to this equilibrium if there are two decision makers, each with a single objective, and the mathematical model is convex).

This method was applied successfully, for example, in the analysis of the location of waste-treatment plants in the Nitra river in Slovakia, as well as in other environmental problems (see [25]). Such applications show that decision makers do not, on the one hand, object to specifying reference points (aspiration and/or reservation levels) as vague and tentative descriptions of their preferences, while, on the other, they usually object to more precise forms of preference specification. Another conclusion was that it is better to use reservation levels as *soft constraints* (i.e., parameters) in an achievement scalarizing function than to use them as *hard constraints*. For example, demanding that all environmental purity standards be strictly satisfied as inequalities might lead to unacceptable levels of investment costs in waste-processing plants, if the river was originally highly polluted. Using the standards as reservation levels and soft constraints results in the possibility of analyzing solutions that come close to fulfilling environmental standards at much lower costs.

3) In summarizing such experiences it might be said that normative rational choice theory is not really useful in supporting negotiation processes because it is too normative, not sufficiently descriptive, and too motivated by simplistic economic applications with a single monetary objective; life, meanwhile, is much more complicated than the economy.

This does not mean that successful applications of hard mathematical modeling for supporting negotiation processes are not known. There are a number of them in the literature, some of which will be described in this publication. Internationally known, for example, is the use of the Regional Air Pollution INformation and Simulation (RAINS) model of the International Institute for Applied Systems Analysis (IIASA) to support multilateral environmental negotiations (see e.g., Amman *et al.*, [2]). However, it must be stressed that the RAINS model, though quite complex and having been developed for more than a decade to *reasonably describe* the specific environmental decision-making situation, is based on a simplifying assumption for the selection of possible equilibria. This assumption is that all participants in the negotiation process agree on a principle for minimizing the joint costs of environmental protection, with suitably defined constraints and side-payments (because the effects of environmental protection in any given country are usually more pronounced in neighboring countries than in the country itself). Such an assumption is, in fact, justified by a known method of negotiation strategies stemming from the political sciences: the most successful results in negotiations are often obtained by using *principled negotiations* (i.e., first negotiating and agreeing on principles, then discussing the details) (see [7]).

The example of the IIASA RAINS model (or models; there are several variants, improved over the years of applications) illustrates another, more general concept: that of a *virtual laboratory*. A virtual laboratory is a system of computerized models, summarizing and approximating knowledge in a given field, used for performing diverse computational experiments and equipped with diverse computational tools supporting these experiments. Each engineering field, for example, bridge construction, has today developed a large number of computerized models that describe the behavior of the objects it is studying; and it is less expensive and less time-consuming to undertake computational experiments with such models when preparing for real experiments than immediately starting real experiments. For this reason, virtual laboratories are actually broadly applied today, although the name is not popular; but looking at them as examples of a general principle is helpful because standardized tools can be developed for them and applied in a cross-disciplinary manner. For example, in order to make experimenting easier with the IIASA RAINS models, interactive methods of vector (multiple criteria) optimization based on reference points have been adapted and further developed (see [25]).

3 Team Decision Situations

Team decision situations are characterized by independent decision makers who form a team with essentially common objectives that are not characterized by independent preferences and that still need mechanisms for resolving differences of opinion. Paradoxically, hard mathematical theory—namely, voting theory—is most useful in this simplest of negotiating situations. Voting theory is also applicable to group decision making, but in team decision making (where the purpose is simply to agree on a course of action, not to aggregate individual preferences), it is not biased by the known paradoxes of this theory.

It is well known that most voting procedures—both widely used and/or axiomatically based—lead to various paradoxes. Arrow's impossibility theorem illustrates this side of axiomatically based voting procedures [3]. But as shown by Nurmi [17], there is a way to escape the danger of such paradoxes. It is better to select a voting procedure that does not necessarily satisfy the first axiom of Condorcet—than that an option winning in most pairwise comparisons should be selected by voting—but otherwise reasonably aggregates the preferences of voters. Such properties can be found in Borda's voting procedure, or in its generalization by Nurmi. The Nurmi generalization consists of giving each voter 100 points (percentage of support) to be distributed arbitrarily among all options voted upon. This is a generalization of the original Borda procedure, where among n options, the best gets $n - 1$ points, the second $n - 2$, etc., until the worst gets zero points.

Against this theoretical background it is appropriate to consider some practical examples. One concerns the ranking and selection of proposals for scientific research projects (grants) by a suitable evaluation and selection committee. This is a frequent application of team decision making; in theory all participants of the committee have common objectives—to choose the best projects and to use in the best way public money allocated for such grants. However, they might differ on their perception of which project is the best. Most of them vote honestly; some try to vote strategically and to influence the voting outcome by giving extreme votes. All voting procedures that depend on the summation of votes or points are sensitive to such strategic distortions of evaluations. However, there is one counting system that is not sensitive: counting not the sum or average but the median evaluation (discarding all extreme evaluations until one or two remain). In practice, the announcement that the counting system will be based on the median works excellently: the evaluators try to be more objective. This has been confirmed by using the median counting system in diverse applications. Unfortunately, this information is not widely known: there are several examples of difficulties with strategic distortions of votes by juries in international competitions, but these juries often simply do not know that they could use median counting in order to prevent strategic distortions.

Median counting is a positive example of using hard mathematics to get the best results. However, another example is negative. Consider an evaluation procedure to decide upon a tender. Polish law requires such an evaluation when using public money for purchasing equipment for sums in excess of €3,000. The evaluation can be either multiobjective or multicriteria, but the law (actually, its official interpretation) requires that weighting coefficients be determined (counted in percentage points of importance of the criteria) and published. It is well known in the theory of multiobjective evaluations that such procedures cannot work reasonably, as they imply that preference is given to extreme options.[3] This theoretical property was frequently observed in practice in Poland, and the organizers of tenders had to use

[3] This phenomenon is sometimes called the Korhonen paradox (private communication) that can be formulated as follows. Suppose you want to choose a partner for life, evaluated by two criteria: sex appeal and intelligence, using the weighted-sum procedure. Suppose one option (a possible partner) is evaluated 10 points for sex appeal, 0 for intelligence, second option 0 for sex appeal, 10 points for intelligence, third option 4.5 points for sex appeal and

additional constraints to eliminate extreme options. However, the obvious theoretical remedy—to use a nonlinear scheme of aggregation, such as a minimal increase in evaluation of criteria when compared to reservation points—is too complicated to be useful in common law.

4 Hard versus Soft Systems Approaches

This leads us obviously to the question of whether hard mathematical approaches are of any use for negotiation processes that might require soft approaches. This is an old controversy in systems analysis; to quote Harold Barnett (private communication) from a discussion at IIASA in 1980: *soft models—hard thinking, hard models—soft thinking*. One can only wonder why this old dispute has erupted again in negotiation theory.

One avenue of criticism of mainstream decision theory that went even further than others claims that individual decisions are *too complicated to be explained by analytical theory*: that decisions are made in a *deliberative way*. This criticism came from the perspective of *general systems theory* [4] as a so-called *soft systems approach*; it also stressed the role of *synergy*—that the whole is bigger than the sum of its parts. It stressed, moreover, the fact that in a more complex cognitive or decision process, there is a eureka moment or *aha effect*, which can also be called a *cognitive enlightenment* effect. Thus, *soft or deliberative decision making* is based on perceiving the whole picture, observing it from various angles, and finding the right decision by expert intuition. This is supported by the argument about *reductionism versus holism*: an analytical approach implies a reductionist concentration on parts, while deliberation helps to concentrate on the whole (the *Gestalt* perception) of a problem.

This criticism was most convincingly presented by Dreyfus and Dreyfus [6]. Among other things, they show that *the way decisions are made changes with the increasing level of expertise* of the decision maker. While a beginner, novice, or apprentice needs analysis to support his or her decisions, specialists, experts, and master experts make decisions instantly, intuitively, or rely on deliberation. In the case of operational or repetitive decision making, the empirical evidence for such a thesis is abundant: for example, when driving a car, a novice has to think before shifting gears, while experts make decisions with their *entire body*. In the case of strategic or creative decision making, when new types of decisions are often needed, the Dreyfus brothers designed and conducted an experiment involving playing chess. They saturated the analytical part of the brain of a chess player with tedious computations in interaction with a computer, while letting him or her at the same time compete in a chess match. Apprentice players could not play when the analytical parts of their minds were saturated, while this did not make any difference to chess masters. The two authors believed these results were related to the way we perceived reality by

4.5 points for intelligence. It is easy to see that the third option will never be chosen, no matter what weighting coefficients are selected in the weighted-sum procedure.

images and Gestalt, and how we recognize patterns, to the typical connections of *pathways in the brain* established by training, to the functions of *the right-hand part of the brain*, and to the concept of intuition. However, they do not analyze the latter concept.

We should be aware, however, that the development of *soft systems approaches* since about 1980 has *hardened paradigmatically* and resulted in methods that might be excellent but are inconsistent in their assumptions. An example is *Soft Systems Methodology (SSM)* (see Checkland [5]). SSM stresses listing diverse perspectives, so-called *Weltanschauungen, problem owners*, and following an open debate representing these diverse perspectives. Actually, when seen from a different perspective, that of hard mathematical model building, SSM—if limited to its systemic core—is an excellent approach, consistent with the lessons derived earlier from the art of modeling engineering systems. More doubts arise when we consider the paradigmatic motivation of SSM, which is presented by Checkland as a general method, applicable in interdisciplinary situations; but a sign of a paradigmatic hardening is his opinion that *soft systems thinking* is broader and includes *hard systems thinking*.

Should SSM also be applicable to itself? It includes two *Weltanschauungen: hard and soft*; accordingly, *the problem owners of hard Weltanschauung should have the right to define their own perspective*. However, hard systems practitioners have never agreed with the definition of hard systems thinking given by Checkland. He defines hard systems thinking as the belief in the statement by Ackoff [1] that *all problems ultimately reduce to the evaluation of the efficiency of alternative means for a designated set of objectives*. On the other hand, hard systems technological practitioners say *no, that they are hard because they use hard mathematical modeling and computations, but for diverse aims, including technology creation, they often do not know what objectives they will achieve*. As a result, hard systems technological practitioners and soft systems researchers simply do not understand each other.

Another related subject is the soft validity of using the adjective *rational* in its hard mathematical sense of rational choice theory. Scientific theories require quite different definitions of rationality, discussed by philosophy of science (e.g., Popper [18]). Even in this field, there are diverse interpretations of this concept: *falsificationism* by Popper, *historicism of changing paradigms* by Kuhn [11], and several others (see [8]) versus *evolutionary epistemology* as initiated by Lorentz [13]; see also Wuketits [26]). We should also recognize that the *sociology of science* today promotes a more relativistic, postmodernist view that denies the concept of rationality and objectivity, arguing that all knowledge is *subjective*—results from a discourse, is constructed, negotiated, relativist, and motivated by the desire for power and money. Such an *episteme* is not internally consistent and expresses only a crisis in sociology at the end of the industrial civilization era. Postmodernist sociology treats science as a social discourse. What happens if we apply this to sociology itself? A paradox: *sociology is a social discourse about itself* (see [10]). Hard science and technology represent different cultural spheres from sociology. But if the postmodernist sociology of science (represented, for example, by Latour [12]) tells hard science and technology that they do not value truth and objectivity, only power and money, then this opinion reflects poorly on sociology and is a sign of its crisis.

From the more normative, Popperian perspective, rationality of a statement concerning the real world implies its empirical testability and falsifiability (at least in principle). When seen from this perspective, the term *rationality* in decision theory has been used rather unfortunately. Interpreted as a mathematical statement, the assumption of utility maximization is logical and consistent with many other developments of decision theory. However, interpreted as a description of reality—that real "rational" decisions are always made in the best-perceived interests of the decision maker—this assumption itself is not rational in the normative, Popperian sense of falsifiability. Popper uses precisely this as an example of an unrefutable theory. If any other concerns—honor, altruism, etc.—can be subsumed into the individual utility, then the assumption of best-perceived individual interests cannot be falsified in principle but expresses an ideology instead of science—at least according to Popper.

Kuhn [11] argues that certain components of ideology are always present in a paradigm of normal science. As described by Kuhn, a paradigm is not necessarily formed or changed because of its empirical testability; its central statements are often chosen in order to simplify analysis and its testability depends on additional assumptions and interpretations. The same applies to the paradigm of utility maximization. If modestly interpreted in its descriptive power, this paradigm is empirically testable and thus rational, even in the sense of Popper. Indeed, a statement such as *an averaged behavior of certain aggregate classes of individuals acting in economic markets is usefully described by assuming that they maximize utility or value functions* is falsifiable and has been tested in many economic studies. Unfortunately, this applies neither to individually descriptive nor to prescriptive (sometimes called normative) interpretations that every individual makes (or could, or should) optimize his or her aggregated, best-perceived interests in every decision.

As a scientific theory, hard rational choice theory is hardly rational—at least according to the requirements of Popper. It cannot be falsified; hence it is only a mathematical language, not a scientific theory. This leads us back to the discussion on normative or prescriptive aspects of decision theory; unfortunately, in order to be useful in applications (and rational as a scientific theory), it must be treated as a descriptive theory and must be subject to falsification attempts. Therefore, *the rationality of choice differs essentially from the rationality of knowledge*.

5 A Rational Theory of Intuition and Its Implications for Negotiations

The controversy regarding hard versus soft approaches has led the author of this chapter to develop and test (empirically, in the sense of falsification attempts) a rational theory of intuition. Intuition was regarded by Plato, Descartes, Kant, and others as an infallible source of inner truth concerning *a priori synthetic judgments*; but it turned out later (with non-Euclidean geometry and with relativity theory) that such judgments are not necessarily true; intuition is fallible. Nevertheless, in a period of knowledge-based economy we need a rational theory of intuition because it is a powerful source of innovative knowledge creation, even if it might be fallible and must thus be checked empirically.

This theory uses facts known from modern telecommunications and computational complexity theory: the ratio of the volume of visual and audio signals is about 100; thus the ratio of complexity of processing them can be assessed as at least 10,000.[4]

Further, the theory then concentrates on a *thought experiment*: let us reflect on the period of human civilization when language was discovered. This discovery was a powerful shortcut in human development, as it simplified reasoning about 10,000 times. It started with easy human communication; it enabled the intergenerational transfer of knowledge, leading to the entire evolution of human civilizations. However, it also suppressed the old, preverbal way of reasoning, which was much more complicated and also much more powerful.

The conclusion is that we still use the old, preverbal way of reasoning subconsciously. As it is preverbal, we lack words to describe its functioning—and for lack of a better word, we call it intuition. As it is done subconsciously, we obtain better results this way when we suppress consciousness—as when sleeping or in Zen meditation. Thus, *intuition can be rationally defined as preverbal, quasi-conscious or subconscious information processing, much more powerful than verbal reasoning, utilizing aggregated experience and training and performed (most probably) by a specialized part of the human mind* (see [23]).

This definition leads to a rational theory of intuition, because it can be falsified: various conclusions resulting from this definition can be put to practical or scientific testing. Many known results from cognitive science, from hemispheric asymmetry of our brains, etc., support this theory; this also resolves some of the controversies about hard versus soft systems analysis. Together with this definition we can recall the distinction of at least two types of decisions, which is especially important for intuitive decisions: repetitive and *operational*, when performing functions to which we are well trained, or creative, and *strategic*, when solving a novel problem.

The theory defines intuitive decision processes and their phases and conditions for stimulating intuition. For example, a *strategic intuitive decision process* can be defined in the following stages:

1. *Recognition*, which often starts with a subconscious feeling of uneasiness (well known, for example, to senior decision makers if they are not satisfied with a piece of information prepared by their staff). This feeling is sometimes followed by a conscious identification of the type of problem involved.
2. *Deliberation or analysis*: for experts, a deep-thought deliberation suffices, as suggested by Dreyfus and Dreyfus. Otherwise, an analytical decision process

[4] This is a lower-bound estimate because a quadratic increase in the complexity of processing large sets of data is one of the mildest nonlinear increases. Thus, the old saying that a picture is worth a thousand words should be changed to a picture is worth at least ten thousand words.

is useful—with *intelligence* and *design* phases as defined by Simon [21],[5] but suspending the final elements of choice.

3. *Gestation and enlightenment*: this is an extremely important phase—we must have time to forget the problem in order to let our subconscious work on it. The expected eureka *effect* might appear but may not be consciously noticed; for example, after a night's sleep it is easier to generate new ideas (one of the reasons why group decision sessions are more effective if scheduled for at least two days).
4. *Rationalization*: in order to communicate our decisions to others we must formulate our reasons. The word *rationalization* is used here in a neutral sense, without necessarily implying self-justification or advertisement, though the meaning of the word often includes at least the latter. For example, when writing a research paper, we obviously rationalize our deeper thoughts—and occasionally also advertise them. Note the similarity of this phase to the classical phase of *choice*.
5. *Implementation*: which might be conscious, after rationalization, or immediate, and even subconscious.

The rational theory of intuition stipulates that intuition is the result of training (it is a preverbal processing of our entire experience). We can be trained or even train ourselves to concentrate on nothingness in order to stimulate our subconscious, preverbal, trained abilities. An athlete concentrating before competition—say, in archery—is using precisely this method.

From this perspective, diverse theoretical and practical conclusions can be derived. Theoretical conclusions are related to *microtheories of knowledge creation* that emerged at the turn of the twenty-first century. The epistemology of the twentieth century concentrated either on knowledge validation or on *macrotheories of knowledge creation* on a grand historical scale, as expressed by Kuhn [11]; but this did not lead to conclusions about how to create knowledge for today and tomorrow for the requirements of the knowledge economy. Therefore, most of the new *microtheories of knowledge creation* came from outside philosophy, from systems science and management science, for example, *Shinayakana Systems Approach, Knowledge Creating Organization with SECI Spiral* and several others (see [14], [16], [24]). All such theories used the concepts and stressed the role of *tacit knowledge, intuition, instincts, and myths*; the rational theory of intuition outlined above has been used to integrate such diverse microtheories of knowledge creation.

However, we are interested in the possible conclusions from a rational theory of intuition for negotiation theory and practice; and such conclusions for negotiation processes can be drawn theoretically and then tested empirically. For example, there is a conclusion—actually used earlier in soft theory and in the protocol for international negotiations—that negotiating sessions would benefit from being limited to an early time of the day, while evenings should be devoted to relaxation with a prohibition on discussing the negotiated subjects. The validity of such conclusions can be

[5] We recall that Simon [21] defined the three main stages of a decision process: intelligence, design, choice. Later, an obvious fourth stage was added: implementation.

and has been easily tested. It is sufficient to prepare for students some difficult simulated negotiation problems concerning hidden possibilities of common gains and then to let them solve such problems in simulated negotiations over a given period, say six hours. We separated the students into two groups: one negotiating for six hours with short breaks, another having a longer break and a night's sleep between two three-hour sessions. Since the problems concerned simulated negotiations, it was easy to judge which set of negotiators did best—the average achievements in the group with a longer break and sleep between sessions increased significantly. Other conclusions on the effect of intuition on negotiations can also be drawn and tested.

In addition there is a general conclusion resulting from the synthesis of hard versus soft approaches. *Hard approaches should not be treated as a final synthesis of knowledge, only as tools for constructing virtual realities or rather virtual laboratories—computerized gaming, simulated negotiation exercises—that might be used for training our intuition.* This conclusion deserves an illustration by a postulate on how best to prepare for difficult international negotiations on complex problems. We should use available knowledge, even in the form of mathematical, formal models prepared by experts. But in order to use this knowledge to prepare negotiations, it must be integrated with the life-long experiences of good negotiators—and this integration must occur on a subconscious level, become part of the intuition of the negotiator. Only then might we be sure that formal knowledge would actually be used in difficult international negotiations.

6 Conclusions

There are several conclusions to this chapter. Here are just a few:

1. It is natural, perhaps even inevitable, to treat the usefulness of game theory in negotiations in the *meta-theory* sense: the basic concepts of this theory are more important than its specific results, since they contribute more to our intuition.
2. In contrast, the results of game theory could also be helpful as tools for constructing virtual realities or virtual laboratories that might be used to train our bargaining intuition. To achieve this result, however, game theory must be treated descriptively, not prescriptively, and must be combined with gaming.
3. The main disadvantage of rational choice theory is that it is too simplistic, motivated originally by describing an average economic decision maker and escaping into prescriptive or normative arguments instead of trying to enrich the description of real-life situations. Real life and individual decision makers are much more complex than economics.

References

1. Ackoff RI (1974) The future of operational research is past. Journal of Operational Research Society 30(2):93–104

2. Amman M, Johansson A, Lueckewille W, Schoepp H, Simon A, Warren R, Gonzales T, Tarasson L, Tsyro S (2001) An integrated assessment model for fine particulate matter in Europe. Water, Air and Soil Pollution 130:223–228
3. Arrow KJ (1983) Social choice and individual values. Wiley, New York, USA
4. Bertalanffy L (1968) General systems theory: foundations, development, applications. Braziller, New York, USA
5. Checkland PB (1985) From optimizing to learning: a development of systems thinking for the 1990s. Journal of the Operational Research Society 36:757–767
6. Dreyfus H, Dreyfus S (1986) Mind over machine: the role of human intuition and expertise in the era of computers. Free Press, New York, USA
7. Fisher R, Ury W (1981) Getting to YES. Houghton-Mifflin Co., Boston, MA, USA
8. Hacking I (ed 1981) Scientific revolutions. Oxford University Press, Oxford, UK
9. Kalai E, Smorodinsky M (1975) Other solutions to the Nash's bargaining problem. Econometrica 43:513–518
10. Kozakiewicz H (1992) Epistemologia tradycyjna a problemy współczesności. Punkt widzenia socjologa [in Polish] (Traditional epistemology and problems of contemporary times. Sociological point of view). In: Niżnik J (ed) Pogranicza epistemologii [in Polish] (The boundaries of epistemology). Wydawnictwo IFiS PAN
11. Kuhn TS (1970) The structure of scientific revolutions. 2nd edition, Chicago University Press, Chicago, IL, USA
12. Latour B (1990) Science in action. In: Postmodern? No, simply a-modern! Steps towards an anthropology of science. Studies in the History and Philosophy of Science 21
13. Lorentz K (1965) Evolution and modification of behavior: a critical examination of the concepts of the "learned" and the "innate" elements of behavior. The University of Chicago Press, Chicago, IL, USA, London, UK
14. Nakamori Y, Sawaragi Y (1990) Shinayakana systems approach in environmental management. In: Proceedings of the 11th World Congress of International Federation of Automatic Control, Tallin. Pergamon Press, 5:511–516
15. Nash J (1953) Two persons cooperative games. Econometrica 21:128–140
16. Nonaka I, Takeuchi H (1995) The knowledge-creating company. How Japanese companies create the dynamics of innovation. Oxford University Press, New York, USA
17. Nurmi H (1999) Voting paradoxes and how to deal with them. Springer-Verlag, Berlin, Germany, New York, USA
18. Popper KR (1975) The rationality of scientific revolutions. In: Harre R (ed) Problems of scientific revolution. 72–101, Oxford University Press, Oxford, UK
19. Raiffa H (1953) Arbitration schemes for generalized two-person games. In: Kuhn HW Tucker AW (eds) Contribution to the theory of games II. Princeton University Press, Princeton, NJ, USA
20. Raiffa H (1982) The art and science of negotiations. Harvard University Press, Cambridge, MA, USA
21. Simon H (1955) A behavioral model of rational choice. Quarterly Journal of Economics, 69:99–118
22. Wierzbicki AP (1992) Multiple criteria games – Theory and applications. X International Conference on Multiple Criteria Decision Making, Taipei, Taiwan
23. Wierzbicki AP (1997) On the role of intuition in decision making and some ways of multicriteria aid of intuition. Journal of Multi-Criteria Decision Analysis 6:65–78
24. Wierzbicki AP, Nakamori Y (2006) Creative space: models of creative processes for knowledge civilization age. Springer, Heidelberg, Germany
25. Wierzbicki AP, Makowski M, Wessels J (2000) Model-based decision support methodology with environmental applications. IIASA–Kluwer

26. Wuketits F M (1984) Evolutionary epistemology – A challenge to science and philosophy. In: Wuketits FM (ed) Concepts and approaches in evolutionary epistemology. D. Reidel Publishing Co., Dordrecht, The Netherlands
27. Zartman I.W., Faure G.O. (eds 2005) Escalation and negotiation in international conflicts. Cambridge University Press, Cambridge, UK

Negotiation Models and Applications

Daniel Druckman

George Mason University, Fairfax, Virginia, USA

1 Introduction

A variety of modeling approaches have been used as tools for analyzing aspects of negotiation and related interactions. Many models represent the mixed-motive feature of negotiation where parties are torn between cooperating to get an agreement and competing to get an acceptable agreement close to their own preferences. This feature is highlighted in the two-person games—particularly the Prisoner's Dilemma—constructed and analyzed in the game theory tradition. The focus of these games is on solutions or outcomes, with attention paid to the distinction between equilibrium (minimizing losses) and optimal (maximizing joint gains) solutions. Similarly, outcomes or end states are the focus of models in the tradition of decision analysis: the bargaining frontier is a tool for identifying optimal or integrative solutions. Other models concentrate more on the negotiation process or throughputs including how impasses are resolved and how a process moves or transitions through stages. Some of these models provide a diagnostic function from which advice is generated; other process models identify the conditions that can either stall the talks (as when the parties are stuck in a stage or have disagreements within delegations) or provide momentum toward an agreement (as when they move quickly from one stage to another or reach consensus within delegations). Examples of each of these types of models are discussed in this chapter.

This chapter is divided into two major sections, each of which contains several parts and is organized according to a set of six questions. The first section focuses on three models of the process. One is an operating model designed to resolve impasses with three mediation functions: diagnosis, analysis, and advice. The model, referred to as a negotiating support system, derives from a broad framework of the negotiating process. Another model construes negotiation as a stochastic process. It focuses on state (or stage) transitions and relates to a framework for analyzing turning points. The third modeling approach emphasizes nonlinear processes in the tradition of system dynamics. This type of model captures small group behavior leading to either consensual or fractious outcomes. The simulation analyses incorporate both social–psychological and information-processing variables. Further details about this approach are provided in Gabbay's chapter in this publication.

The second section presents two models of outcomes. The steps needed to perform a decision analysis are shown. These steps lead to the identification of an optimal solution. The discussion of game theory illuminates alternative forms for analyzing interdependent decision making. These are complementary forms that emphasize timing of actions (extensive form), lack of information (strategic form), or groups (coalition form). Both of these outcome-oriented approaches have made practical contributions: a cottage consulting industry has developed around decision analysis. The 1994 and 2005 Nobel Prize for Economics recognized the practical contributions of game theory. An extended example of a policy-relevant analysis is developed in Güner's chapter in this volume. This application elucidates the technical considerations that are entertained by a modeling analysis. The five modeling approaches are compared in terms of six features in a concluding section.

2 Modeling the Process: From Frameworks to Models

A framework is a broad conceptualization of a process. It serves primarily to organize thinking about complex phenomena and may take the form of a visual representation (e.g., [11], [12]) or a path through the chronology of a process (e.g., [4]). Models provide precision in the form of operational definitions for concepts, mathematical equations, statistical ratios, or computer programs. Three examples of attempts to move from broad frameworks to models are work on negotiation support systems (NSS), on departures or transitions during the negotiation process, and on small group decision processes. Each is discussed in turn according to a common format.

2.1 Negotiation Support Systems

What are the objectives? Negotiation support systems (NSS) operate in electronic environments that facilitate the execution of negotiating or mediating functions. An example is the current work on e-mediation [5]. This R & D project consisted of developing, implementing, and evaluating a conception of mediation that includes the functions of diagnosis, analysis, and advice. These functions are intended to facilitate the resolution of negotiating impasses.

What is the analytical tool? The key analytical tool takes the form of a sequence of computer-aided decisions: a diagnostic grid that answers the question, Where are we now?; an analysis that identifies the source of the impasse; and advice that answers the question, What should we do? These decisions are based on negotiators' answers to questions organized into five sections defined by a framework: issues, parties, delegations, situations, and process. The questions derive from research findings about negotiating flexibility. This NSS tool is intended to mediate bilateral disputes over multiple issues.

What is the mode of analysis? A question of interest is whether the operating NSS model works in practice. This question is addressed both with cases and by conducting a series of experiments. Nine historical cases were used to validate the diagnostic function. (The analysis and advice functions were not addressed in this exercise.) Analysts with specialized knowledge about these cases answered the NSS questions. The resulting diagnoses were compared to the actual outcomes of the cases, showing correspondence in eight of the nine cases. Further diagnostic work provided insight into the one case that was misjudged. (The computer decision of impasse did not correspond to the agreement obtained.) The computer captured the long impasse, which was due to delegation activities and that was eventually resolved in favor of one of the parties. The agreement resulted from an external event.

Three experiments conducted to date served to evaluate the impact of the three mediation functions on simulated negotiating behavior. The NSS was compared to several nonelectronic conditions, including a live mediator, as well as in alternative configurations (separate versus joint use of the NSS between rounds of the simulated negotiation). Significantly more agreements (across seven issues) were obtained by negotiators in the NSS conditions. The system performed best when the opposing negotiators used it together. Joint use of the system outperformed a live mediator scripted to perform the same three mediation functions; however, the live mediator produced more agreements than separate use of the NSS. These findings support the three-function conception of mediation.

What are some of the applications? The experimental work has shown that the NSS is helpful in achieving agreements in simulated negotiations. The electronically augmented simulation environment can also be used for training. Third parties and support staffs experience the implementation of mediation functions when impasses occur. But the NSS is also a valuable support tool in real-time negotiations. In addition to the mediation functions, it allows for "what if" evaluation of alternative tactical options. This is done by answering suites of questions in different ways and comparing the effects on diagnoses and analyses. These are some of the applications that remain to be tried.

What is the key advantage of the approach? This and other NSS provide efficient platforms for implementing mediation functions that have been shown to be effective. Under certain configurations (joint use) and situations (simulated negotiations) these systems may even outperform a live mediator. In addition to the obvious advantages of speed and efficiency, the e-mediation system performs analyses and provides advice based on the results of published research studies.

What are some of the limitations? This particular NSS is currently suited only to a bilateral negotiation format, the questions asked are limited to multiple and sometimes binary choices, and the way in which negotiators use the advice during the process of negotiating has not been explored. More important, perhaps, are findings that show an aversion to computer aids: simulation participants indicated a preference for noncomputer formats, despite the improved performance engendered by the NSS. Perhaps a more human-like system would improve user-friendliness.

2.2 Transition rates as stochastic processes

Objective. The objective of stochastic-process modeling is to identify transitions that occur during an interaction process. A recent project applies this modeling to the negotiation process, focusing on events that precede entry into and exiting from states. The states can be thought of as negotiation stages; the entry and exiting from them can be considered as departures. The analysis of negotiating departures has implications for change in a process toward achieving agreements or being stuck in impasses.

Analytical tool. The key analytical tool consists of two exponential functions. One is referred to as a decaying function, the other as a growth function. The decay function captures the assumption that rates of change out of a state (stage) decline as the length of time in that state increases. The function takes the following form:

$$q_d(t) = q_d(0) \cdot e^{-at}, \tag{1}$$

where $q_d(0)$ is its value at the time of the initiating event, and where a is a measure of the decaying rate.

The growth function is based on the assumption that transition rates increase over time in a systematic fashion. This is expressed as follows:

$$q_g(t) = q_g(0) \cdot e^{at}, \tag{2}$$

where $q_g(0)$ is its value at the time of the initiating event, and where a is a measure of the growth rate.

These functions call attention to the reversed processes of a decline and an increase in transition rates through time. Initiated events set in motion a process of declining attention to a past (initiated) event; terminating events set in motion a buildup of attention to a forthcoming (terminating) event. This is a general process that should also apply to negotiation: initiated events are the precipitants in our turning points framework [4], transition rates are departures, and terminating events are deadlines. These processes can be explored by providing estimates for the parameters of the above equations through experimentation.

Mode of analysis. The question asked is whether the above functions describe a negotiation process as it unfolds through a series of precipitants, departures, and consequences. Departures are construed by the model as bounded states with entering and exiting as transitions into and out of them. Time spent in each state as well as rates of change can be assessed with experimental data. Transition rates can be calculated in a way that is similar to that done for concession rates in bargaining and then transformed into probabilities. The two functions, decay and growth, can be compared in terms of how effective they are in predicting transition rates with or without preannounced or suddenly imposed deadlines, referred to as terminating events (see Krause *et al.* [7], for an example of the way that experimental data are used to evaluate formal models). A number of other research questions about transitions (or turning points) can be addressed with this stochastic modeling approach. For example, does the type of initiated event or precipitant make a difference? (Our framework distinguishes among three types: external, substantive, and procedural.)

Another question is whether the type of consequence of a departure influences transition rates. An escalatory consequence may increase the rate as negotiators move more quickly out of previous states. A de-escalatory consequence may decrease the rate, causing negotiators to remain in a state longer and thus either delaying or foreclosing the transition to another state. These hypotheses about consequences suggest that what happens after departures can also influence transition-rate frequency in a negotiation that is modeled as a continuous stochastic process. They are the bases for programmatic research that remains to be done.

Applications. Coleman [2] reviewed studies showing that the decaying function (Equation 1) captures such processes as a decline in marriageability from the point of first eligibility, a decline in job changes as age increases, and a decline in residential mobility since the last move was made. In each of these examples, rates of change decrease as a function of time since the last change (or entry into a state) occurred. In negotiation, the function may describe a decline in departures from the time since the last departure or initiating event occurred. The growth function (Equation 2) is illustrated by an auction where bidding could occur until a candle flickered out and the last bid before this terminating event was the bid taken. Similarly, an increase in negotiating departures may be expected to occur as a deadline approaches. Applicability to negotiation processes would seem to follow from the examples in other domains.

Key advantages. The functions require estimates of movement into and out of states. They emphasize the importance of time and terminating events as influences on rates of change. These are relatively easy to estimate for most negotiations. An advantage is reduced uncertainty about the likelihood of change in negotiation: the key predictor is time in state. Another advantage is that the frequency of departures can be influenced by altering terminating events.

Limitations. The functions assume that processes proceed in a linear fashion toward an end state. Feedback mechanisms that encourage learning—and, then, ideas that lead to unexpected changes—are not taken into account.[1] Nor is guidance provided in terms of how states are to be defined. There is considerable controversy among negotiation theorists about the number and types of stages through which a process passes (or must pass) on the way to agreement. The stages are more elusive than the state of years in marriage, living at a particular residence, or time spent in a job. They usually refer to particular accomplishments (agenda setting, formula definition) rather than maintenance of a particular status. The transitions are regarded more as progress (turning points) than simply as change of status. The implications of these differences for stochastic analysis remain to be explored.

[1] This limitation is addressed in Gabbay's chapter in this publication. His model attempts to capture such nonlinear transitions or turning points as the sudden emergence of consensus or major schisms. The model also considers the nonlinear impacts of external events (precipitants) on the process (departures and consequences.)

2.3 Nonlinear system dynamics

Objective. A third type of process model has been used to capture decision making in small groups as a function of various influences. An example of application is presented in the chapter by Gabbay. Focusing primarily on consensual group decisions, he uses nonlinear differential equations to show how various forces influence decisions. The forces are variables discussed in the social–psychological (e.g., group cohesion) and cognitive (e.g., incoming information) literatures. The modeling process is shown to have implications for international negotiation.

Analytical tool. The primary analytical tool is referred to as coupled nonlinear differential equations. The equations are shown in an Appendix to Gabbay's chapter. I will only summarize them here. The model highlights three forces referred to as self-bias, group influence, and information flow. Each of these forces is divided into component parts consisting of preferences and commitments (self-bias force), coupling strength and latitude of acceptance (group-influence force), and aspects of messages and context (information-flow force). Linear and nonlinear equations are used to capture rates of change in group members' positions, pairwise coupling for group influence, and valences. An equation (referred to as an equation of motion) is written for each group member. These equations then form a system of coupled, nonlinear ordinary differential equations. Parameters of these equations are estimated for analysis.

Mode of analysis. The analyst performs sensitivity analyses, where a variety of parameter values (e.g., coupling strength) are explored for effects on such outcomes as whether decisions are consensual or the result is a deadlock. The analyses are guided by assumptions about member commitments, symmetry of bilateral influence, and latitudes of acceptance. Initial evaluations are based on simple situations—a group of two members whose decisions are bifurcated as either consensual or deadlock. Various levels of coupling strength (degree of mutual influence) are explored, showing a clear relationship between strength—as weak, intermediate, and strong—and decision consensus. For each outcome, there is a discernable equilibrium solution. Further analyses take into account the effects of incoming information, groups with more than two members, and various types of network structures. An interesting insight emerged from the analysis of incoming information: high sensitivity to context increases the chances that a group will undergo frequent transitions between consensus and deadlock—in other words, higher transition probabilities. This variable can be further specified by distinguishing among such aspects as attentiveness to information, salience or "loudness" of the information, and the intrusion of external events. These are possible additional parameters of the model.

Applications. This type of model is particularly relevant for studying the internal group dynamics of negotiation. A number of insights with practical implications have been suggested by the analyses. One concerns the difference between revealing divisions within delegations and presenting a united front. Another concerns giving advice to third parties about when to increase or decrease pressure on negotiators in order to resolve an impasse. A third implication is for mediator strategies concerning meeting separately with the parties or encouraging direct communication between them. And, a fourth useful insight highlights shifts that may occur between contentious and problem-solving behavior in negotiation. The conditions for rapid rather than gradual changes are identified by the modeling analyses. A connection is made here to the stochastic models discussed in the previous section. Both types of model deal with transition rates. A difference is that the stochastic models focus on time in states or stages, including initiating and terminating events; the dynamic system models highlight bifurcated processes such as competitive or cooperative behavior. To date, however, these practical implications have not been channeled to policymakers.

Key advantages. Like other modeling approaches, system dynamic models add precision to empirical analyses of negotiation. They also make explicit (in the form of differential equations) the assumptions that guide hypothesis testing: in this case, the equations capture nonlinear processes. Unlike the other modeling approaches discussed in this chapter, the system-dynamic approach reveals the sudden emergence of group consensus or schisms due to small changes in parameter values. By doing so, the models provide a formal theory to guide the empirical work on turning points in negotiation [4]. They also have heuristic value as evidenced by the many ideas generated for research on small-group decision making. Some of these ideas would seem to have implications for policy as well.

Limitations. Three limitations are evident. One is that, to date, the models have concentrated on bifurcated outcomes, known also as “switch-on, switch-off” variables. Negotiation outcomes are often more nuanced and include complex multi-issue packages. The fit between this family of models and negotiation processes seems relatively weak. They seem to capture generic small group processes, which are only a part of negotiation. Another problem is the limited focus on certain social–psychological and cognitive variables, particularly those related to positions, cohesion, and incoming information flows. Missing are such variables as the distinction between distributive and integrative outcomes, tactics, emotions, cultures, power, and the role of relationships. A third limitation is a lack of empirical evaluation. There is uncertainty about an appropriate platform for data collection. The respective contributions of laboratory experiments and case studies need to be evaluated. These are the next steps for these modelers.

3 Modeling Outcomes: Decision Analysis and Game Theory

Two closely related modeling approaches are decision analysis and game theory. Both approaches use the parties’ different preferences for outcomes to derive a solution considered as a “best” agreement on single or multiple issues. A difference between the approaches is their focus on single parties (decision analysis) or both (all) parties (game theory). The one-sided approach of decision analysis asks, How can a particular set of preferences be reconciled with another, different or conflicting set of preferences to produce an optimal outcome? The analysis is illustrated in the first part of this section. The two- (or more-) sided approach of game theory asks, Which joint decision is likely to produce an optimal outcome? A key contribution of game theory is the idea of interdependent decision making. This is illustrated with a detailed example in the chapter by Güner in this publication.

3.1 Decision analysis

Objective. An objective of decision analysis for negotiation is to identify an optimal agreement. This is also known as an integrative agreement, defined as an outcome that maximizes the preferences of all the parties. The analysis provides parties with a slate of alternative possible agreements, allowing them to compare different kinds of bargaining strategies. They can, for example, gauge improvements that can be made over compromising (each party gives something up) on the issues.

Analytical tool. The key tool for comparing alternative outcomes is the bargaining frontier. This tool serves to evaluate alternative agreements (or packages) in terms of benefits achieved for each party. The benefits are depicted in terms of the extent to which each outcome maximizes the joint "return," operationalized in terms of an ordering of points that depicts gains for each side. The points are arranged on a concave surface (frontier) that shows just how far joint gains can be pushed for a particular negotiation. The best agreement is the one closest to the point furthest out on the frontier. This point defines an integrative agreement (for example [19], *Figure 1*).

Mode of analysis. The analysis consists of a series of steps intended to evaluate each of several possible outcomes in terms of its position on the bargaining frontier. The nine steps are as follows:

1. Identify the issues and articulate the positions of each party on them.
2. For each party, indicate the impact of each issue on its interests—for example, the issue of compensation has a large impact on the interest or criterion of expenditures but a negligible impact on the interest of human rights.
3. Express the relative importance of the interests served by each issue as a weight ranging from 0 to 10 and divide by the sum of the weights (a normalized weight).
4. Consider possible compromise agreements, given the initial positions of both (all) parties.
5. Quantify the options by assigning a value between 0 and 100 for each possible outcome (e.g., 0 for other side gets its preferred position, 50 for compromise, 100 for own side gets its preferred position).
6. Multiply each value by a normalized weight.
7. Develop packages consisting of outcomes on each of the issues (own side's position, other's position, compromise).
8. Evaluate each package in terms of its location on the bargaining frontier.
9. Choose the package closest to the point that maximizes each party's preferences, which is also the point furthest out on the frontier.

The sequence of steps identifies the outcome that improves the benefits for each party over a compromise agreement. In most applications, this involves a mix of compromise and alternation between one or the other side's preferences.

Applications. There have been a variety of applications of decision analysis to problems of negotiation. These include both bilateral and multilateral negotiations.

Raiffa's [10] application to the Panama Canal talks contributed to the United States (US) delegation's planning before and during the talks, which culminated in an agreement in 1977. Ulvila's [19] application to the Philippines' base rights talks identified a package that combined compromise on some issues and alternation on others (the US position accepted on some issues, the Philippines' position accepted on others). That package was shown to maximize the benefits to both parties, effectively expanding the bargaining frontier furthest. With regard to multilateral negotiations, Ulvila and Snider [20] evaluated four alternative packages on 11 criteria (interests) for 10 parties on a set of oil tanker issues. The assessments were provided by Coast Guard personnel. These authors identified a new package that resulted in favorable values for all countries. It was the package that surfaced as the agreement. Other multilateral applications include the General Agreement on Tariffs and Trade (GATT), the Single European Act of the European Council, and the United Nations Conference on Environment and Development (UNCED). The analysis of UNCED's financial resource issues led to suggestions for a funding mechanism for future environment-development projects and agreements on acceptable funding target levels [18]. Friedheim [6] performed a massive multiattribute utility (MAU) analysis of the UN Conference on the Law of the Sea (UNCLOS) (discussed in conjunction with a different model in the chapter by Antrim in this publication) in order to apply and test a number of propositions from negotiation theory.

Decision analysis has also been found to be a useful tool for making career choices. I used this approach to evaluate alternative job choices. By comparing results obtained from an evaluation of the jobs at the time of the offer with an evaluation of the jobs three years into the future, I was able to identify areas for bargaining. Those areas were issues for demands that would increase the value of the new job in light of a strong alternative (staying at the current job).

Key advantages. The applications highlight a key advantage of decision analysis, which is to provide advice to negotiating delegations on possible outcomes that increase benefits over compromise agreements. The calculations are easy to learn and understand and can be performed with a hand calculator. They can readily be updated with changing preferences and performed before (prenegotiation) or during the talks (between rounds). This approach has been found to be useful for planning with regard to the making of demands, accepting offers, and developing strategies for joint problem solving. Joint use of this tool could facilitate a negotiation. Indeed, it would be interesting to assess the extent to which applications of decision analysis are effective in transforming competitive negotiations into problem-solving discussions.

Limitations. Several limitations are apparent. One is whether accurate information can be obtained about each party's preferences: parties are often delegations divided on preferences; it may be easier to assess preferences for accessible delegations or those from the same country as the decision analyst. Further, gauging changes in preferences may be difficult. Analysts must monitor the negotiating delegations on a regular basis. Another limitation is that the analysis ignores interdependencies in parties' choices and among the issues. It also takes the focus off an interaction process in favor of expressed preferences and interests. These limitations are addressed by game theory (on interdependent decision making) and process analysis (changing evaluations based on new information). We turn now to game-theoretic modeling.

3.2 Game theory

Objectives. Game theory is an analytical tool used to study situations of strategic interdependence where decision makers' actions affect each other. The application of game-theoretic frameworks to negotiation problems can help to improve and generate explanations and predictions. The improvement comes both in the form of negotiation outcomes as well as refutation or verification of "accepted" insights and explanations. If not applied, however, game theory can remain an inaccessible mathematical mode of analysis, solving only abstract theoretical problems.

Analytical tool. The identification of players, players' actions (and their sequence), strategies, information conditions, and payoff functions constitute the main tools of game-theoretic analysis. The establishment of a correspondence between real-world interactions and these theoretical concepts is necessary to comprehend negotiations at a practical level. The analyst must answer the following questions: Who are the decision makers that, by defining the game structure, have an impact upon the benefit received by players? Who are the players? What choices and ranges of choices are available to them? Do they interact with or without knowing previous moves? When do they have to make a decision? What gains and costs occur at the end of their interactions? The answers will depend upon the specific negotiation problem at hand and upon the modeler's skill at simplifying complex phenomena. The identification of gains and costs or sequences of decisions may not be clear at the outset. The analyst generates explanations by assuming relations among a few variables. Modeling efforts discipline thoughts, verify connections, prevent inconsistencies, and produce transparent arguments [9].

Mode of analysis. There are three forms of game-theoretic analysis: extensive form, strategic form, and coalitional form. The first level requires the specification of players, actions available to them, information conditions, the sequence of actions, and payoffs. The second concentrates upon players, strategies, and payoff conditions without attention to information availability. The third leaves out strategies and focuses on what groups of cooperative players can achieve in a game. In practice, if timing of actions and information conditions are important in a negotiation, then the extensive mode of analysis is appropriate. If, however, players ignore others' decisions when they act, the strategic form becomes more attractive. When there are groups of negotiators cooperating among themselves and obtaining a specific value, the attention turns to possible final distributions of resources among these groups and their members. The coalitional form can then be used to understand a particular negotiation problem [8]. If, for example, a country is deterred from taking an action under another country's threat, then the extensive form is suitable. There is a clear sequence of actions and information conditions in a deterrence situation: for example, a player issues a threat and the target, informed of the player's previous move, decides whether to act against the threatening player's interests.

Applications. A number of applications are developed in several of the chapters in this volume. Güner's chapter derives policy implications from an analysis of a current international dispute between Greece and Turkey over territorial waters. Indeed, the game-theoretic literature is noted for drawing practical implications from normative analyses. Well-known examples are Schelling's [14] work on coordination problems and negotiating tactics, Snyder and Diesing's [17] use of matrix games for analyzing a host of actual international conflicts, Brams' [1] work on applying game theory to superpower conflict, and Shubik's [15] analyses of strategic behavior in relations among nuclear powers. Many of the applications derive from the notion of solutions to the bargaining problem. These solutions invoke ideas about optimality and equilibria (see Schellenberg and Druckman, [13], for a review).

Key advantages. The implications of game-theoretic analyses are often difficult to work out with verbal (ordinary-language) arguments. In one sense, the equilibria compel the modeler to show how variations in parameters produce different decisions, that is, which equilibria produce which actions on the basis of the values that costs of war and probabilities of victory can take. (See Gabbay's chapter in this publication for an analysis of the ways in which group processes may influence the stability of equilibrium solutions.) They also delimit the range of variables one should use at the step of interpretation. Some practitioners may think that the implications are not worth the strenuous modeling effort since their own evaluations seem to produce the same interpretations. Two points here are in order: first, the equilibria are correctly deduced from a set of clear assumptions, and, second, the point of view that these game-theoretic conclusions are already known may not be justified [8]. One (a non–game theorist) can judge, for example, that Greece would not disrupt the status quo if a war is inexpensive for Turkey. However, Greece can extend its territorial waters under such a condition, given that its war costs are reduced. This interpretation is suggested by Güner's analysis in this volume. (See also [16], for another application to this conflict.)

Limitations. In general, game theory applications to real-world interactions can imply no solution or multiple solutions. In repeated games, for example, it is quite possible to generate a large number of equilibria, and, in games in coalitional form, it is possible to have no solution such as the core. This can lead to interpretations where the former implies competing explanations while the latter suggests no explanation. Perhaps a more serious limitation is that different evaluations of game rules are equally applicable to the same negotiation problem. Payoffs assigned to the same outcomes for the same bargaining problem can differ across analyses. Nevertheless, this is not an insurmountable problem.

4 Conclusions

Eight features of the five models discussed in this chapter are compared in *Table 1*. The primary dimension of difference is between models that focus on the process (NSS, stochastic modeling, system dynamics) and those concerned with outcomes (decision analysis, game theory). However, the table summarizes other differences

Table 1. Features of the Five Modeling Approaches

Features	NSS	Type of Modeling: Stochastic models analysis	System decision dynamics	Decision analysis	Game theory
Purpose	Impasse resolution	Monitoring and forecasting transitions	To generate insights about group decision making	Identifying integrative agreements	Discovering equilibria, evaluating options
Focus	Tactics, strategies (advice)	Process dynamics	Consensual and dissensual group decisions	Preferences on issues, outcomes	Inter-dependent choices/ moves
Tool	Electronic decision aid	Transition-rate equations	Coupled nonlinear differential equations	Bargaining frontier	Strategic choice matrices
Data	Self-report survey	Time in states (stages)	Simulated parameter values	Outcome preferences and interest articulation	Preference ranking of choices
Validation	Case and experimental analyses	Comparative case studies	Experiments and case studies	Real-time applications	Compare recommended to actual decisions in a modeled case
Applications	Simulated negotiations and scenario evaluation	Patterns of change in life choices	Design of policy-making groups	Actual bilateral and multi-lateral negotiations	Policy advice and evaluation
Advantages	Speed and efficiency	Reduced uncertainty; influencing change	Formal theory to guide empirical work; nonlinear processes	Improved outcomes	Clearly deduced options/ strategies
Limitations	Possible aversion to the computer format	Limited to linear processes; defintion of states not clear	Focus on bifurcated variables; limted range of variables explored	Accuracy of preference information	Binary choices provide a limited range of outcomes; nuances may be missed

among the models. A key difference among them is the analytical tool. The tool also determines the kind of data used for analyses and validation. Although a variety of applications are conceivable, only a few have been tried to date. Perhaps the most practical impact on policy-related matters has come from decision analysis and game theory. Those documented applications have less to do with the complex technical requirements of the approaches than with the resulting advice: clearly stated options for action and guidelines for choosing among them. As with other types of models, skilled analysts perform the tasks while skilled practitioners apply the results. This collaboration is vital for the successful implementation of modeling. (See [3], for a discussion of the craft of research consulting.)

The differences among the modeling approaches suggest that they may be used in complementary ways. One example is that the analyses performed by an NSS can include options for packages generated by a decision analysis. Another is that that the advice generated by an NSS or the options derived from a decision or game-theoretic analysis can contribute to transitions out of a state or increase the rate of process transitions analyzed by stochastic and system dynamics models. They may also influence the rate of movement out of states in which little progress is evident. A third is that the implications from a game-theoretic analysis—such as the interpretation of the equilibria from the example in Güner's chapter—may strengthen alternatives or reframe the issues as suggested in the advice offered by an NSS. A fourth is that the insights about small group decision making generated from system-dynamic simulations can be included as part of the advice offered by an e-mediation system. These examples illustrate the value of combining the models. They address a key question posed by research on negotiation: What is the relationship between processes and outcomes? They also address a concern of negotiation practitioners and policymakers who ask: What can be done during the negotiating process to encourage certain kinds of outcomes?

References

1. Brams SJ (1986) Superpower games. Yale University Press, New Haven, CT, USA
2. Coleman J (1973) The mathematics of collective action. Aldine, Chicago, IL, USA
3. Druckman D (2000) The social scientist as consultant. American Behavioral Scientist 43:1565–1577
4. Druckman D (2001) Turning points in international negotiation: a comparative analysis. Journal of Conflict Resolution 45:519–544
5. Druckman D, Druckman JN, Arai T (2004) e-Mediation: evaluating the impacts of an electronic mediator on negotiating behavior. Group Decision and Negotiation 13:481–511
6. Friedheim R (1993) Negotiating the new ocean regime. University of South Carolina, SC, USA
7. Krause RM, Druckman D, Rozelle R, and Mahoney R (1975) Components of value and representation in coalition formation. Journal of Peace Science 1:141–158
8. Kreps D (1990) A course in microeconomic theory. Princeton University Press, Princeton, NJ, USA

9. Powell R (1999) The modeling enterprise and security studies. International Security 24:97–106
10. Raiffa H (1982) The art and science of negotiation. Harvard University Press, Cambridge, MA, USA
11. Randolph L (1966) A suggested model of international negotiation. Journal of Conflict Resolution 10:344–353
12. Sawyer J, Guetzkow H (1965) Bargaining and negotiation in international relations. In: Kelman HC (ed) International behavior: a social-psychological analysis. Holt, New York, USA
13. Schellenberg JA, Druckman D (1986) Bargaining and gaming. Society 23:65–71
14. Schelling T (1960) The strategy of conflict. Harvard University Press, Cambridge, MA, USA
15. Shubik M (1993) Models of strategic behavior and nuclear deterrence. In: Tetlock P *et al.* (eds) Behavior, society, and international conflict. Vol. 3, Oxford University Press, New York, USA
16. Smith R, Sola M, Spagnolo F (2000) The prisoner's dilemma and regime-switching in the Greek-Turkish arms race. Journal of Peace Research 37:737–750
17. Snyder GH, Diesing P (1977) Conflict among nations. Princeton University Press, Princeton, NJ, USA
18. Spector BI (1993) Decision analysis for practical negotiation application. Theory and Decision 34:183–199
19. Ulvila JW (1990) Turning points: an analysis. In: McDonald JW, Bendahmane DB (eds) U.S. base rights overseas. Westview, Boulder, CO, USA
20. Ulvila JW, Snyder W (1980) Negotiation of international oil tanker standards: an application of multiattribute value theory. Operations Research 28:81–96

Part II

Models of International Negotiations

Nash's bargaining approach as the mother of all models of negotiations as well as the Stahl–Rubinstein bargaining theory have in common that they offer solutions to bargaining problems but do not really describe negotiation processes. In fact, both give advice on how to agree immediately. There are also descriptive models of real negotiations of the past or at present that provide insight into what has happened or what may be expected to happen. They may be interesting for researchers for various reasons, but practitioners may consider them user-unfriendly or even irrelevant.

Both types of approaches are represented in this part. Whereas the chapter by Gabbay deals with general concepts, in the subsequent chapters by Avenhaus and Krieger, Rudnianski and Bestougeff, and Güner, concrete international negotiations are analyzed with the help of various formal models—vector valued payoffs, incomplete information, and multiparty aspects are some of their ingredients. What their value might be to theoreticians and practitioners is discussed.

A Dynamical Systems Model of Small Group Decision Making*

Michael Gabbay

Information Systems Laboratories, Inc., Seattle, Washington, USA

1 Introduction

Key negotiations questions—whether to engage in or break off negotiations, what position to stake out, and whether to accept a deal—are crucial matters of state, and such decisions are often taken out of the hands of the bureaucracy and made by a small circle of top leaders. In this chapter, we present a model of small group decision-making dynamics that draws on social, cognitive, and political psychology. The model can address questions involving whether or not a decision-making group will reach consensus or deadlock in a given policy debate, the basic policy direction that will be chosen, and the stability of the decision in the face of new events and information. The model's foundation in social psychology and the use of a dynamical systems methodology represents an alternative perspective to most formal treatments of group decision making and negotiations that take a rational choice and game-theoretic approach. In addition, the social network over which group influence is exerted is a fundamental aspect of our approach.

The model, described in Part 2, is guided by social and cognitive psychology theories of attitude change [5] and small group dynamics [15, 26]. The model describes how each group member's position changes because of the interaction of the forces that arise when a member's current opinion is different from his own natural preference, the position of other group members, and the position supported by new information. Such forces are also central to the "decision units" analysis framework for foreign policy decision making in which the questions of how strongly an individual identifies with the group and the responsiveness to information are key contingencies [12]. The conceptualization of each member's position changing in response to "forces" naturally lends itself to the mathematical formalism of dynamical systems theory. A critical feature of the model is the presence of nonlinearity, which

* I am indebted to Michael Farrell for guiding me through the relevant social psychology and small group dynamics literature. I thank Margaret Hermann and Merle Lefkoff for valuable discussions and Daniel Druckman for helpful comments. I also thank William Zartman and Rudolf Avenhaus for extending me an invitation to participate in the PIN Program.

allows it to capture behaviors such as the sudden emergence of consensus or major schisms due to relatively small changes in conditions; path-dependent effects; the sharpening of factional lines due to an external threat; the greater facility of a group leader to bring sharply divergent factions closer together when acting as a broker; and the stability of equilibrium states of either consensus or deadlock with respect to perturbations by events and incoming information. These results are presented in Part 3.

Group processes are an underelucidated area of negotiations research and theory [22]. A group's internal dynamics affects how it sets goals, develops negotiating positions, seeks allies, and responds to actions by the other side. Understanding the impact of internal dynamics has also been identified as a prime question in need of further investigation with regard to ripeness theory, which addresses the conditions under which potentially successful negotiations may be initiated [27, 24]. Zartman [27] states that it is not sufficient for the parties in a conflict situation to be objectively in a mutually hurting stalemate in order for the situation to be ripe for negotiations; the parties must themselves perceive that this is so. In addition, they must have a sense of optimism that a negotiated solution is possible. These conditions stress the importance of subjective perceptions and attitudes and therefore call for a model of group decision making that has a basis in social psychology when addressing ripeness. Finally, group decision making in a situation of severe intragroup disagreement bears some kinship to negotiations and the results of our model of decision making within a party may serve as a guide to what a similar approach to modeling negotiations between parties would yield. In Part 4, we discuss the application of the model to international negotiations, focusing on its key predictions.

2 Model Description

In this part, we present a nonmathematical overview of the model. A fuller description of the model is given in [8] and a mathematical summary is contained in the *Appendix*. The model describes the evolution of group member positions for a single policy area, such as the state's strategic posture toward an adversary, including whether or not to pursue negotiations. The group member policy positions are arrayed along a one-dimensional continuum which we refer to as the *position spectrum*. This axis should correspond to the most salient dimension over which the decision makers' positions can be differentiated such as hawk versus dove. Although a one-dimensional representation is undoubtedly a considerable simplification of the complexities of a real-world decision-making situation, it can in fact be a parsimonious and useful one. For basic questions, such as whether to engage in a conflictual or cooperative strategy toward an opponent, it is more straightforward to tag decision makers along a hawk–dove continuum than to rate their positions on a multiplicity of individual issues. The hawk–dove spectrum is of particular relevance to negotiations, as the internal debate of most groups can be characterized as one between hawks who favor contentious tactics and doves who lean toward negotiating and problem solving [23]. The effect of this internal debate on negotiation outcomes has been experimentally investigated by Jacobson [17] who used a one-dimensional extremist–moderate scale to characterize his subjects.

We assume that a given group member's instantaneous position is subject to change under the influence of three separate forces: (i) the self-bias force, (ii) the group influence force, and (iii) the information flow force. The net effect of these forces determines the rate of change of a decision maker's current position.

2.1 Self-bias force

Each group member comes to the debate with an ingrained attitude that biases him or her toward a particular spot on the position spectrum independent of the arguments and pressures of the rest of the group. This bias position, which we refer to as the member's *natural preference*, is a reflection of the member's underlying beliefs, attitudes, and worldview of relevance to the matter at hand. For instance, a hawk would tend to advocate a tough response to a perceived threat or provocation, whereas a dove might call for a more conciliatory response. It is certainly possible for one's natural preference to change over time, but we will assume it to be fixed, which is often a reasonable assumption, especially over the course of a relatively short-lived decision-making episode. Although the natural preference is not evolved dynamically in the model, we can of course change it "manually" in order to accommodate rapid shifts in natural preference in response to major events of signal importance.

If a member's position is shifted from a natural preference because of group pressures or new information, she will experience a psychological force that resists this change, which we refer to as the *self-bias force*. In general, the further one's position is displaced away from the natural preference, the larger this force will be. The self-bias force can be viewed as a form of cognitive dissonance, which results from a divergence between the member's current position and that to which she is naturally inclined because of her innate worldview or bureaucratic role. It is simplest to assume that the self-bias force increases in proportion to the discrepancy from the natural preference (although more complex dependencies can be justified). For a given discrepancy, the strength of the self-bias force is determined by the member's *commitment*; the more committed a member is to her natural preference, the more difficult it becomes to shift her away from it. In the model, the two parameters associated with the self-bias force for a given member are the natural preference and commitment, where the term parameter refers to a quantity that characterizes some aspect of the group members but is not evolved by the endogenous dynamics of the model, although parameters may be allowed to vary exogenously. The group member's instantaneous positions are the dynamical variables that are evolved endogenously by the model.

2.2 Group influence force

The *group influence force* is the total force acting to change a given member's position through the influence of the other group members. It is taken to be the sum of all the forces resulting from that member's pairing with each of the other group members. In political decision-making groups, the relationship between any pair of members will typically be asymmetric, so that one member exerts a greater degree of influence on the other than vice versa. The amount of that influence depends on their relative status and their personal relationship. A member's formal status within the group depends on his official rank or position in the group hierarchy, the amount of resources he controls by virtue of his bureaucratic role, and his expertise in the policy area under debate [18]. But a member can also wield influence as a result of informal factors such as the respect he commands because of, for instance, a reputation for personal integrity. And, of course, a pair of members may have an informal relationship that can enhance or degrade their mutual influence such as friendship, common factional membership, or personal animus.

The force that a member feels through the persuasive efforts of a second member, which we call the *coupling force* so as to distinguish it from the total group influence force, is a function of the discrepancy between their positions. The form of the coupling force is shown in *Figure 1*, which plots the force on a member, the "receiver," due to the persuasive messages sent by another member, the "sender." If the discrepancy is relatively small, the force changes linearly with discrepancy and a maximum force is then reached, after which the force tails off, asymptoting to zero for very high discrepancies. The basic shape of the curve is similar to that posited in social judgment theory (see [16, 5]) in which there is a zone around the receiver's position called the "latitude of acceptance," for which the amount of attitude change increases as the position difference increases because the messages are perceived as fair and reasonable. Beyond a certain discrepancy, however, the amount of attitude change decreases with rising discrepancy, as the messages are increasingly seen as distorted and unreasonable—this zone is the "latitude of rejection." Social judgment theory therefore predicts a nonlinear dependence for the amount of attitude change as a function of the message discrepancy, as shown in *Figure 1.*

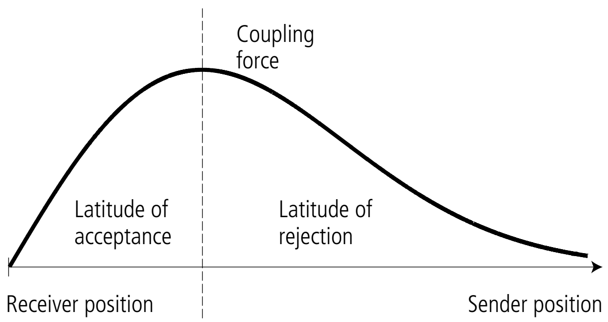

Fig. 1. Coupling force on receiver as function of distance from sender's position, showing nonlinear dependence

There are two basic model parameters that characterize the coupling force which member j, acting as the sender, exerts on member i, acting as the receiver: i's *latitude of acceptance* and the *coupling strength* of j on i. The latitude of acceptance scales the width of the coupling force curve as can be seen from *Figure 1* and corresponds to the discrepancy at which the coupling force reaches its peak. The latitude of acceptance, when considered as a unit of length along the position spectrum, can be used as a gauge for the level of disagreement; if the distance between two members is well within a latitude of acceptance, they are essentially in agreement, whereas if their positions are roughly three or more latitudes of acceptance apart, then a deep rift exists between them. For the purposes of interpreting the model results, we will typically assume that a rough consensus exists if all members are within one latitude of acceptance from each other. For convenience, we assume that the latitude of acceptance is the same for all group members.

The coupling strength scales the height of the coupling curve, so that for a fixed discrepancy a larger coupling strength implies a larger force. There is a coupling strength value for each directional pair (i.e., sender j's persuasion on receiver i, which depends on the particulars of their relationship as noted above). Folded into the coupling strength parameter is another parameter, called the *group coupling scale*, that is the same for the whole group and that assesses the overall pressure for the group to agree. In other words, it weighs the importance of the group influence force relative to the self-bias and information flow forces. The group coupling scale is dependent on situational variables. We would expect, for instance, that as the level of external threat to the group increases, the pressure to unite and reach consensus grows and so the group coupling scale increases. Another factor that could increase the group coupling scale is the presence of a short time limit in which to make a decision. In addition, the group coupling scale depends on the rate of communications between members. The group influence force can also model the effects of group norms, such as the desirability of appearing hawkish, by incorporating an offset parameter in the coupling force, which models the tendency of group members to strive to outdo each other in the direction of the norm [8].

2.3 Information flow force

The self-bias and group influence forces account for the dynamics that arise from the structure of preference distributions and influence relationships, but the group will also be exposed to information that becomes known to one or more members during the decision-making process. Incoming information can affect a person's current opinion without causing a permanent change in his or her underlying attitudes as expressed by the natural preference. This immediate impact of information upon a group member's instantaneous position is modeled by the *information flow force*, which enables the model to capture the more transitory back-and-forth effects that occur as a result of an incoming stream of ambiguous information. Each piece of information is attributed by a member as supporting some particular position along the position spectrum known as the *message position*, which need not be the same for all members. The information flow force on a given member due to a given message can be taken to be a function of the discrepancy between the message position and the member's current position analogous to what is done for the group influence force, although it need not have the same functional form.

The key parameter in the information flow force is the *contextual sensitivity* which gauges how responsive a group member is to information from the environment [11, 13]. A more contextually sensitive member will change his position in response to the vicissitudes of events than a contextually insensitive one who will maintain his position regardless of changing circumstances. The contextual sensitivity can also be dependent on situational factors that characterize the channels through which decision makers receive information. In order to investigate the effects of information in a generic way, it is useful to treat the information flow as a stochastic process where we assume that each member will be exposed to a stream of random incoming information that will cause his opinion to wander back and forth. Members who are more contextually sensitive will bounce around more than less sensitive ones. This approach can be used to address the stability of states corresponding to group consensus or dissensus; states that are stable equilibria in the absence of information flow may not be so when subject to random perturbations from incoming information, as we shall see in Part 3.

2.4 Model summary

Here we summarize the main features of the model as a set of qualitative hypotheses or "rules." These rules are mathematically implemented using coupled nonlinear differential equations as described in the *Appendix*.

1. Each group member's *position* is affected by three forces: the self-bias force, the group influence force, and the information flow force.
2. The self-bias force is the tension that the member feels when his current position is different from his *natural preference*:
   a) It acts in the direction of the natural preference.
   b) Its magnitude is proportional to the difference between the member's current position and natural preference.
   c) Its strength depends on the member's *commitment* to his natural preference.
3. The group influence force is the tension that a member feels when his position is different from that of others in the group:
   a) The total group influence force on a member is the sum of the coupling forces resulting from his pairings with all the other members.
   b) For a given pairing, the coupling force on a member acts in the direction of the other member's position.
   c) The coupling force magnitude increases approximately linearly for small position differences but weakens for differences greater than the member's *latitude of acceptance*.
   d) The coupling force that member j exerts on member i is scaled by the *coupling strength* which characterizes factors such as how often j communicates with i on the policy matter at hand, their relative status, i's perception of j's credibility, and the importance that i attaches to group influence.

4. The information flow force is the tension on a member when a message representing a new piece of information supports a position that is different from the member's current position:
   a) It acts in the direction of the *message position*.
   b) It increases with increasing discrepancy and weakens for large discrepancies (if taken to be of the same form as the coupling force).
   c) Its strength depends on the member's *contextual sensitivity*.
5. Events can change the structural properties of the group as determined by exogenous factors. Examples include:
   a) A heightened external threat increases the coupling strength between members.
   b) A policy failure leads to a loss of status of its advocates, decreasing their coupling strengths on others.
   c) A dramatic event causes an immediate shift in natural preferences.
   d) A change in group membership.

3 Survey of Model Dynamics

In this section, we illustrate the model dynamics using selected examples that show that the model produces intuitively reasonable results and interesting insights into group decision making. The centrality of bifurcation as the transition between consensus and deadlock is shown using a simple two-person group. Bifurcations typically imply that the system can undergo a large change for a small change in parameters, which in a political setting can lead to a sudden, dramatic, and perhaps surprising, transition. Furthermore, the particular type of bifurcation that our model exhibits provides for the existence of path-dependent effects and for the instability of decisions to incoming information.

3.1 Two-person model results

Bifurcation phenomena

We use a group of two members in which the members have equal commitments; the mutual influence between them is symmetric, so that the coupling strengths between them are the same; and their latitudes of acceptance are the same. We set the commitment of each member to one. We also set their latitudes of acceptance to one. This implies that the distance between their positions, and hence the level of disagreement, is reckoned in units of latitude of acceptance. *Figure 2* shows the behavior of the two-man group as a function of the coupling strength in the absence of the information flow force. We take the natural preference of member 1 to be -1.5 and that of member 2 to be 1.5, so that the difference between their natural preferences is 3 which represents a wide gulf of disagreement that, if unresolved, would not allow a common decision to be reached. If we were to take the position spectrum as representing a hawk–dove continuum with hawks falling on the negative side and doves on the positive, then member 1 would be considered to have a solidly hawkish attitude and member 2 would be decidedly dovish. In *Figures 2(a)*, *(b)*, and *(c)*, the initial positions are taken so that each each member starts out the discussions at time $t = 0$ at their respective natural preferences. *Figure 2(a)* shows the case of relatively weak coupling. It is seen that the member positions barely budge and so they remain deadlocked. In *Figure 2(b)* the coupling is strong, enabling a consensus to be forged—the final positions are less than 0.5 units apart. For an intermediate value of the coupling strength, *Figure 2(c)* shows that the coupling force is still insufficient to overcome the initial disagreement. If, however, for historical reasons, the members start out in or near consensus rather than at their natural preferences then this level of coupling is sufficient to maintain the consensus, as is seen in *Figure 2(d)*.

The difference between *Figures 2(c)* and *(d)* is an instance of *bistability*, where there are two stable equilibrium states of the system and the one that results depends on the initial conditions. In cases where a decision on a new policy or a response to a crisis situation is under consideration, it is appropriate to assume that the members start discussions at positions correlating with their natural preferences. In situations where an existing policy is under consideration, then the members have preexisting positions which result from previous debate and decisions and need not be the same as their natural preferences.

Path dependence

The above results reflect the bifurcation structure of the system. The transition from the case where the group deadlocks to that where it reaches consensus does not occur smoothly as the coupling strength is increased. Rather there is a sharp transition to consensus once a certain critical value is exceeded. This can be seen in the upper curve in *Figure 3*. Here the final difference in the member positions, which we call the *discord*, is plotted as a function of the coupling strength as it is increased. For each value of the coupling strength, the system is evolved for an amount of time typically sufficient to reach an equilibrium state; this final state is then used as the initial condition when the coupling is incremented to a higher value. Starting from low coupling strengths on the left-hand side, we see that the deadlock condition remains until a coupling strength of about 2.2 is exceeded, after which there is a sharp transition to the consensus state. The lower curve shows what happens as the coupling strength is reduced from the high values at the right-hand side of the plot: the system remains in a consensus state until the strength drops below 1.5, resulting in an abrupt shift to the deadlock state. It is evident from the figure that in the zone between coupling strengths of 1.5 and 2.2, both the deadlock and consensus states are possible which corresponds to the bistability seen in *Figure 2* plots *(c)* and *(d)*. For any location in this zone, all the parameters of the system—the commitment, natural preferences, coupling strength, and latitude of acceptance—are the same and yet there are two possible outcomes, consensus and deadlock. The outcome that results depends on the group's immediate past history in terms of its members' positions. The system therefore displays a type of path dependence that naturally arises from the dynamics. Specifically, it stems from the particular type of bifurcation that this

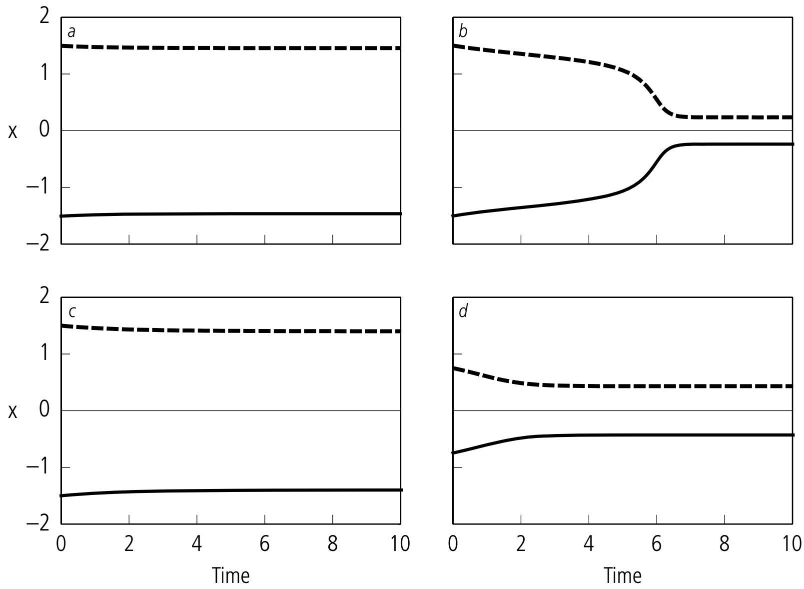

Fig. 2. Symmetric two-person group examples showing effect of coupling strength and initial conditions. *(a)–(c)* are for natural preference initial conditions: *(a)* weak coupling ($k = 1$); *(b)* strong coupling ($k = 3$); *(c)* intermediate coupling ($k = 1.8$). *(d)* near consensus initial conditions for intermediate coupling ($k = 1.8$). Parameter values (see *Appendix* for notation; indices are omitted if a given parameter has the same value for all group members): $c = 1,\ p_1 = -1.5,\ p_2 = 1.5,\ l = 1$

system undergoes, known as a *saddle-node bifurcation*. Physical systems marked by saddle-node bifurcations display the memory effect shown in *Figure 3* and known as *hysteresis*.

Effect of information on decision stability

The two-person results in the preceding section did not include the possible effects of incoming information upon the group deliberations. In this part we allow for the impact of such information. We will treat it as a stochastic input that causes a member's position to bounce around randomly. This enables us to address questions of decision stability, that is, the robustness of a group consensus, given that there will always be new information coming in that will cause members' opinions to fluctuate. These fluctuations may be small but they can, in some circumstances, shatter a consensus. The issue of decision stability is significant as it bears on how zealously a decision will be pursued and the chance for its reversal.

Figure 4 shows cases for a two-person group where each person is subjected to a random information flow force in addition to the self-bias and group influence forces. Each member is exposed to an independent information flow. The degree to which a

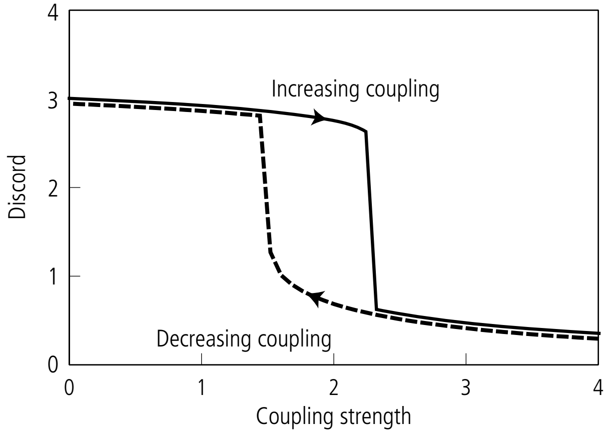

Fig. 3. Discord in symmetric two-person group exhibits hysteresis as coupling strength is varied. Parameter values: $c = 1,\ p_1 = -1.5,\ p_2 = 1.5,\ l = 1$

group member's position changes in response to the information is given by his or her contextual sensitivity as discussed in *Part 2.3*. In *Figure 4(a)*, the group starts off in a rough consensus state since they are about one latitude of acceptance apart. Both members have the same contextual sensitivity and, although their positions bounce around from the equilibrium values, the consensus endures. In *Figure 4(b)*, the contextual sensitivity of the dovish member is now quadrupled, which results in the larger fluctuations in his position as compared with the hawkish member. We see that at time $t = 12$, an open rift develops as the consensus breaks down and the member's positions return to essentially their natural preferences, which are 1.5 and -1.5. In *Figure 4(c)*, the members again have equally low contextual sensitivities, but they start out in the deadlock state and there they remain. The same holds true for *Figure 4(d)* where the dovish member's sensitivity is quadrupled. Thus, in this situation, the consensus state proves more fragile than the deadlock state.

This potential for decision instability in the face of events requires that, in the deterministic case (no information flow), the system resides in the bistable zone where both the consensus and deadlock states exist for the same parameter values (technically, this instability is with respect to finite not infinitesimal perturbations). Since the amount of jitter in one's position is gauged by one's contextual sensitivity, we would expect that the instability of a decision state as measured via a transition probability would increase as the contextual sensitivity is increased. This is indeed the case as is shown in *Figure 5*. Note that in this example, for any given contextual sensitivity, there is a substantially greater probability of the consensus state flying into a rift than for the reverse transition. At high sensitivity values, there is a significant chance of either transition, which indicates that the group would bounce around between the consensus and deadlock states.

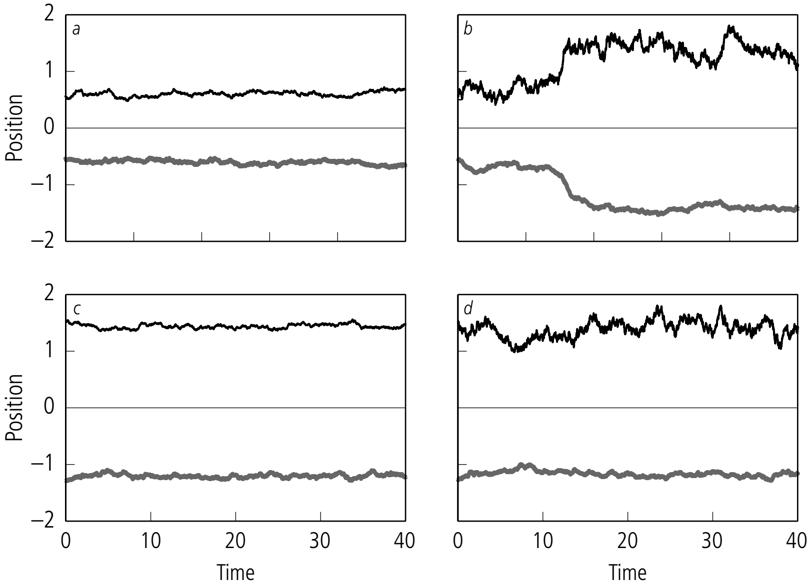

Fig. 4. Effect of stochastic information flow on symmetric two-person group: *(a)* equal contextual sensitivities, ($s_1 = s_2 = 0.05$), consensus initial conditions; *(b)* decision instability results when the contextual sensitivity of the dovish member (upper curve) is raised ($s_2 = 0.2$); *(c)* same contextual sensitivities as *(a)* but with deadlock initial conditions; *(d)* same contextual sensitivities as *(b)* with deadlock initial conditions. Parameter values: $c = 1,\ p_1 = -1.5,\ p_2 = 1.5,\ k = 1.55,\ l = 1$.

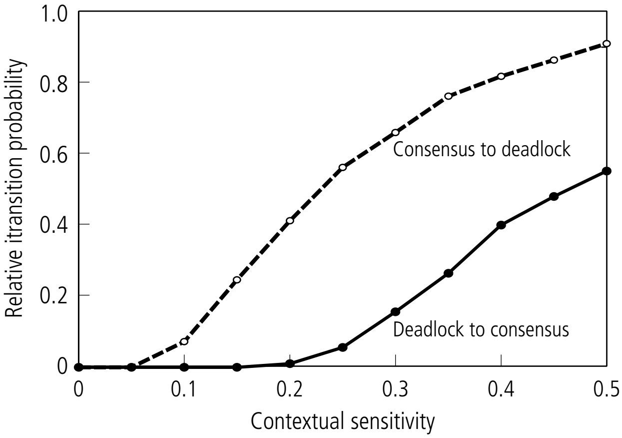

Fig. 5. Dependence of probability of transition between consensus and deadlock states on contextual sensitivity for two-person group. Parameter values: $c = 1,\ p_1 = -1.5,\ p_2 = 1.5,\ k = 1.55,\ l = 1$. Probabilities were obtained at each point based on 500 runs of duration 10 time units

3.2 Group structure effects

In this part, we will concern ourselves with the consequences of the group structure, where the term structure refers to the coupling network between group members and the distribution of members' natural preferences. There are two basic types of group network that we consider here: the *all-to-all* network, in which every group member can directly influence every other member (not necessarily equally); and the *broker* network, in which one key member, typically the group leader, acts as a conduit of influence between two opposing factions. The all-to-all network is appropriate when there is a significant degree of direct communication between each pair of group members; for instance, if they all convene in common meetings to discuss the policy under consideration and each has at least some chance to express his or her views. The broker network is relevant when there is a group member who links two or more sides or bureaucracies either by virtue of the hat he wears under the formal political structure of the group or deliberately via manipulation and control of information and access. The broker network also applies when the factions are sufficiently at odds so as to choose not to communicate with each other. The presence or absence of a broker is a key question in the decision units analysis framework of small group decision making [25, 10].

All-to-all network results

Figure 6 shows a six-person group that is homogeneous except for the member preferences, which break down into two diffuse camps of hawks and doves. All the members are coupled to each other equally. The figure shows that strong enough coupling will produce a consensus as in the two-person group above.

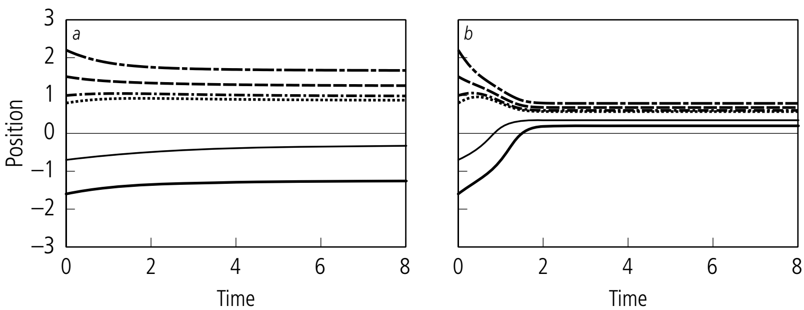

Fig. 6. Examples of six-person group with symmetric all-to-all coupling: *(a)* weaker coupling ($k = 0.3$) results in deadlock; *(b)* stronger coupling ($k = 1$) leads to consensus

The above example shows that for strong enough coupling, the group reaches consensus. This might lead one to believe that an increase in the coupling will always drive the group closer together. The model, however, shows an interesting effect in which an increase in overall group coupling need not make the factions come appreciably closer; rather, the factions unify individually, thereby producing the counterintuitive effect of accentuating the inter-faction rift. *Figure 7* shows simulations of a group consisting of four hawks and two doves. The coupling strengths between all the members are the same with a value of 0.5 for *Figure 7(a)* and 2.0 for *Figure 7(b)*. It can be seen that the higher coupling strength tightens each faction but does not close the gap between the factions and that the most moderate elements of each faction are now farther apart.

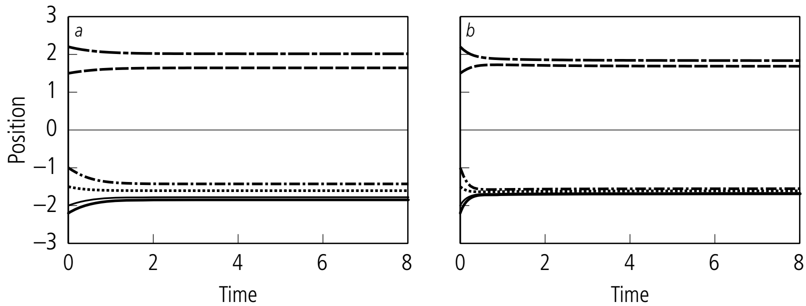

Fig. 7. Effect of increasing coupling strength for a homogeneously-coupled six person group consisting of four hawks and two doves: *(a)* weaker coupling strength ($k = 0.5$); *(b)* stronger coupling strength ($k = 2.0$)

One can quantify this effect using the concept of self-categorization from social categorization theory [14, 19]. Self-categorization theory states that the members of the in-group (one of the factions in our case) seek to conform most closely to the in-group mean while maximizing the distance from the out-group (the opposing faction). We define a metric for self-categorization as the ratio of the (absolute) difference between the in-group and out-group mean positions to the in-group discord, the latter being the difference between the most extreme members of the faction. So the self-categorization of a given faction measures how far apart the two factions are relative to the spread of positions within a faction (it can also be viewed as an index of ethnocentric bias). *Figure 8(a)* shows the dependence of the self-categorization of the hawks and doves of our six-member group upon the coupling strength. It increases as the coupling strength increases, which implies that the factional lines are becoming more clearly drawn rather than bringing the groups closer, as one might intuitively expect. The separation between the factional means, shown in *Figure 8(b)*, barely decreases until a critical value of about 3.2 is reached, when consensus emerges and the self-categorization drops rapidly (after which it becomes constant and is no longer a meaningful measure once the group has reached consensus and resides in the linear regime of the coupling force).

The above dynamic points to an important effect which could arise in conflict situations: if an external party were considering putting pressure on one of the sides of

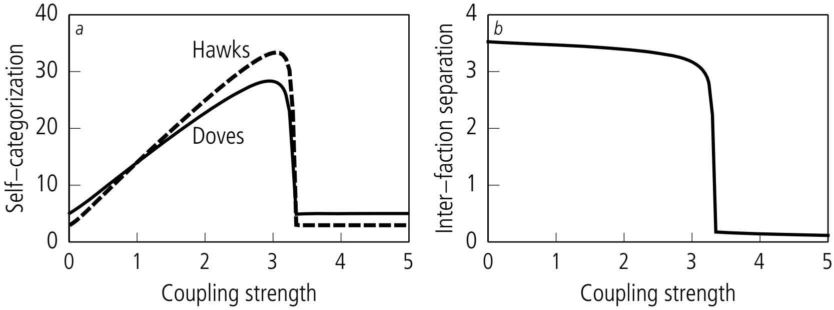

Fig. 8. (a) Effect of increasing coupling strength on factional unity as measured by the self-categorization index for doves (solid) and hawks (dashed). (b) Separation between the mean positions of the factions

a conflict in order to bring them to the negotiating table, it is possible that such pressure, if not sufficiently strong, would heighten factional divisions within that side's government, thereby inhibiting, rather than facilitating, its ability to make a decision to negotiate. As the coupling strength is dependent on the rate of communications between members, we would also expect an increase in communications to result in similar behavior. It should be noted that this effect is due to the presence of nonlinearity in the model and would not have been predicted by a linear model. This effect will be present only for very strong levels of disagreement between factions; for low to moderate disagreement, the standard intuition that increasing pressure should mitigate factional divisions holds.

Comparison of all-to-all and broker networks

An important aspect of our modeling approach is its ability to address effects that arise from the influence network structure. *Figure 9* plots the group discord as a function of the leader status and leader natural preference for the all-to-all and broker coupling networks. The five-member group consists of hawk and dove factions of two members apiece and a leader. The natural preferences of the factions are far apart with the most "moderate" members of each separated by three latitudes of acceptance. Each faction member is assigned a status of one but the leader's status and natural preference are variable (recall that the coupling strength from one member to another is a function of their relative status). In the all-to-all network, all members are assumed to be connected with each other. In the broker network, each faction member is connected to his fellow faction member and to the leader, who plays the broker role. The light region in the figure represents a high discord situation in which the factions do not overcome their sharp divisions and deadlock pertains. The dark region is a zone where the factions are either in consensus or have moved considerably nearer to it. We see in both cases that if the leader's status is low enough then the group will deadlock regardless of the leader's natural preference. In the all-to-all case, for example, a leader whose status is one (i.e., equal to the other group members), would not be able to appreciably close the gap between the factions. We observe that the dark, lower discord region is larger for the broker network; it extends to lower status levels and is wider for a given status. This implies that the broker network provides a more conducive structure for a leader to bring the factions toward consensus than the all-to-all network. In the latter, the leader is a big fish in a big pond, whereas in the former, he is a big fish in two smaller ponds. Although this makes intuitive sense, it may not have been expected from initial qualitative intuition alone, as it also could have been argued that, in the all-to-all network, there would be more pressure to agree on each side due to the presence of the opposing faction. Indeed, in a linear model, the all-to-all network would always be more effective at moving the factions closer and, in our model, this is the case for low to moderate inter-faction differences. But for severe inter-faction disagreement—such that the tail of the coupling function in *Figure 1* is in play—the broker network more readily facilitates consensus. Our model, therefore, provides a criterion as to when to expect each type of network to be more effective, based on the level of initial disagreement.

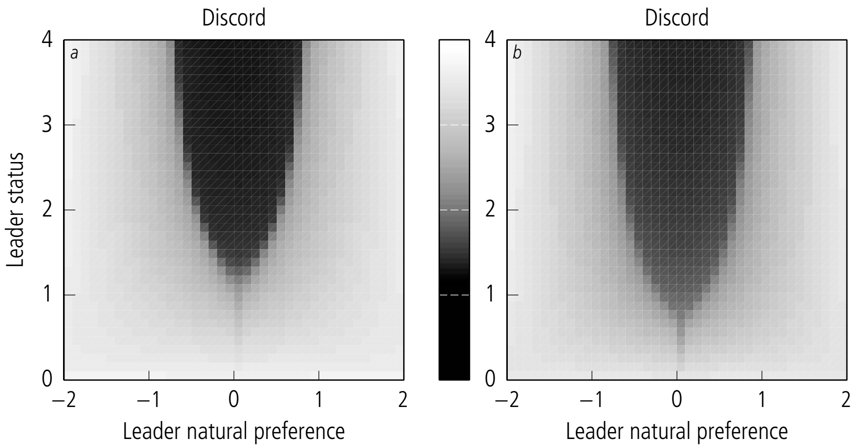

Fig. 9. Discord as a function of leader natural preference and status for a group consisting of two factions of two members each; *(a)* all-to-all network; *(b)* broker network. All group members have equal status of one except for the leader. The faction member natural preferences are ± 1.5 and ± 2 for the two doves (+) and two hawks (−)

4 Discussion

4.1 General remarks

The model presented above combines social psychology theories of attitude change with a mathematical formalism based on dynamical systems theory. It displays a range of interesting behaviors such as sharp transitions between deadlock and consensus; hysteresis, which is a type of path dependence; decision instability in the face of incoming information; the possible heightening of factional divisions as an external threat or pressure is increased; and the enhanced ability of a leader to close the gap between factions in a broker network topology. The nonlinear dynamical phenomenon of bifurcation arises naturally in the model and helps to enable predictions of a qualitative nature. Results that stem from the bifurcation structure are more likely to be robust than are those that depend on the specific form of the model. This is so because all systems that exhibit a certain type of bifurcation will display similar behaviors that do not depend on the exact mathematical form of the system.

In terms of its relationship with previous work, our model bears kinship with social influence network theory ([6], [7]) and dynamic social impact theory [21]; the former for its emphasis on opinion change over a social network and the latter for its emphasis on nonlinear dynamics. Our emphasis on bifurcation phenomena is likely most familiar to formal modelers from catastrophe theory [20]. Given that the application of catastrophe theory to social science has a controversial past, it is worth noting that the bifurcations in our model stem from plausible choices of individual-level attitude change mechanisms rather than via an a priori assumption of a particular type of bifurcation or catastrophe. Consequently, we do not need to justify our model by appealing to a broader paradigm that claims that diverse swaths of social phenomena can be categorized into a relatively small number of catastrophes.

The model is implemented using differential equations and is essentially deterministic. In these respects, it is similar to Richardson-type arms race models [20], which are often thought to be primarily of descriptive rather than predictive value. Although our model does provide descriptive insight into group decision making, it also has value for predictive applications in which policymakers and analysts seek to evaluate ongoing real-world conflict or negotiations situations. This is the case because there are distinct regimes of qualitatively different behavior in the model and predictions can be made on the basis of these regimes rather than on a precise quantitative implementation. Loosely speaking, these regimes are "linear" and "nonlinear," corresponding, respectively, to low/moderate and high levels of disagreement. Thus, the model can provide heuristic guidance about what qualitative dynamics to expect, and its predictive capacity is not restricted to purely calculating the value of some abstract quantity.

As this is a new approach to modeling group decision making, it has not been subject to a systematic validation effort. An implementation scheme involving an objective, empirical fitting of the model parameters from data would not, in general, be possible in real-world political decision-making settings. However, a scheme whereby expert observers make subjective estimates of model quantities is possible. In the case of the position spectrum, intuitive terms for the level of disagreement such as moderate, substantial, and severe could be coded in terms of units of latitudes of acceptance. The development of such a systematic procedure would enable model validation on historical case studies. Laboratory validation of the basic qualitative phenomena predicted by the model is feasible. Efforts toward developing techniques for empirically estimating certain model quantities would be helpful, prime among these being the position spectrum and the latitude of acceptance. This would enable better identification of the level of discord and, consequently, the behavioral regime that characterizes the group. In general, the parameters of the model would benefit from further elaboration on how they depend on factors both endogenous and exogenous to the group.

4.2 Implications for negotiations

There are two ways in which our model could be applicable to the theory and practice of international negotiations: the first is in its direct capacity as a model of small group decision making within a given party; the second is by drawing upon the similarity of decision-making situations marked by sharp factional disagreements to high-conflict negotiation situations.

In its direct application to a leadership group, the model could be used to assess the direction, the level of consensus, and the stability of the basic policy that the group would pursue given the mix of hawks, doves, and moderates and their associated influence network. The model is applicable as long as the broad policy options under debate lie roughly along a conflictual/cooperative spectrum that can be correlated with the hawkish or dovish attitudes of each of the group members. For example, the position spectrum can correspond to the strategic options that affect the fundamental intensity of the conflict ranging through levels of escalation, status quo, and conciliation. This would be the appropriate scale over which to look at the decision to seek negotiations, as such a move, if sincere, represents a core strategic choice. Once negotiations have started, the decision making surrounding the choice of a negotiating position and the reaction to the opposing party's positions and actions can be addressed by the model. Of course, the group structure itself could change and the model is capable of investigating the impacts of such changes, including changes in group membership, the loss of status of key members, or a sudden shift in the natural preference of a group member.

Consideration of the model dynamics can provide heuristic guidance as to how internal dynamics can play out in the negotiations process. As an example, Pruitt and Kim [23] suggest that it is of greater avail for a side to reveal its internal hawk–dove debate in negotiations rather than present a unified front, and there is experimental evidence in support of this contention [17]. In addition, they state that this effect is greatest when the hawk and dove factions are about equal in power. Our model could investigate how the dynamics of a given side's decision-making elites, the "receiving group," differ, depending on whether they receive independent streams of messages coming from the other side's hawks and doves or whether they are presented with an average position coming from a unified front. For instance, consider a receiving group that is currently in a rough consensus of its own hawks and doves but falls in the bistable regime where the deadlock state is also an equilibrium solution. Our model allows for a dynamic in which two incoming streams of competing hawk and dove messages from the other side could in fact rupture this rough consensus, preventing the receiving group from reaching decisions regarding negotiations, whereas a consistently moderate incoming message would preserve the consensus (this assumes a nonlinear form of the information flow force analogous to that of the group influence force). On the other hand, if the receiving group's hawks and doves were sufficiently united, this rift would not then be expected to develop. Thus, the model points out that the strategy of revealing hawk–dove differences may be more appropriate when the receiving group can be approximated as a unitary actor, but could backfire if the receiving group is characterized by sharp divergences in natural preferences that could undermine the consensus for negotiations.

The model can also benefit the analysis of conflict situations in which ripeness is of concern by enhancing the account of internal dynamics within ripeness theory where again the hawk–dove debate is of central importance [27]. For instance, the model could help identify circumstances under which increasing pain or pressure from a third party might hinder the ability of a group to decide to negotiate rather than encourage it as would be expected. This could occur if the factions on a given side were in sharp disagreement—separated by three latitudes of acceptance or so—and an increase in pain or external pressure would bring the intragroup conflict into sharper relief as in *Figure 8*, impairing their ability to reach a consensus around negotiating. This would represent an additional type of the "resistant reactions" discussed by Zartman. On the other hand, if the factions were about a latitude of acceptance or so from each other, then an increase in pain or pressure would have the expected effect of drawing them closer. Here we see how qualitative considerations of the level of dissensus within the group can lead one to expect different reactions; if the level of disagreement is high, the group is in the zone where the nonlinear effects dominate and one would be alerted to the possibility of counterintuitive reactions whereas for more moderate disagreement, the group would be expected to behave in accord with standard "linear" intuition.

Given that our approach is based on social psychology theories of attitude change and persuasion and explicitly accounts for network structure, it is well suited to be adapted to the network model of negotiations in which the links between the various actors involved in the negotiations are considered [22]. One possible result that may carry over from the decision-making model is that discussed in connection with *Figure 9* in which a leader in a broker network is better able to bring opposing factions closer than in the all-to-all network where the sides can communicate directly. This finds analogy in negotiations where it is believed, counter to initial intuition, that it is better for a mediator to caucus separately at first with the two parties in high conflict situations rather than allow the sides to talk directly [23].

A nonlinear dynamical systems approach may shed light on ripeness theory in terms of the relationship between the two sides of the conflict. Ripeness arises when the parties go from a situation in which they each seek unilateral solutions to one in which they perceive themselves to be in a hurting stalemate and view a negotiated solution as a possible way out of the conflict [27]. The perception of a stalemate need not imply that hopes of unilateral solutions on both sides have been abandoned, however. If this were so, then a negotiated solution would be the only one possible, and agreement would be assured. Thus, we can view the transition to ripeness as a change in the outlooks of the parties from one in which only unilateral solutions are sought to one in which both unilateral and bilateral solutions are possibilities. A qualitative change in the nature of the solution space is the definition of a bifurcation, suggesting that we regard the transition to ripeness as such. If we can identify the type of bifurcation, then we can deduce the associated observable behaviors.

We can use the results of our decision-making model as a rough guide to the implications of the above bifurcation hypothesis. Given the change from a single stable solution to a bistable condition, we speculate that the onset of ripeness in a conflict situation corresponds to a saddle-node bifurcation. In *Figure 3*, this is analogous to the transition from the high discord-only state to the state where both high- and low-discord situations are stable solutions. To pursue the analogy further, bifurcations occur as a parameter is varied, and in the negotiations case we can likewise speak of the parties as being coupled to each other, the degree of coupling being parameterized by the coupling strength. The coupling strength could depend on a number of factors but a key factor is the level of pain that the conflict is inflicting on the sides [27]. So we can think of an increasing level of pain as increasing the coupling strength between the sides, eventually resulting in a situation ripe for negotiations. This notion of coupling between sides is in accord with the idea that the perception of stalemate results in acceptance of the interdependence of the parties [23]. It should be noted that, as in group decision making, the parties need not be coupled together with the same strength as assumed in the symmetric case of *Figure 3*. Asymmetries could result from differences in relative power or pain. In addition, there are other factors besides pain that would increase the coupling strength such as the amount of third-party pressure or the level of trust between the parties.

The level of discord between the parties can be related to the nature of the tactics they employ. Pruitt and Kim [23] state that there are two limiting forms of negotiation—contentious tactics and problem solving. If negotiations have been initiated then we associate the high discord state with contentious behavior and the low-discord state with problem-solving behavior. In the high-discord state, the parties are several latitudes of acceptance apart and have little effective influence on each other. The high-discord state can also correspond to a high-conflict condition in which negotiations have not been started.

With respect to *Figure 3*, the high-discord state is the only possible one at low-coupling strengths, as the sides are committed to seeking unilateral solutions. Negotiations may be initiated here but they would not reflect an underlying sincerity to resolve the conflict bilaterally and would fail. As the pain level increases, the bifurcation to ripeness occurs and the low-discord state appears as a possible solution. If negotiations have begun, the parties will remain in the contentious, high-discord state until new information of sufficient significance shifts them to the low-discord, problem-solving state. The amount of this requisite "push" decreases as one moves deeper into the ripeness zone. For sufficiently large coupling strengths, only a small push would be needed, corresponding to Goodby's [9] notion of ripeness as a metastable condition in which a small external force exerted by a third party can yield major changes in the situation. Druckman [4] has found that external influences such as third-party interventions often produce abrupt changes in security negotiations.

The central prediction implied by the bifurcation hypothesis is that there is a bright line between the contentious and problem-solving forms of behavior in high-conflict situations; one would not in general expect to observe a gradual transition between these two extremes. This claim could be validated by laboratory or case-study analysis. The “turning points” framework for analyzing negotiation processes should be useful in investigating this prediction, which may be related to the finding that security talks are significantly more likely to be marked by abrupt process departures than are political and trade talks [4]. Anecdotal evidence from professional mediators suggests that rapid shifts from contentious to collaborative behavior do indeed occur between teams of negotiators (Lefkoff, personal communication). The resolution of elite crises by agreement between rival elite factions has also been noted to occur very quickly [1]. Further support for the claim that high-conflict situations have a binary “on–off” type character comes from the analysis of enduring international rivalries. Diehl and Goertz [2] find that when such rivalries are examined over a long period rather than in the context of specific crises, they are apt to exhibit a quick lock-in of the conflict, a prolonged period of high conflict (which can vary around a baseline level), and a rapid de-escalation rather than a smooth, gradual change.

A secondary prediction stems from the nonlinear dynamical phenomenon in which perturbations to the system take longer to decay in the vicinity of a bifurcation than when the system is far from the bifurcation point. As one approaches the ripeness transition from the unripe zone, this might manifest itself by evidence of more deliberation of proposed negotiation initiatives rather than outright rejection. Lastly, it is important to recall that these predictions of the ripeness as bifurcation hypothesis are motivated by analogy with the decision-making model and to put them on firmer ground requires the formulation of a specific model of conflict and negotiations.

In conclusion, the social psychology of negotiations and conflict is a rich field that encompasses a diverse range of phenomena [3, 23] and merits much greater attention from formal modelers than it has been afforded to date. Of course, our model requires further development in order to systematically apply it to actual cases. It is, however, of more immediate value as a heuristic tool for understanding small group decision making. As it provides a different perspective from that of game theory, its use as simply a conceptual framework may help both researchers and practitioners gain insight into negotiations dynamics that are not readily addressed by a traditional game-theoretic framework.

Appendix: Mathematical summary of the model

The basic elements upon which the model operates are:

- The group members whose positions we take into account; each group member is labeled by an index $i = 1, \ldots, N$
- The position spectrum over which member opinions can range; the *instantaneous position* of member i at time t is $x_i(t)$

The instantaneous position of each member is acted on by three basic forces:

- The *self-bias force*, $S_i(x_i)$
- The *group influence force*, $G_i(x_i)$
- The *information flow force*, $I_i(x_i)$

The basic elements of the self-bias force are:

- The *natural preference*, p_i, which expresses the i^{th} member's ideological or bureaucratic predisposition
- The member's *commitment*, c_i, to his natural preference

The key elements of the group influence force are:

- The *coupling force*, $H_{ij}(x_j - x_i)$, which is the influence force that member j exerts on i because of the discrepancy between their positions
- The *coupling strength*, k_{ij}, which expresses the strength of the influence of j upon i due to their relationship; it scales the height of the coupling curve
- The *latitude of acceptance*, l_i, which corresponds to the discrepancy beyond which the impact of an argument or message on i starts to wane and the coupling force begins to tail off

The key quantities appearing in the information flow force are:

- The *contextual sensitivity*, s_i, which gauges the member's responsiveness to incoming information
- The *message position*, m_i, which is the position supported by an incoming message or event as interpreted by the i^{th} member
- The *message weight*, $w_i(t)$, which member i accords to the message at time t
- The message *valence function*, $v_i(m_i - x_i)$, which describes how the impact of the message varies with its discrepancy from the member's position

In terms of the three forces, the dynamical equation that governs the rate of change of the i^{th} member's current position is written as:

$$\frac{dx_i}{dt} = S_i(x_i) + G_i(x_i) + I_i(x_i). \tag{1}$$

The group influence force is the sum of the pairwise coupling forces between members. The information flow force is the sum of the forces due to each message, m_i^n, $I_i(m_i^n, x_i)$. Equation (1) then becomes

$$\frac{dx_i}{dt} = S_i(x_i) + \sum_{j=1}^{N} H_{ij}(x_j - x_i) + \sum_{n} I_i(m_i^n, x_i). \tag{2}$$

The self-bias force is taken to be of linear form:

$$S_i(x_i) = -c_i(x_i - p_i). \tag{3}$$

The functional form of the coupling force corresponding to the nonlinear shape of *Figure 1* is

$$H_{ij}(x_j - x_i) = k_{ij}(x_j - x_i) \exp\left(-\frac{(x_j - x_i)^2}{2l_i^2}\right). \tag{4}$$

The information flow force due to a single message is

$$I_i(m_i, x_i) = s_i w_i(t) v_i(m_i - x_i). \tag{5}$$

The valence function can have the same form as the coupling force although there may be contexts in which other forms are merited.

Using Equations (3)–(5) in Equation (2), the equation of motion becomes,

$$\begin{aligned} \frac{dx_i}{dt} &= -c_i(x_i - p_i) + \sum_{j=1}^{N} k_{ij}(x_j - x_i) \exp\left(-\frac{(x_j - x_i)^2}{2l_i^2}\right) \\ &\quad + s_i \sum_n w_i^n(t) v_i(m_i^n - x_i), \quad i = 1, \ldots, N. \end{aligned} \tag{6}$$

Thus, the full model is a system of N coupled, nonlinear ordinary differential equations.

For initial conditions, a typical choice would be for each member to have an initial position corresponding to his or her natural preference, so that

$$x_i(0) = p_i, \tag{7}$$

where $t = 0$ is the start time. However, there are circumstances when the decision-making process concerns revisiting an already extant policy, for which the members may have initial positions other than their natural preferences.

A decision rule could be specified based on the final positions of the members, such as a weighted sum. Rough consensus is achieved if all the group member final positions are within a latitude of acceptance from each other.

References

1. Burton MG, Higley J (1987) Elite settlements. American Sociological Review 52:295–307
2. Diehl PF, Goertz G (2000) War and peace in international rivalry. University of Michigan Press, Ann Arbor, MI, USA
3. Druckman D (1983) Social psychology in international negotiations. In: Kidd RF, Saks MJ (eds) Advances in applied social psychology. Volume 2, Erlbaum, Hillsdale, NJ, USA
4. Druckman D (2001) Turning points in international negotiation. Journal of Conflict Resolution 45:519–544
5. Eagly AH, Chaiken S (1993) The psychology of attitudes. Harcourt, Fort Worth, TX, USA
6. Friedkin NE (2001) Norm formation in social influence networks. Social Networks 23:167–189
7. Friedkin NE, Johnson EC (1999) Social influence networks and opinion change. Advances in Group Processes 16:1–29

8. Gabbay M (2004) A nonlinear spatial model of the dynamics of small group decision making. Paper presented at the annual meeting of the American Political Science Association, Chicago, IL, USA, September
9. Goodby J (1996) When war won out: Bosnian peace plans before Dayton. International Negotiation 1:501–523
10. Hermann CF, Stein JG, Sundelius B, Walker SG (2001) Resolve, accept, or avoid: effects of group conflict on foreign policy decisions. In: Hagan JD, Hermann MG (eds) Leaders, groups, and coalitions. Blackwell, Malden, MA, USA
11. Hermann MG (1984) Personality and foreign policy decision making: a study of 53 heads of government. In: Sylvan DA, Chan S (eds) Foreign policy decision making: perception, cognition, and artificial intelligence. Praeger, New York, USA
12. Hermann MG (2001) How decision units shape foreign policy: a theoretical framework. In: Hagan JD, Hermann MG (eds) Leaders, groups, and coalitions. Blackwell, Malden, MA, USA
13. Hermann MG, Hermann CF (1989) Who makes foreign policy decisions and how: an empirical inquiry. International Studies Quarterly 33:361–387
14. Hogg MA (2001) Social categorization, depersonalization, and group behavior. In: Hogg MA, Tindale RS (eds) Blackwell handbook of social psychology: group processes. Blackwell, Malden, MA, USA
15. Hogg MA, Tindale RS (eds) (2001) Blackwell handbook of social psychology: group processes. Blackwell, Malden, MA, USA
16. Hunter JE, Danes JE, Cohen SH (1984) Mathematical models of attitude change. Volume 1, Academic Press, Orlando, FL, USA
17. Jacobson D (1981) Intraparty dissensus and interparty conflict resolution. Journal of Conflict Resolution, 25:471–494
18. Maoz Z (1990) National choices and international processes. Cambridge University Press, Cambridge, UK
19. McGarty C, Turner JC, Hogg MA, David B, Wetherell MS (1992) Group Polarization as conformity to the prototypical group member. British Journal of Social Psychology 31:1–20
20. Nicholson M (1989) Formal theories in international relations. Cambridge University Press, Cambridge, UK
21. Nowak A, Latane B (1994) Simulating the emergence of social order from individual behavior. In: Gilbert N, Doran J (eds) Simulating societies: the computer simulation of social phenomena. University College London Press, London, UK
22. Pruitt DG and Carnevale PJ (1993) Negotiation in social conflict. Brooks/Cole Publishing, Pacific Grove, CA, USA
23. Pruitt DG, Kim SH (2004) Social conflict: escalation, stalemate, and settlement. McGraw-Hill, New York, USA
24. Stedman SJ (1991) Peacemaking in civil war. Lynne Rienner Publishers, Boulder, CO, USA
25. Stewart PD, Hermann MG, Hermann CF (1989) Modeling the 1973 Soviet decision to support Europe. American Political Science Review 83:35–59
26. ’t Hart P, Stern EK, Sundelius B (1997) Beyond groupthink. University of Michigan Press, Ann Arbor, MI, USA
27. Zartman IW (2000) Ripeness: the hurting stalemate and beyond. In: Stern PC, Druckman D (eds) International conflict resolution after the cold war. National Academy Press, 2000, Washington, DC, USA

Formal Methods for Forecasting Outcomes of Negotiations on Interstate Conflicts

Rudolf Avenhaus and Thomas Krieger

University of the Federal Armed Forces Munich, Neubiberg, Germany

1 Introduction

Conflicts between states can be subdivided into several phases which characterize the intensity of the tension between the states involved. According to [7] these phases are stable peace, unstable peace, major tension, and war. In general, the transitions between the different phases are caused by decisions of one or several government or states that are party to the conflict. In principle, governments have three kinds of decisions at their disposal (see e.g., [11]), namely, cautious actions, de-escalating actions, and escalating actions.

In order to evaluate their possible actions, decision makers take into account their own system of criteria and the consequences of their actions. In doing so, they have to consider the consequences of their actions for their adversaries and their adversaries' foreseeable reactions. Since developing interstate conflicts can be solved only rarely by a single decision, the decision makers have to anticipate complicated sequences of decisions from all parties. The number of alternatives, as well as their consequences, depends upon the number of parties, their possible actions, their criteria, and the number of decisions to be taken sequentially.

Sometimes, in cases that seem to be simple, many quite different aspects have to be taken into account. The evaluation of such a large amount of information—for example, in the course of continuing state of affairs analyses in alliances, interstate organizations, or national foreign affairs and defense ministries—may become very difficult without formal methods that are suited to the reduction of complexity.

We consider here only conflicts where the parties are willing to try to resolve their conflict by means of negotiations. In such cases it may happen—as our examples will demonstrate—that other parties enter the scene for various reasons. Perhaps they are called or offer themselves as mediators or they want to pursue their own interests. Our assumption of being willing to negotiate, however, does not exclude the threat of military action.

As long as the purpose of the analysis of real conflicts is to *understand* what was or is going on, more or less formalized methods may be applied. If, however, one wants to *predict* the outcome of a conflict or even *give advice* to the conflict parties, one needs formalized methods; otherwise one is not able even to define what might be considered a fair solution. Game theory as a tool for conflict analyses has developed methods for the quantitative analyses of negotiations; therefore, if the necessary information is provided, it is, in principle, possible to predict outcomes of interstate conflicts that are resolved with the help of negotiation, perhaps not in a unique way, and perhaps only with some specified probability. It is the purpose of this work to present such a method and to demonstrate it with the help of two concrete examples.

Given an interstate conflict situation, and given the principal willingness of the parties to resolve this conflict peacefully, any formal analysis has to start with the description of the *actors*, their possible actions and reactions, that is, their *strategies*, their information states, and their assessment of possible *outcomes*, which may be vector-valued, for example, characterized by multiple objectives. The payoff vectors are assumed to be known to all the players. Since, in general, negotiations represent *processes* over time, we describe these processes as noncooperative two-person games in extensive form.

In addition, the information states, for example, concerning the knowledge of other parties' moves and unexpected outside events that may have an impact on the development of the conflict under consideration also have to be taken into account.

We do not make any attempt to form scalar payoffs from the vectors since such a procedure would require the application of interrogation techniques (see [3]), which we do not believe can be applied to interstate conflicts that are unique in every single case. Nevertheless, scalarization is important enough for the formal aspects of the theory to be used here (see [5]).

It should be mentioned that, not surprisingly, the elements of the formal analysis of interstate conflicts to be solved by negotiations are the same ones developed in the course of the Processes of International Negotiations (PIN) Project (see, e.g., [4]). The so-called *Analytical Framework for Negotiations Analysis* consists of the components, actors, strategies, process, structure, and outcome. It has been applied for ten years to the study of many international negotiations of various kinds. The correspondence between this framework and extensive form games has been described in more detail by Avenhaus [1].

In Part 2 of this chapter we give an outline of the formal representation of conflicts, introducing the normal or strategic form game as well as the extensive form game with vector valued payoffs, and we also present the equilibrium concept.

In Part 3 we apply the formal methods of Part 2 to the United States (USA)–Iraq conflict as it stood in November 2002. We will proceed as follows: first we will briefly recall the history of the conflict and then examine the decision options and the payoff vectors to the players. The determination of the Pareto equilibria and the discussion of the results concludes Part 3.

Part 4 deals with the modeling of the Rambouillet Negotiations that took place in early 1999 involving Yugoslavia and representatives of the North Atlantic Treaty Organization (NATO). From a formal point of view, we proceed as in Part 3.

In the final part we discuss the advantages and disadvantages of our method: tacit assumptions are examined and the problem of multiple equilibria is raised again. A general statement on the applicability of our method concludes the chapter.

2 Formal Representation of Conflicts

Before turning to our applications we present a short outline of the formal representation of interstate conflicts with the help of noncooperative games in extensive form with vector-valued payoffs. For this purpose we start with a few remarks on vector optimization; thereafter, we introduce noncooperative games in normal and extensive forms in a nonformal way. Mathematical details can be found in textbooks such as [9] or in [5].

2.1 Vector optimization

To solve a classical decision problem means that we are looking for the best among several possible decisions according to a well-defined objective. Mathematically speaking, such a maximization problem is of the form

$$\max_{x \in S} f(x) ,$$

where $f(x)$ is the real-valued objective function and S the feasible region. The purpose of this optimization problem is to determine a point $x \in S$ (if it exists) such that f has the highest objective value.

As we know from daily life, and even more from important real-world problems, for many decisions several objectives have to be considered. Thus, one has the problem

$$\max_{x \in S} f_1(x), \ldots, \max_{x \in S} f_n(x)$$

for given functions $f_1, \ldots, f_n$. As in general no $x \in S$ exists that maximizes all functions at the same time, we have to define what we mean by solving the optimization problem defined above.

One way to solve such a problem is to define the decision maker's scalar value or utility function U, for example, a linear combination of single objectives, and solve the problem

$$\max_{x \in S} U(f_1(x), \ldots, f_n(x)) .$$

The difficulty is that in many cases it is not possible to obtain a mathematical representation of the decision maker's scalar value or utility function U. It is about such problems that we are concerned here. Despite the fact that we may never know the decision maker's utility function, the problem still has to be solved.

Therefore, without explicit knowledge of the decision maker's value or utility function, we have no other choice but to somehow find the "space of tradeoffs" among the objectives of a vector optimization problem. The mathematical optimization concept used here is called the *Pareto optimum*. It means that an element $x_0 \in S$ is a solution of the vector optimization problem if and only if it is not possible to move feasibly from x_0 to increase an objective without decreasing at least one of the others. Formally it means that there exists no $x_1 \in S$ with $f_i(x_1) \geq f_i(x_0)$ for all $i \in \{1, \ldots, n\}$ and $f_i(x_1) > f_i(x_0)$ for at least one $i \in \{1, \ldots, n\}$.

Summarizing a multicriteria decision problem involves the set of alternatives S, the objective functions $f_1, \ldots, f_n$, and, most importantly, the preference structure of the decision maker (see for instance, [19] or [18]). For the latter we use the preference structure given by the Pareto concept. This solution concept is popular in the vector optimization area and in the field of economic theory simply because it is very general.

2.2 Vector games

In game theory, the behavior of rational players in interaction with other rational players is studied. Players are considered to be rational if they maximize their objective function(s) given their beliefs about the behavior of the other players. Since the applications in this chapter deal with two-person games we introduce the concepts and ideas only for cases of two players, although all concepts and ideas can be extended to general n-person games (see for instance, [5]). In noncooperative game theory there are two ways to represent a social interaction as a game, namely, the normal or strategic form and the extensive form.

In the *normal or strategic form* one lists all the possible strategies of each player together with the payoff vector that results from the strategy choices of the players. An example of a two-person vector-valued normal form game is depicted in *Figure 1*, where the game is represented by the payoff matrices of the two players.

I \ II	a	b
(l_1, l_2)	$\begin{pmatrix}3\\6\end{pmatrix}^* \quad {}^*\begin{pmatrix}2\\9\end{pmatrix}$	$\begin{pmatrix}1\\5\end{pmatrix} \quad \begin{pmatrix}0\\3\end{pmatrix}$
(l_1, r_2)	$\begin{pmatrix}1\\7\end{pmatrix}^* \quad {}^*\begin{pmatrix}4\\7\end{pmatrix}$	$\begin{pmatrix}1\\5\end{pmatrix} \quad \begin{pmatrix}0\\3\end{pmatrix}$
(r_1, l_2)	$\begin{pmatrix}2\\5\end{pmatrix} \quad {}^*\begin{pmatrix}5\\0\end{pmatrix}$	$\begin{pmatrix}2\\5\end{pmatrix}^* \quad {}^*\begin{pmatrix}5\\0\end{pmatrix}$
(r_1, r_2)	$\begin{pmatrix}2\\5\end{pmatrix} \quad {}^*\begin{pmatrix}5\\0\end{pmatrix}$	$\begin{pmatrix}2\\5\end{pmatrix}^* \quad {}^*\begin{pmatrix}5\\0\end{pmatrix}$

Fig. 1. The payoff matrix of a two-person vector-valued normal form game

The set of pure strategies of the i-th player, $i \in \{1, 2\}$, is denoted by Φ_i. Thus, we have

$$\Phi_1 = \{(l_1, l_2), (l_1, r_2), (r_1, l_2), (r_1, r_2)\} \quad \text{and} \quad \Phi_2 = \{a, b\}.$$

A strategy combination is a pair $(\phi_1, \phi_2) \in \Phi_1 \times \Phi_2$. The payoff vector to player i which results from a strategy combination (ϕ_1, ϕ_2) is denoted by $H_i(\phi_1, \phi_2)$. In this example both payoff vectors are two-dimensional. We always write in the lower-left-hand corner of an entry of the payoff matrix the payoff vector for player I and in the upper-right-hand corner the payoff vector for player II, which leads, for instance, to $H_1((l_1, l_2), b) = \binom{1}{5}$ and $H_2((l_1, l_2), b) = \binom{0}{3}$.

A Pareto equilibrium expresses the idea that the unilateral deviation of a player from an equilibrium strategy combination cannot improve his payoff vector. This has to be explained in detail. A vector A is *better than* vector B ($A \succ B$) if and only if each component of A is greater or equal than the corresponding component of B and at least one component of A is greater than the corresponding component of B. Using $A := (a_1, a_2, \ldots, a_n)^T \in \mathbb{R}^n$ and $B := (b_1, b_2, \ldots, b_n)^T \in \mathbb{R}^n$ the definition says

$$A \succ B \text{ iff } a_i \geq b_i \text{ for all } i \in \{1, \ldots, n\} \text{ and } a_i > b_i \text{ for at least one } i \in \{1, \ldots, n\} .$$

If neither $A \succ B$ nor $B \succ A$, we call these vectors indifferent and write $A \sim B$. That means especially that $\binom{100}{1} \sim \binom{1}{1.001}$, although a decision maker might intuitively say that the first vector is "better" than the second one, as the deviation in the second component is very small. But this would not meet our equilibrium concept.

A strategy combination (ϕ_1^*, ϕ_2^*) constitutes a *Pareto equilibrium* (see [17]) if and only if

$$\begin{aligned}&\text{there exists no } \phi_1 \in \Phi_1 \text{ with } H_1((\phi_1, \phi_2^*)) \succ H_1((\phi_1^*, \phi_2^*)) \textit{ and}\\ &\text{there exists no } \phi_2 \in \Phi_2 \text{ with } H_2((\phi_1^*, \phi_2)) \succ H_2((\phi_1^*, \phi_2^*)) .\end{aligned}$$

If the payoff vectors of both players consist of only one component, that is, are real-valued, the Pareto equilibrium reduces to the Nash equilibrium [10]. For the games considered in this chapter a Pareto equilibrium always exists; thus, we need not define mixed strategies and expected payoff vectors (see, for instance, [9]).

From a practical point of view we present an algorithm that allows us to determine all Pareto equilibria in a given two-person vector-valued normal form game represented by a payoff matrix.

Take one specific strategy combination (i.e., choose a special entry in the payoff matrix) and consider player I's payoff vector. If it is not dominated by all other payoff vectors in the same *column* (i.e., if there is no vector in the same column with a greater or equal payoff in *any* component and with a greater payoff in *at least* one component), put an asterisk on the vector. Now do the same with player II's payoff vector for the same strategy combination: if it is not dominated by all other payoff vectors in the same *row*, put an asterisk on the vector. Then any entry of the payoff matrix at which I's as well as II's payoff vector have an asterisk constitutes a Pareto equilibrium, and conversely.

For the game in *Figure 1* we obtain the four Pareto equilibria

$$((l_1, l_2), a), ((l_1, r_2), a), ((r_1, l_2), b) \quad \text{and} \quad ((r_1, r_2), b) .$$

There are some vector-valued normal form games that we can simplify by eliminating redundant strategies. Two strategies $\phi_1, \tilde{\phi}_1 \in \Phi_1$ of player I are called payoff-equivalent if and only if $H_i(\phi_1, \phi_2) = H_i(\tilde{\phi}_1, \phi_2)$ for all $\phi_2 \in \Phi_2$ and $i = 1, 2$. Analogously we can define what are payoff-equivalent strategies of player II. In the game depicted in *Figure 1* we see that the strategies (r_1, l_2) and (r_1, r_2) are payoff-equivalent. When there are payoff-equivalent strategies, we can simplify the bimatrix-representation by merging payoff-equivalent strategies and replacing each set of payoff-equivalent strategies by a single strategy. The result of this simplification is called the *reduced* vector-valued normal form game.

An *extensive form game* is the most explicit description of a game. It describes the sequence of moves, all possible states of information, and the choices at different stages for all players of the game. An example of a two-person vector-valued extensive form game is depicted in *Figure 2*.

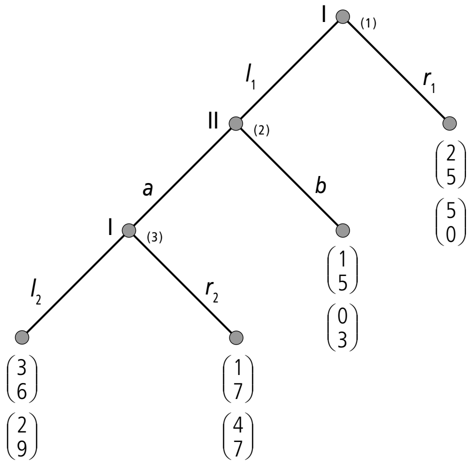

Fig. 2. A vector-valued extensive form game (the nodes are numbered 1, 2, and 3)

The sequence of possible moves in an extensive form game can be represented by a game tree describing the structure of the game. Each possible situation in a game is referred to as a node. In general we have to distinguish nodes where a player has to act and nodes at which the game finishes. The latter are called terminal nodes. Each nonterminal node is labeled with the name of the player who has to act at this node. Sometimes it will be useful to number the nodes as well. In that case we write this number in parentheses on or into a node. A move (decision or action) of a player leads from one node to the next node. At a terminal node the players receive a payoff vector. In *Figure 2* the first payoff vector belongs to player I and the second payoff vector to player II.

How will a game like that in *Figure 2* be played? First, player I has to make a decision. If he chooses r_1 the game ends with the payoff vector $\binom{2}{5}$ for player I and $\binom{5}{0}$ for player II. If player I chooses l_1 next, player II has to make a decision on node 2. If he chooses b the game finishes and in case of the decision a the game continues. Now player I has to decide on node 3 and the game will always end after the move of player I.

A pure strategy of player $i \in \{1,2\}$ in an extensive form game describes for *each* node of player i a unique decision that will be taken as to whether or not this node is reached during the course of the game. For the game in *Figure 2* we obtain the set of strategies

$$\Phi_1 = \{(l_1,l_2),(l_1,r_2),(r_1,l_2),(r_1,r_2)\} \quad \text{and} \quad \Phi_2 = \{a,b\}\,.$$

It should be noted that game-theoretical modeling requires a very strict—not necessarily intuitive—definition of pure strategies: decisions have to be defined also at those nodes that are not reached because of decisions at earlier nodes. If we consider, for instance, the strategy combination $((l_1,l_2),b)$ then the game will end after the decision of player II with the payoff vectors $\binom{1}{5}$ for player I and $\binom{0}{3}$ for player II. Why must player I explain what he would do at node 3 when the game never reaches this node (because of decision b of player II)? It can be shown that this strict definition is necessary and important if one considers deviations from a given strategy combination which is necessary for the determination of Pareto equilibria. Details can be found in [5].

With these preparations we can determine for a given extensive form game the set of strategies and the payoff vectors for each player for a given strategy combination. In other words we can transform the extensive form game into a normal form game. The vector-valued normal form representation of the game in *Figure 2* is depicted in *Figure 1*.

A Pareto equilibrium of an extensive form game is defined as a Pareto equilibrium in the corresponding normal form representation. That means, the game in *Figure 2* also possesses the four Pareto equilibria $((l_1,l_2),a)$, $((l_1,r_2),a)$, $((r_1,l_2),b)$ and $((r_1,r_2),b)$ that were determined earlier. It should be noted that in a normal form representation of an extensive form game especially, the dynamic structure of the latter game is omitted. One concentrates only on the strategic aspects of the game.

Further elements of an extensive form game, which we have to use in one of the applications that will be presented subsequently, are shown in the game of *Figure 3*.

The game starts with a chance move. At nodes that are labeled "chance" the outgoing branches are chosen with a given probability. Next, if we consider extensive form games, we can distinguish games according to the information that players have about the actions chosen by the other players. An information set, shown as an oval in *Figure 3*, can be used to model the lack of information of that player about the course of a game. This means that for the game in *Figure 3*, player I does not know which node of the information set is actually reached. He only knows that he stands in one of the nodes of the information set. This implies that player I does not know the outcome of the chance move.

An extensive form game with trivial information sets (i.e., those consisting of single nodes), is called a game with perfect information. Otherwise it is called a game with imperfect information. We will see that the game representing the USA–Iraq conflict has imperfect information and that the game representing the Kosovo conflict has perfect information.

Extensive form games with real-valued payoffs (i.e., one-component payoff vectors), can be solved with the help of a backward induction procedure [6]. This

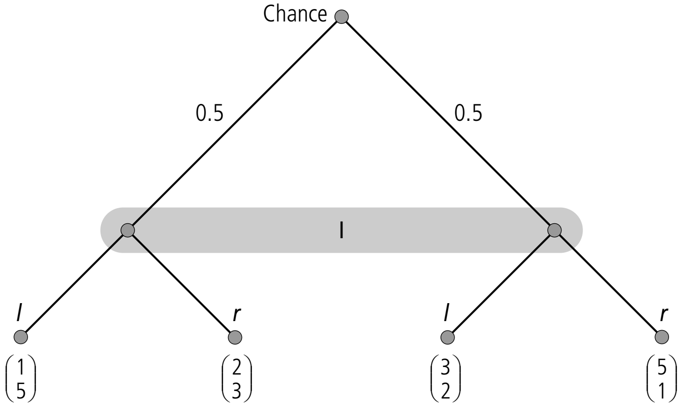

Fig. 3. A vector-valued one-person extensive form game with chance move and an information set

procedure leads to a Nash equilibrium [10], which is a special case of the Pareto equilibrium concept introduced above. Unfortunately, this procedure cannot be applied to the games discussed here because of the lack of the transitivity of the indifference relationship of our payoff vectors. For example, define $A := \binom{3}{6}$ $B := \binom{1}{7}$ and $C := \binom{2}{5}$. We then have $A \sim B$ and $B \sim C$ but not $A \sim C$.

If one tries to determine Pareto equilibria with a procedure similar to that used for the real-valued games, one sees that different effects may occur which make such a procedure problematic. A vector-valued backward induction procedure can be found in [5].

As the numbers of Pareto equilibria in extensive form games might be very large, their use poses a problem to applications, whether it be for predictive or normative purposes. In fact, one usually talks only about a *solution* of a game, if just one equilibrium exists. Therefore, in case of multiple equilibria we are looking for equilibrium selection procedures. It should be mentioned, however, that we do not want to exaggerate this, as multiple equilibria may reflect the complexity of the conflict under consideration.

Therefore, we consider here only one, very natural selection concept, namely, the subgame perfect Pareto equilibrium. This concept was introduced for real-valued games by Selten [15]. A subgame perfect Pareto equilibrium is a Pareto equilibrium (for the entire game) whereby the restriction of the equilibrium to *each* subgame constitutes a Pareto equilibrium in that subgame. The strategy combination $((r_1, l_2), b)$ constitutes a Pareto equilibrium of the game in *Figure 2*. For the restriction of that strategy combination to the subgame starting at node 2 we obtain (l_2, b), as the decision r_1 cannot be made in this subgame. But this strategy combination cannot be a Pareto equilibrium of the remaining game, as a deviation of player II will lead to the payoff vector $\binom{2}{9}$ whose components are greater than the corresponding components of $\binom{0}{3}$. Only the strategy combinations $((l_1, l_2), a)$ and $((l_1, r_2), a)$

are subgame perfect Pareto equilibria. The popularity of this refinement concept lies in the fact that if one player deviates from his equilibrium strategy the behavior of all players will be in equilibrium again.

Furthermore, it should be noted that all players know the game tree and the payoff vectors on the terminal node for *each* player. This assumption is called common knowledge.

3 The Conflict between the United States and Iraq in 2002

Let us briefly recall the history of the conflict between the United States and Iraq as it stood in the fall of 2002. In August 1990, led by the dictator Saddam Hussein, Iraq had occupied the Emirate of Kuwait. This was not acceptable to the international community, and following a mandate of the United Nations (UN), an alliance under the leadership of the USA liberated Kuwait in the so-called Gulf War in January and February 1991 ("Operation Desert Storm"), invaded the south of Iraq but neither conquered Baghdad nor overthrew Saddam Hussein's regime.

In subsequent negotiations performed under the auspices of the United Nations, international inspections of Iraq were agreed, among other measures, and these were carried out under the supervision of the International Atomic Energy Agency (IAEA) based in Vienna, which had been active in Iraq before the Gulf War in the framework of the Treaty on the Non-proliferation of Nuclear Weapons. Because of continuing problems between the Iraqi authorities and the international inspectorate, which was unable to carry out the inspections as it deemed necessary, the inspections ultimately ceased in 1998. Nevertheless, suspicion continued because of the obstacles placed in the way of the inspectors by Saddam Hussein who, it was claimed, was developing and producing all kinds of weapons of mass destruction (WMD).

After the events of September 11 in 2001, and the subsequent attempt by US-led forces to eliminate the Taliban regime in Afghanistan, the USA accused Iraq of supporting Islamic terrorism throughout the Western world. The USA in 2002 thus began to evaluate the possibilities of countering this twofold threat—support of terrorism and possession of WMD. The UN continued to urge Iraq to let inspectors reenter the country and perform their inspections properly.

Assuming the situation as described in the fall of 2002, we intend to perform a *predictive* outcome as discussed in the introduction.

3.1 Decision options and payoffs

In October 2002 the conflict between the USA as the leading Western power and Iraq may be described simply as a noncooperative two-person game, the decision options and payoffs of which are given as follows.

In the beginning the decision options (and in parentheses their representation as shown in *Figure 4*) are:

- No war, UN inspections in Iraq (no war, inspections).

- War against Iraq with the UN mandate (war with UN).
- War against Iraq without the UN mandate (war without UN).

Knowing the decision of the USA, Iraq's decision options, in the instance where the USA does not start a war, imply:

- Stop or do not start any support of terrorism, destroy any existing weapons, and do not start any development or production of weapons of mass destruction (no terror support, no WMD).
- Stop or do not start any support of terrorism *and* start or continue the development or production of weapons of mass destruction (no terror support, WMD).
- Support terrorism *and* stop (i.e., destroy any weapons) or do not start any development or production of weapons of mass destruction (terror support, no WMD).
- Support terrorism *and* start or continue the development or production of weapons of mass destruction (terror support, WMD).

If the USA starts a war against Iraq, we think that Saddam Hussein's Iraq had no choices of the kind described above. We did not consider the possibility that Iraq might refuse to accept the new UN Resolution and to let international inspectors again enter their country. This was confirmed on 14 November 2002.

For the sake of simplicity it is assumed that UN inspectors will detect whether or not weapons of mass destruction are being developed or produced in Iraq. Furthermore, it is assumed in the former case that development will stop and existing weapons will be destroyed. However, the same would not be appropriate with regard to the support of terrorism if it were not part of the inspectors' mandate. Instead, we simply assume that with probability $1 - \alpha$ no support of terrorism is confirmed, while support is confirmed with probability $1 - \beta$. In line with standard statistical criteria we assume the idealized decision procedure to be *unbiased* (i.e., we assume $\alpha + \beta < 1$) (see for instance, [14]).

We assume also that a war involving the USA and Iraq, with or without a UN mandate, will end Saddam Hussein's regime and will put an end to his eventual support of terrorism and the development and production of weapons of mass destruction.

The extensive form of this noncooperative two-person game is represented graphically in *Figure 4*.

At the terminal nodes of the game tree the payoff to both players are given, which are defined as follows.

The four criteria for the USA and their three possible payoffs (-1 worst, 0 neutral, $+1$ best) are:

- To end Saddam Hussein's regime ($+1$ in case of success, otherwise -1).
- To stop support of terrorism ($+1$ respectively -1).
- To stop development and production of weapons of mass destruction ($+1$ respectively -1).
- To maintain United States (US) standing in the Islamic world (-1 in case of war without a UN mandate, 0 in case of war with such a mandate, $+1$ in case of no war).

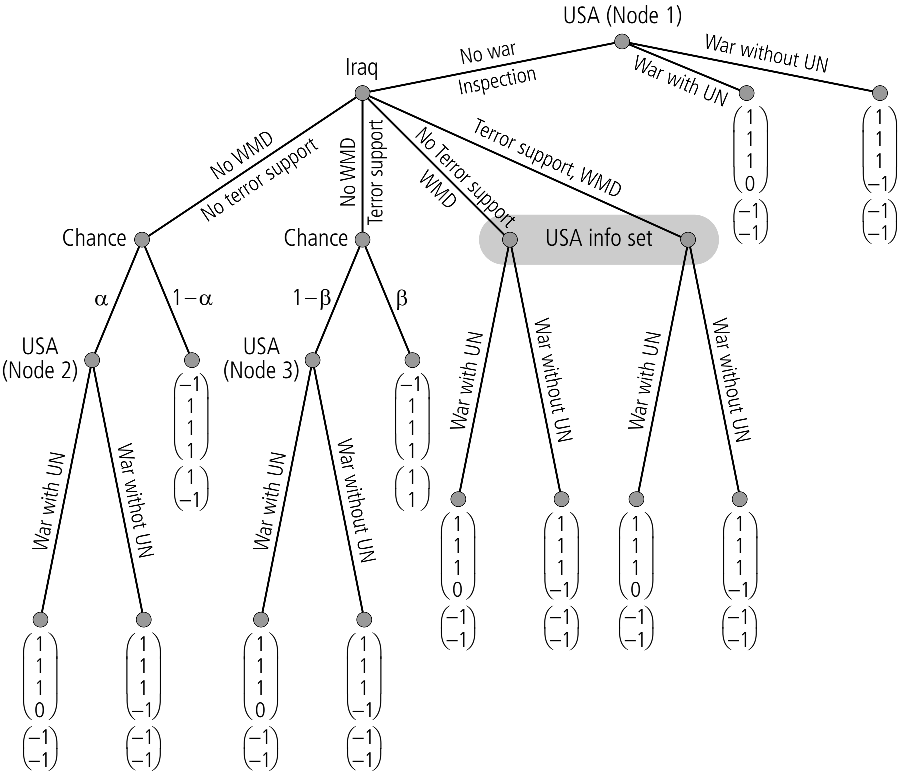

Fig. 4. The USA–Iraq game in extensive form. The four-component vector describes the USA's payoff, and the second two-component vector describes Iraq's payoff for each possible outcome

It should be emphasized that there is no ordering of these criteria and that we do not see any possibility of combining them into a single scalar criterion. There is, however, some dependence: we assume that ending Saddam Hussein's regime implies an end to the support of terrorism as well as to development and production of weapons of mass destructions.

Because of the lack of any information of that kind, the costs of a war—in terms of human and financial losses—are not taken into account as a criterion. Moreover, the question of regional stability with Saddam Hussein's regime, or without it after a war, was not considered important.

The two criteria for Iraq, as assumed to be viewed by Saddam Hussein's regime, and their possible payoffs are:

- Maintenance of power of the present regime ($+1$ respectively -1).
- Defense of the Islamic world, in particular against the (assumed) hegemonic ambitions of the USA ($+1$ respectively -1).

With these two criteria we try to explain the decisions of Iraq which appeared unreasonable at first sight, as Saddam Hussein must have known that a war with the USA would terminate only with the end of his regime. Why did he run such a risk if he could maintain power for a long time if he surrendered support of terrorism and the development and production of WMD? Our interpretation of his possible choice of one of these decisions is that he uses the first one, support of terrorism, in order to meet his defense of the Islamic world criterion, and the second one, WMD, in order to maintain his power, as he thought that he could not be attacked once he possessed WMD in sufficient quantities.

One important modeling aspect should be mentioned here: the choice of -1, 0, and $+1$ for the payoffs is crucial for our results, as, because of the possible errors (with probabilities α and β) we have to consider expected payoffs. As probabilities of that kind do not occur in our subsequent example, only the order of the payoffs is relevant.

3.2 Pareto equilibria and discussion

In this part we shall determine all subgame perfect Pareto equilibria of the USA–Iraq game using two different approaches. First, we determine all Pareto equilibria of the normal form representation of the game and subsequently select from this set the subgame perfect Pareto equilibria. In the second approach, we use properties of the game tree as well as the payoff vectors to determine the subgame perfect Pareto equilibria with the help of a backward induction procedure. Both approaches lead to the same results. More mathematical details may be found in [5].

For the computation of Pareto equilibria we use the following properties A, B, and C of some of the payoff vectors of the two players,

$$A : \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \succ \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} B : \begin{pmatrix} 1-2\alpha \\ -1 \end{pmatrix} \succ \begin{pmatrix} -1 \\ -1 \end{pmatrix} C : \begin{pmatrix} 2\beta-1 \\ 2\beta-1 \end{pmatrix} \succ \begin{pmatrix} -1 \\ -1 \end{pmatrix} , (*)$$

which are fulfilled if and only if $\alpha < 1$ and $\beta > 0$.

The normal form representation of the USA–Iraq game, which is represented graphically in *Figure 4*, is depicted in *Table 1*.

To obtain a US strategy we must combine the decision: "No war, inspections" at node 1, with one of the decisions encoded in the abbreviations $1, \ldots, 8$. This leads to the first eight columns in *Table 1*. As the same combination with the decisions "War with UN" and "War without UN" at node 1 always leads to the same terminal node of the game tree and thus to the same payoff vectors, we have combined it in one column. With the help of the properties A, B, and C as given in $(*)$ and the unbiasedness of the idealized test procedure, $\alpha+\beta < 1$, we get the following results.

In case of nonvanishing error probabilities (i.e., $\alpha > 0$ and $\beta > 0$), we obtain many Pareto equilibria that are characterized by the shaded areas in *Table 1*. The four dark-shaded areas are the strategy pairs, which are subgame perfect Pareto equilibria

USA / Iraq	No war and inspections at node 1								War with UN at node 1		War without UN at node 1	
	1	2	3	4	5	6	7	8	1	2,...,8	1	2,...,8
No terror support, No WMD	$2\alpha-1$ 1 1 $1-\alpha$ $1-2\alpha$ -1	$2\alpha-1$ 1 1 $1-\alpha$ $1-2\alpha$ -1	$2\alpha-1$ 1 1 $1-\alpha$ $1-2\alpha$ -1	$2\alpha-1$ 1 1 $1-\alpha$ $1-2\alpha$ -1	$2\alpha-1$ 1 1 $1-2\alpha$ $-\alpha$ -1	$2\alpha-1$ 1 1 $1-2\alpha$ $-\alpha$ -1	$2\alpha-1$ 1 1 $1-2\alpha$ $-\alpha$ -1	$2\alpha-1$ 1 1 $1-2\alpha$ $-\alpha$ -1	1 1 1 0 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 -1 -1 -1
Terror support, No WMD	$1-2\beta$ $1-2\beta$ 1 β $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 β $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 $2\beta-1$ $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 $2\beta-1$ $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 β $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 β $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 $2\beta-1$ $2\beta-1$ $2\beta-1$	$1-2\beta$ $1-2\beta$ 1 $2\beta-1$ $2\beta-1$ $2\beta-1$	1 1 1 0 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 -1 -1 -1
No terror support, WMD	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 -1 -1 -1
Terror support, WMD	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 0 -1 -1	1 1 1 0 -1 -1	1 1 1 -1 -1 -1	1 1 1 -1 -1 -1

Decisions of the USA			Abbreviation
War with UN at node 2	War with UN at node 3	War with UN at node 4	1
		War without UN at node 4	2
	War without UN at node 3	War with UN at node 4	3
		War without UN at node 4	4
War without UN at node 2	War with UN at node 3	War with UN at node 4	5
		War without UN at node 4	6
	War without UN at node 3	War with UN at node 4	7
		War without UN at node 4	8

Table 1.

Top: Normal form of the USA–Iraq game represented in *Figure 4*. The shaded areas are strategy pairs, which are Pareto equilibria in case of $\alpha > 0$ and $\beta > 0$

Left: Abbreviations of decisions of the USA and which can be seen with the help of the extensive form game given in *Figure 4*. In case of vanishing error probabilities (i.e., $\alpha = \beta = 0$), we obtain as subgame perfect Pareto equilibria only the strategy pairs in the first row that are dark-shaded.

In order to determine the equilibria with the help of the backward induction procedure, we first reduce the extensive form game represented graphically in *Figure 4* to the extensive form game as given by *Figure 5*. For this purpose, we consider the expected payoff vector for the subgames starting at those nodes which are labeled with "chance" for each player. Furthermore, because of property A in ($*$) the USA would choose the decision "War without UN" neither at node 1 nor at nodes 2 and 3, nor at the information set. This decision can be canceled everywhere. Finally, because of properties B and C in ($*$) Iraq would never choose one of the decisions which leads to the information set of the USA.

The subgame perfect Pareto equilibria of the *reduced* game of *Figure 5* are, in the case of $\alpha > 0$ and $\beta > 0$, all combinations of decisions of both players at their respective nodes, whereas in the case of $\alpha = \beta = 0$ the equilibria are the combinations "War with UN, no terror support, and no WMD" and "No war and inspections, no terror support, and no WMD". It should be noted that in the case of "War with UN" the decision of Iraq also has to be specified for formal reasons, even though it is irrelevant in practice. It should also be noted that in this simple game the problems of applying backward inductions to vector-valued games, as discussed in the previous section, do not occur.

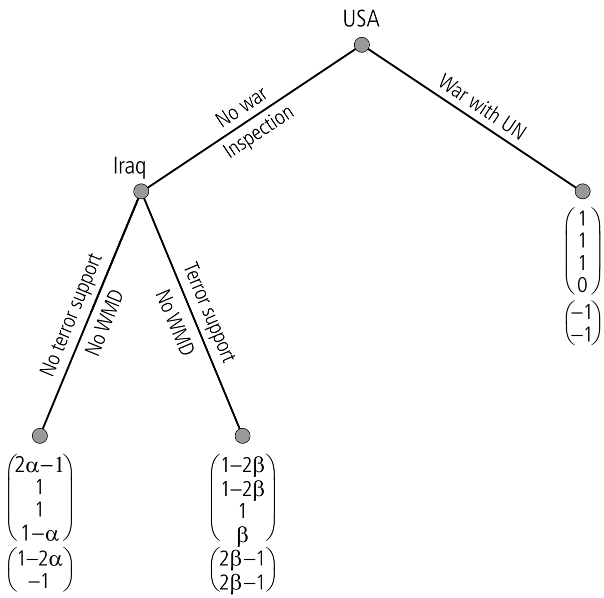

Fig. 5. The *reduced* USA–Iraq game in extensive form

The subgame perfect Pareto equilibria for the *original* game can be obtained if one adds the decision "War with UN" at nodes 2 and 3 in *Figure 4* and at the information set of the USA.

The paths that result from the equilibrium strategies of the two players are described as follows

- In the case of $\alpha > 0$ and $\beta > 0$
  – USA at node 1: war with UN;
  – USA at node 1: no war, inspections $\longrightarrow$
    Iraq: no terror support, no WMD $\longrightarrow$
    No war with probability $1 - \alpha$ and war with UN with probability α;
  – USA at node 1: no war, inspections $\longrightarrow$
    Iraq: terror support, no WMD $\longrightarrow$
    No war with probability β and war with UN with probability $1 - \beta$;
- In the case of $\alpha = 0$ and $\beta = 0$
  – USA at node 1: war with UN;
  – USA at node 1: no war, inspections $\longrightarrow$
    Iraq: No terror support, no WMD $\longrightarrow$
    No war.

Let us repeat the results of the game-theoretic analysis presented so far. We obtain the following subgame perfect Pareto equilibria:

- In the case of $\alpha > 0$ and $\beta > 0$, the equilibrium strategies of the USA imply war with a UN mandate with certainty as well as war with a UN mandate with some positive probability, depending on which equilibrium is actually chosen.
- In the case of $\alpha = \beta = 0$, the equilibrium strategies of the USA imply war with a UN mandate as well as no war, depending on which equilibrium is actually chosen.

Before evaluating these findings let us consider two interesting special cases of the USA–Iraq game that arises if one were to omit the first or the fourth criteria in the payoff vectors of the USA. We give here only the results and omit the mathematical analysis that can be performed with the help of *Table 1*.

At the start we omit the first criterion of the USA (i.e., ending Saddam Hussein's regime). This seems to be an interesting special case since some observers consider this criterion to be motivated primarily by President Bush's domestic political objectives. As recent history tells us, the mere fact of the existence of a dictator is not a cause for war for the USA. We obtain:

- In the case of $\alpha > 0$ and $\beta > 0$, the equilibrium strategies of the USA imply war with a UN mandate with certainty, as well as war with a UN mandate with a positive probability, depending on which equilibrium is actually chosen.
- In the case of $\alpha = \beta = 0$, the equilibrium strategy of the USA implies no war at all and that of Iraq supports neither terrorism nor the development and production of weapons of mass destruction.

Next, we omit the fourth criterion of the USA (i.e., the maintenance of its standing in the Islamic world), as there are voices in the USA that question the value of this standing. For this game we obtain, independently of the constellation of α and β, that the equilibrium strategies of the USA imply war with a UN mandate and war without a UN mandate, respectively.

Let us now discuss these results in the sense that we are trying to *predict* possible outcomes of this conflict—which may be interpreted as a modest form of recommendation to the conflict participants. Let us repeat that we still assume that we are in October 2002, even though several years have passed and history has not stood still.

If we consider all criteria, as well as the general case of $\alpha > 0$ and $\beta > 0$, then we have a problem: we cannot present a solution of the game as we have to consider four subgame perfect Pareto equilibria. Which of the two represents the predicted outcome, or which one should be recommended? As we do not want to apply formal equilibrium selection theory (see for instance [2]), beyond what we have already done, namely, considering only subgame perfect Pareto equilibria, we tried to use substantive arguments and arrived at the following conclusions.

We saw in the case of $\alpha > 0$ and $\beta > 0$, and without the first criterion of the USA, that there are still three subgame perfect Pareto equilibria. However, for small β, there is only the second component of the US payoff vector $(1 - 2\,\beta)$; in the case of the decision no war of the USA and terror support of Saddam Hussein, Iraq is slightly worse-off than in the case of war (1), whereas in all other cases all payoffs to the USA are equal or better. We assume that the community of states represented in the UN will succeed in convincing the USA not to insist on ending Saddam Hussein's regime.

Following this path, Saddam Hussein's Iraq still has the possibility of either supporting terrorism or not. Therefore, it depends on the probabilities α and β as to whether or not the USA will start a war with a UN mandate, with probability α in the case of no support of terrorism and with probability $1 - \beta$ in the case of support from the Iraqi side.

4 The Rambouillet Negotiations in 1999

The beginning of the decline of the Socialist Federal Republic of Yugoslavia (SFRY), which began after the death of Marshall Tito in 1980 in the second half of the 1980s, had its origin in the abrogation of the autonomy of Kosovo in 1989–1990. This represented the first step in the realization of the greater Serbian state, envisaged by Serbian President Milosevic at the expense of the other nations in the former Yugoslavia.

In July 1990 the Albanian deputies of the Kosovo Parliament declared the independence of the Republic of Kosovo in reaction to the establishment of a centralistic Serbian state. As a consequence, Belgrade abandoned the parliament and the government of its southern province. In September 1990, the unrecognized Republic of Kosovo established a constitution and elected Ibrahim Rugova as president. His

Democratic League of Kosovo (LDK) won the underground elections in May 1992, which were not recognized by Belgrade.

The most important objective of the LDK was the creation of an internationally recognized neutral Republic of Kosovo. The Albanians, who represented 90 percent of the 2 million inhabitants of Kosovo, argued that Kosovo was a constitutive part of the SFRY and therefore had the same rights of self-determination as the other Slavic nations that had formed Yugoslavia.

The Republic of Kosovo from 1990–1997 had a form of shadow government that established some institutions, particularly in areas such as health and education, that existed in parallel to the equivalent Serbian institutions. But in the middle of the 1990s a polarization arose. While the LDK did not want to use force, a second bloc wanted to create an "Albanian intifada." The liberation army of Kosovo (UCK) was transformed from 1996 to 1999 from a guerilla movement into a people's army.

In spring 1998 operations involving Serbian forces and the UCK in Kosovo resulted in large numbers of Albanian refugees, and by September 1998 these totaled around 265,000. NATO began planning the use of military actions to end the fighting. The UN Security Council passed resolutions in March and September 1998, but they had no effect on President Milosevic. NATO then started preparing for air strikes in October 1998. It should be noted that the resolutions included the threat of military action if Yugoslavia did not comply, but Russia and China did not interpret this as a formal basis for these actions.

Influenced by the actions of NATO, the Yugoslav government negotiated the Holbrooke–Milosevic agreement of October 1998, that foresaw the end of military operations in Kosovo and the return of refugees to their villages under the supervision of the Organization for Security and Co-operation in Europe (OSCE) observers. Despite these agreements, the situation in Kosovo became worse in the winter of 1998–1999. There was fighting and additional tens of thousands of refugees; at the end of January the conflicting parties received an ultimatum to begin negotiations within a week.

On 6 February 1999 negotiations between the Western community of states and Yugoslavia started at Rambouillet, near Paris. The meeting was chaired by the French and British ministers of foreign affairs, Hubert Vendrine and Robin Cooke; the negotiations were directed by the Austrian Ambassador in Yugoslavia, Wolfgang Petrisch, the US mediator, Christopher Hill, and the Russian Ambassador, Boris Majorski.

The meeting lasted until 23 February without success. On 22 March US Ambassador Holbrooke tried for the last time, unsuccessfully, to convince President Milosevic to comply. On 23 March 1999, NATO began air strikes against Yugoslavia that continued until 20 June.

More details of the Kosovo conflict can be found in [12] and [13], where the semiquantitative so-called scenario bundle method is used to forecast long-term developments and possible solutions of the conflict. A presentation of this method is given in English in Selten [16].

4.1 Decision options and payoffs

In contrast to the previous analysis, we consider here the events that happened long before we began to look at them; in other words we are performing a *descriptive* analysis.

We assume that at the beginning of the negotiations in Rambouillet in February 1999 the strategic alternatives of the Western community of states (for simplicity's sake, NATO) and Yugoslavia were known to both parties, and that they posed:

- A credible threat of ground war (TGW)
- A credible threat of air strike (TAS)
- No action (NO1)

for NATO at a first level. Subsequently, the alternatives:

- Giving in (G), meaning an end to ethnic cleansing and acceptance of the stationing of NATO ground forces.
- Not giving in (NG)

for Yugoslavia, and finally, depending on Yugoslavia's decision, the alternatives of:

- Ground war (GW)
- Air strike (AS)
- No action (NO2)

for NATO.
The game tree describing this situation is represented graphically in *Figure 6*.

The outcome of all the strategic combinations of the two players are valued by NATO with the help of the following five criteria. It should be noted that the order is arbitrary (i.e., does not mean any relative valuation of the criteria):

- End of ethnic cleansing;
- Ability to act;
- Regional threat potential (i.e., potential for political destabilization of Balkan states);
- Cost;
- Law of nations.

Whereas the outcomes are valued by Yugoslavia using only two criteria:

- (Milosevic's) maintenance of power;
- Integrity of Yugoslavia (i.e., Kosovo remaining part of Serbia).

All criteria are ordered in such a way that only five values are possible, with -2 being the worst, and $+2$ being the best outcome for the parties concerned.

As an example we assume NATO threatens an air strike, Yugoslavia does not give in, and NATO decides to launch the air strike. This leads to values $(1, 0)$ for Yugoslavia and $(-1, 1, 1, -1, -2)$ for NATO.

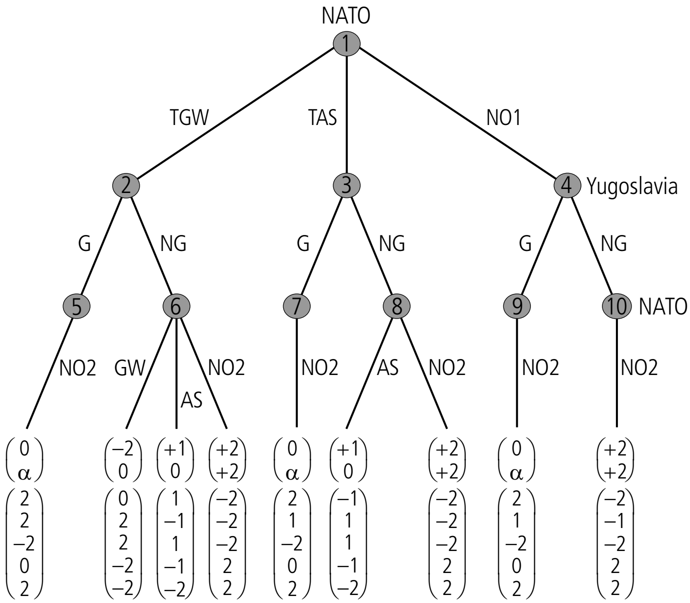

Fig. 6. Rambouillet negotiation game in extensive from. Abbreviations are explained in the text. At the bottom, the first two-component vector describes Yugoslavia, and the second five-component vector describes NATO's payoff for each possible outcome

An explanation of the values of the criteria that were chosen and that are not obvious follows. It should be emphasized, however, that we consider the values highly subjective and in part questionable. The same holds true for the criteria themselves.

With respect to the ending of ethnic cleansing, except for the obvious case, we take 1 for AS and 0 for GW as in the latter case, ethnic cleansing could be accelerated. Furthermore, we take 1 for (TGW, NG, AS) and -1 for (TAS, NG, AS), as TAS is assumed not to be taken as seriously by GW.

For the ability to act we take 2 for (TGW, G) or (TGW, NG, GW) as it demonstrates this capability, 1 for (TAS, G) or (TAS, NG, AS), -2 for threatening, and subsequently NO2 for obvious reasons.

For costs we take 2 for (NG, NO2) for obvious reasons, 0 for (G, NO2) because of cost of stationed forces, -1 for (NG, AS) and -2 for (NG, GW) for obvious reasons.

Finally, for the integrity of Yugoslavia criterion, we take α for (G) depending on the expectation that in the long term it may lead to separation ($\alpha < 0$), but that this need not be so ($\alpha \geq 0$), 2 for (NG, NO2) as ethnic cleansing will leave only the Serbian population in Kosovo, which then stays with Yugoslavia, 0 for (NG, GW) or (NG, AS) as developments cannot be foreseen.

An explanation of the Yugoslavian and NATO strategies is given in *Tables 2* and *3*.

Table 2. Explanation of the eight strategies of Yugoslavia

If Yugoslavia plays strategy number	and the game is in node 2	3	4
	then Yugoslavia decides for		
1	G	G	G
2	G	NG	G
3	G	G	NG
4	G	NG	NG
5	NG	G	G
6	NG	NG	G
7	NG	G	NG
8	NG	NG	NG

Table 3. Explanation of the 18 NATO strategies. Since the structure of these strategies might not be easily understood, for NATO the Cartesian product of the sets of possible actions at nodes 1 and 5 to 10 (i.e., those nodes at which NATO has to decide), is given as follows: $\{TGW, TAS, NO1\} \times \{NO2\} \times \{GW, AS, NO2\} \times \{NO2\} \times \{AS, NO2\} \times \{NO2\} \times \{NO2\}$

If NATO plays strategy number	and the game is in node 1	5	6	7	8	9	10
	then NATO decides for						
1	TGW	NO2	GW	NO2	AS	NO2	NO2
2	TGW	NO2	GW	NO2	NO2	NO2	NO2
3	TGW	NO2	AS	NO2	AS	NO2	NO2
4	TGW	NO2	AS	NO2	NO2	NO2	NO2
5	TGW	NO2	NO2	NO2	AS	NO2	NO2
6	TGW	NO2	NO2	NO2	NO2	NO2	NO2
7	TAS	NO2	GW	NO2	AS	NO2	NO2
8	TAS	NO2	GW	NO2	NO2	NO2	NO2
9	TAS	NO2	AS	NO2	AS	NO2	NO2
10	TAS	NO2	AS	NO2	NO2	NO2	NO2
11	TAS	NO2	NO2	NO2	AS	NO2	NO2
12	TAS	NO2	NO2	NO2	NO2	NO2	NO2
13	NO1	NO2	GW	NO2	AS	NO2	NO2
14	NO1	NO2	GW	NO2	NO2	NO2	NO2
15	NO1	NO2	AS	NO2	AS	NO2	NO2
16	NO1	NO2	AS	NO2	NO2	NO2	NO2
17	NO1	NO2	NO2	NO2	AS	NO2	NO2
18	NO1	NO2	NO2	NO2	NO2	NO2	NO2

4.2 Pareto equilibria and discussion

The extensive form game of *Figure 6* can be transformed into the normal form game depicted in *Table 4*. Furthermore, this game can be reduced to a normal form game with the strategies $1, \ldots, 8$ for Yugoslavia and the strategies $2, 4, 6, 11, 12, 13$ for NATO, as the other strategies are a payoff-equivalent to one of the former strategies. The reduced normal form game is shown in *Table 5*.

With the help of the algorithm formulated in the second section it is now quite easy to determine all Pareto equilibria, where we have to distinguish the three cases $\alpha < 0, \alpha > 0$ and $\alpha = 0$. In *Table 6* all reduced normal form games are depicted as well as the Pareto equilibria in the shaded areas. We can now see many Pareto equilibria.

As a consequence, nearly all paths of the game may lead to equilibria. Essentially, only statements can be made about what *cannot* happen if pairs of equilibrium strategies are chosen by the players.

At first sight this might be considered a disappointing result as no prediction or advice could have been derived if this analysis had been performed in February 1999. One conclusion, however, can be (or could have been) drawn:

> For $\alpha \geq 0$ ground war by NATO is *not* part of a Pareto equilibrium strategy and therefore would have been neither predicted nor proposed.

This can be seen, as the strategy pairs $(5, 2), (6, 2), (7, 2)$ and $(8, 2)$ of *Table 6* are the only ones shaded in the case of $\alpha < 0$. Exactly these four strategy pairs lead to the resulting ground war by NATO.

Let us assume, just for the purpose of illustration, that both players have only one criterion. For both Yugoslavia and NATO, this is the end of ethnic cleansing. The extensive form game can be solved by a backward induction procedure. It turns out that an equilibrium pair of strategies leads to the sequence of decision:

> threat of groundwar by NATO $\rightarrow$ not giving in by Yugoslavia $\rightarrow$ air strike by NATO.

At least the second- and third-step decisions have been made in reality.

Finally, let us mention that we do not present the second approach for determining the equilibria based on the extensive form, as in using this method the number of equilibria is essentially not reduced (i.e., it does not give additional insight).

5 Summary of Results and Recommendations

We have illustrated an approach that is intended either to further understanding of conflicts and negotiations or to help decision makers predict outcomes of negotiations of interstate conflicts, with two examples, namely, 1) the conflict between the USA and Iraq as it existed in the fall of 2002 and 2) the negotiations for a solution to the Kosovo conflict at Rambouillet in early 1999. Let us summarize the steps of the general game-theoretical analysis procedure for this kind of international negotiation:

Table 4. Strategic or normal form of the extensive form game represented graphically in *Figure 6*. A, B, C, D, E, F, G, H, and I are the NATO vector payoffs explicitly given in *Figure 6* from left to right

NATO Yugoslavia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α
2	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	F 1 0	G 2 2	F 1 0	G 2 2	F 1 0	G 2 2	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α
3	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2
4	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	A 0 α	F 1 0	G 2 2	F 1 0	G 2 2	F 1 0	G 2 2	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2
5	B -2 0	B -2 0	C 1 0	C 1 0	D 2 2	D 2 2	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α
6	B -2 0	B -2 0	C 1 0	C 1 0	D 2 2	D 2 2	F 1 0	G 2 2	F 1 0	G 2 2	F 1 0	G 2 2	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α	H 0 α
7	B -2 0	B -2 0	C 1 0	C 1 0	D 2 2	D 2 2	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	E 0 α	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2
8	B -2 0	B -2 0	C 1 0	C 1 0	D 2 2	D 2 2	F 1 0	G 2 2	F 1 0	G 2 2	F 1 0	G 2 2	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2	I 2 2

Table 5. The reduced normal form game of the game in *Table 4*

NATO Yugoslavia	2	4	6	11	12	13
1	A 0 α	A 0 α	A 0 α	E 0 α	E 0 α	H 0 α
2	A 0 α	A 0 α	A 0 α	F 1 0	G 2 2	H 0 α
3	A 0 α	A 0 α	A 0 α	E 0 α	E 0 α	I 2 2
4	A 0 α	A 0 α	A 0 α	F 1 0	G 2 2	I 2 2
5	B -2 0	C 1 0	D 2 2	E 0 α	E 0 α	H 0 α
6	B -2 0	C 1 0	D 2 2	F 1 0	G 2 2	H 0 α
7	B -2 0	C 1 0	D 2 2	E 0 α	E 0 α	I 2 2
8	B -2 0	C 1 0	D 2 2	F 1 0	G 2 2	I 2 2

- List all parties to the conflict, their possible actions and reactions, and evaluate all possible outcomes of the negotiations, in general, by multiple objectives.
- Describe the negotiations with the help of an extensive form game (i.e., draw a game tree that shows which of the parties chooses which alternatives in which order). Ultimately, take into account limited information and chance moves.
- If possible without major effort, determine the equilibrium points of this game by an appropriate backward induction, or if this turns out to be too complicated—even though possible in principle—transform the extensive form game into a normal one and determine all equilibrium points.

This was mentioned in the introduction to this chapter, and our examples demonstrate that in realistic models of major international negotiations the number of equilibria may become large very quickly. Whereas, in our cases, the equilibria could still be

Table 6. From left to right is depicted the reduced normal form game from *Table 5* with the corresponding equilibria in the shaded areas for cases $\alpha < 0$, $\alpha > 0$ and $\alpha = 0$. Y is the abbreviation for Yugoslavia

NATO Y	2	4	6	11	12	13
1	**A 0 α**	A 0 α	A 0 α	E 0 α	E 0 α	H 0 α
2	**A 0 α**	A 0 α	A 0 α	**F 1 0**	**G 2 2**	H 0 α
3	**A 0 α**	A 0 α	A 0 α	E 0 α	E 0 α	**I 2 2**
4	**A 0 α**	A 0 α	A 0 α	**F 1 0**	G 2 2	**I 2 2**
5	B -2 0	**C 1 0**	**D 2 2**	E 0 α	E 0 α	H 0 α
6	B -2 0	**C 1 0**	**D 2 2**	**F 1 0**	**G 2 2**	H 0 α
7	B -2 0	**C 1 0**	D 2 2	E 0 α	E 0 α	**I 2 2**
8	B -2 0	**C 1 0**	D 2 2	**F 1 0**	G 2 2	**I 2 2**

NATO Y	2	4	6	11	12	13
1	**A 0 α**	**A 0 α**	A 0 α	E 0 α	E 0 α	H 0 α
2	**A 0 α**	**A 0 α**	A 0 α	**F 1 0**	**G 2 2**	H 0 α
3	**A 0 α**	**A 0 α**	A 0 α	E 0 α	E 0 α	**I 2 2**
4	**A 0 α**	**A 0 α**	A 0 α	**F 1 0**	G 2 2	**I 2 2**
5	B -2 0	**C 1 0**	**D 2 2**	**E 0 α**	E 0 α	H 0 α
6	B -2 0	**C 1 0**	**D 2 2**	**F 1 0**	**G 2 2**	H 0 α
7	B -2 0	**C 1 0**	D 2 2	**E 0 α**	E 0 α	**I 2 2**
8	B -2 0	**C 1 0**	D 2 2	**F 1 0**	G 2 2	**I 2 2**

NATO Y	2	4	6	11	12	13
1	**A 0 α**	A 0 α	A 0 α	E 0 α	E 0 α	H 0 α
2	**A 0 α**	A 0 α	A 0 α	**F 1 0**	**G 2 2**	H 0 α
3	**A 0 α**	A 0 α	A 0 α	E 0 α	E 0 α	**I 2 2**
4	**A 0 α**	A 0 α	A 0 α	**F 1 0**	G 2 2	**I 2 2**
5	B -2 0	**C 1 0**	**D 2 2**	E 0 α	E 0 α	H 0 α
6	B -2 0	**C 1 0**	**D 2 2**	**F 1 0**	**G 2 2**	H 0 α
7	B -2 0	**C 1 0**	D 2 2	E 0 α	E 0 α	**I 2 2**
8	B -2 0	**C 1 0**	D 2 2	**F 1 0**	G 2 2	**I 2 2**

identified manually, this will not be possible in more complicated models. Computer-assisted tools will have to be used and could be developed in the future for this purpose.

What are the limits of the applicability of models of this kind, except for the technical limits of the type previously mentioned? First and most importantly, it has to be stressed what so far has been assumed tacitly: all parties have to use the same game tree for their analyses. They can either do this by talking to each other—without agreeing beforehand on a solution—or they know each other (i.e., their strategies, objectives, and outcomes) so well that they arrive independently at the same result. This is indeed a very strong assumption that may not be fulfilled in many cases.

It should be acknowledged that there are game-theoretical techniques to deal with these kinds of problems. These so-called games with incomplete information (see e.g., [8], which takes into account the lack of information about the adversaries' preferences), can be solved only under rather restrictive assumptions. Thus, they do not really offer solutions to these problems. Nevertheless, they do indicate a direction for the future.

As our examples convincingly indicated, there may be many equilibria, none of which may be preferred to another. It should be mentioned in passing that the backward induction procedure selects all subgame perfect equilibria. However, in general it does not significantly reduce the total number of equilibria. There are further selection concepts that do not really help. Frequently, one defines a solution of a game as a unique equilibrium point, either because it is the only one or because it is selected for good reasons. With this understanding one may say that most of the vector-valued games do not have any solution. What can be said in cases like these?

Clearly, it is counterproductive to ask for more information in the sense of more criteria. We saw that in this process the number of equilibria increases rather than decreases. Useful information is a knowledge of factors describing the relative weights of criteria so that the players' payoffs can be formulated as scalar. However, if one takes into account that these factors also have to be known to the other side, one realizes that this is a strong requirement indeed.

Even worse, if the analyst presents the equilibria to the decision maker and the latter chooses one of them without any real justification, and if at the same time this happens on the other side, then both may arrive at the worst possible outcomes. The mere fact of several or many equilibria may represent a form of complexity of the problem that does not have a simple solution, that is, which may no longer be treated successfully with quantitative tools of the kind discussed here.

Sometimes, one can only conclude what does not happen in equilibrium. Let us just mention that in the case of our Kosovo example, we showed that a ground war initiated by NATO was excluded in equilibrium.

Game-theoretical models, in particular, extensive form games with vector-valued payoffs, do represent useful tools for the analysis of negotiations of international conflicts extending over some time. In any case, the results of the analysis can provide valuable information that should be taken into account by decision makers. These models, however, do not represent recipes that tell decion makers precisely and under all circumstances what to do in complex situations.

References

1. Avenhaus R (2001) Game theory and the elements of negotiation processes. PIN Points Network Newsletter 16:4–5
2. Borm P, Tijs S, van den Aarssen J (1998) Pareto equilibria in multiobjective games. Math Meth Oper Res 60:302–312
3. Keeney RL, Raiffa H (Last Edition) Decisions with multiple objectives: preferences and value tradeoffs. Wiley & Sons, New York, USA
4. Kremenyuk VA (ed 2002) International negotiation: analysis, approaches, issues. 2nd edition. Jossey-Bass Inc. Publishers, San Francisco, CA, USA
5. Krieger Th (2003) Nicht-kooperative extensive n-Personenspiele mit vektorwertigen Auszahlungen. Ph-D-Dissertation, University of the German Federal Armed Forces Munich. A shortened version of this dissertation has been published under the title, On pareto equilibria in vector-valued extensive form games [in German]. Math Meth Oper Res 58:449–458
6. Kuhn HW (1953) Extensive games and the problem of information. In: Kuhn HW, Tucker AW (eds) Contribution to the theory of games I. Princeton University Press, Princeton, JN, USA
7. Lund M, Mehler A (1999) Peace-building and conflict prevention in developing countries: a practical guide. CPN Guidebook, Brussels, Belgium
8. Morrow JD (1994) Game theory for political scientists. Princeton, Princeton University Press, Princeton, NJ, USA
9. Myerson RB (1991) Game theory: analysis of conflict. Harvard University Press, Cambridge, MA, USA
10. Nash JF (1951) Non-cooperative games. Ann Math 54:286–295
11. Neuhold, HP (ed 1989) Krise und Krisenmanagement in internationalen Beziehungen [in German]. Steiner, Stuttgart, Germany
12. Reiter E (ed 2000) Der Krieg um das Kosovo 1998/1999 [in German]. v. Hase & Koehler, Mainz, Germany
13. Reiter E, Selten R (eds 2003) Zur Lösung des Kosovo-Konfliktes. Die Anwendung der Szenario-Bündelanalyse im Konfliktmangement [in German]. Nomos, Baden Baden, Germany
14. Rohatgi VK (1976) An introduction to probability theory and mathematical statistics. Wiley & Sons, New York, USA
15. Selten R (1975) Reexamination of the perfectness concept for equilibrium points in extensive games. Int J Game Theory 4:25–55
16. Selten R (1999) Game theory and economic behavior. Vol I, 291–325, Edward Elgar, Cheltenham, UK
17. Shapley LS (1959) Equilibrium points in games with vector payoffs. Nav Res Log Quarterly 6:57–61
18. Steuer RE (1985) Multiple criteria optimization: theory, computation and application. Wiley & Sons, New York, USA
19. Yu PL (1985) Multiple-criteria decision making. Plenum Press, New York, USA, London, UK

Bridging Games and Diplomacy

Michel Rudnianski[1] and Hélène Bestougeff[2]

[1] IIM–CNAM, Paris, France
[2] CODATA, Paris, France

1 Introduction

The gap between models and reality is an issue covering almost every field of human knowledge, from physics to linguistics, and from economics to biology. On the one hand, quantitative people are mainly concerned with developing tools and consistent theories, even at the price of neglecting aspects of reality that do not fit in the frame or would lead to technical intricacies that might impede well-stated conclusions. On the other hand, observers and practitioners immersed in the "real world," tend to ignore quantitative tools and theories for reasons ranging from lack of understanding to hostility.

This is particularly true when one considers the field of negotiations, where behaviors are deeply rooted in the complexity of human nature and decision making is related to power.

To bridge the gap it is necessary to revisit models and their relationships to the real world, not only in terms of relevance and tractability but also in terms of self- and mutual perceptions on both sides.

More often than not, decision makers do not, in practice, base their choices on optimization techniques but on the fact that these choices seem good or bad. Based on that observation, the model presented in this chapter uses a bounded rationality approach based on acceptability thresholds. This approach offers three advantages:

- It requires only coarse information about the states of the world.
- It is quite handy to use, even for nonspecialists (although the theory is rooted in complex mathematical developments), since it enables an easy distinction between acceptable and unacceptable states of the world for the negotiating parties.
- It is relatively robust.

Moreover, it will be seen that under some conditions, this approach, termed games of deterrence (as it was originally developed for analyzing nuclear deterrence before being applied in several domains such as traffic control and business process modeling), does not degrade the quality of decisions but, on the contrary, may refine standard game theory approaches such as the Nash equilibrium.

Negotiations will be considered as a sequence of interactions between negotiating parties modeled by games of deterrence. Each party may take the initiative of a unilateral action and/or react to an initiative from another party. We shall consider two levels of interactions:

a) Sets of requests and answers to those requests.
b) Sets of unilateral actions or threats that affect the negotiation process.

The first level of interaction will be illustrated by the Icelandic fishery dispute, through which matrix games, and the core concept of strategy playability, will be introduced. We will then switch to the second level, by extending the concepts to sequential and fuzzy two-player games. We will also introduce N-player games for analyzing coalitions and dynamic games in the form of multistage games—to develop the concept of diplomacy that will be illustrated in the particular case of the Spratly Islands dispute. Finally, the chapter will present a path toward cooperation between game theorists and negotiators.

2 Modeling the Real World

Negotiation is a process of social interaction and communication. A variety of approaches and modeling paradigms or archetypes of negotiation process have been developed among which are *prescriptive* studies, concerned with procedures for achieving good outcomes, and *descriptive* studies, focused on understanding how people negotiate ([1], [3]).

Defining the negotiation stages precisely depends on the viewpoint emphasized. This chapter is more concerned with the sequencing of computational aspects of negotiations. The main distinction is made between unilateral actions (first stage) and interactions (second stage), which slightly differs from the three-stage representation proposed by Faure and Zartman who focus more on interactions between the parties.[3]

More precisely, the following characterization is taken:

1. A prenegotiation stage which includes formulation and analysis of the negotiation problem, the context of the problem, and knowledge about the participants.
2. A negotiation stage where offers and objectives take place.

The second stage is where a real formal model can be developed. This implies that some behavioral studies have been conducted in the first stage, especially in natural language communication. Moreover in an international context, individual differences and social influences must be taken into account, as they can easily lead to misunderstanding.

[3] These three stages are: the prenegotiation stage, which aims at reframing the problem to meet acceptance; the potential agreement formula stage, which deals with the final overall architecture package; and the stage for finding an acceptable balance, which deals with details and precise figures on each basis issue.

Linguists generally describe a language at four levels: morphology, syntax, semantics, and pragmatics [10]. If formal models of morphology and syntax are rather operational, semantics and pragmatics are still difficult to process. Recall that pragmatics is the way a sentence is interpreted in a given practical context. For instance, the sentence “it is pretty cold” could mean an unfriendly atmosphere in a negotiation room or a freezing temperature outside.

In a primitive exchange where a single action is required, a simple—nonambiguous—sentence, is uttered. However, in a negotiation process, ambiguity can be a trick to gain time or force new information to emerge. Wording is never innocent. Formalizing “a state of affairs,” so that sentences in natural language are translated into a logical expression, is a strong reduction process. Just think of such words as: some, several, and most, which do not easily translate into standard quantifiers. But still, analysis of a negotiation problem must lead to a clear formulation of offers, objectives, and utilities.

Too often, the needs and capacities of practitioners are not totally taken into account, while the transition and its inherent limitations from the reality to formal models are not understood.

Let us consider a simple two-party negotiation which can be described by:

- A set of feasible offers;
- A set of objectives resulting from explicit rules that each party can apply on offers;
- A set of utilities, which allow a quantitative appraisal of the objectives.

Immediately, practitioners will raise questions such as:

- Is the process a “one-shot” procedure (in which case time is not considered)?
- What if time modeling is needed?
- Are the players playing simultaneously, or one after the other?
- Is knowledge time-dependent or not?
- How are offers, objectives, and utilities defined?
- What about consistency? Can, for instance, a player have two inconsistent objectives?
- How are natural language subtleties taken into account? For instance, can all the following sentences: “I go to war,” “I may go to war,” “I should go to war,” “I will go to war,” “I don’t think that I won’t go to war,” be all reduced to one sentence?
- How are the rules devised to calculate objectives? How do they express the intuitive idea of strategy? Can they be expressed by a computable function or an algorithm? If so, is it tractable (i.e., is the computation time to get a result “reasonable”?)[7].

All these questions are legitimate, and the list is certainly not exhaustive. Let us elaborate on two elements, which have been extensively studied but are still under discussion: logical consistency and time.

Since Aristotle it has been important to avoid inconsistencies, considered as a source of cognitive disaster: it is not possible to assert that a sentence or a state of affairs is both *true* and *not true*. Admitting simultaneously "to be at war" and "not to be at war" is at first sight impossible, as our world is consistent and so is the knowledge we have about it. Therefore contradictions seem impossible in a formal system. But with a closer look, such a conclusion seems too simplistic, as obviously nature is full of contradictions. Avoiding them reduces the empirical adequacy of the formal systems. Mathematicians and logicians have studied this problem and proposed some solutions.

On the one hand, introduction of modal logics and Kripke possible worlds semantics [5] enables us to consider truth values with respect to "different worlds." Thus, a proposition can be *false* in one world and *true* in another. Hence, notions such as *necessary* is a true proposition and *possibly* a true proposition. But, although defining those notions is rather easy and intuitive, achieving "good" formal properties based on them might be a different story [13].

On the other hand, plausible reasoning, which can be related to probabilities, is another way of handling apparent inconsistent situations [12].

One can also work with locally consistent sets, which do not exclude general inconsistencies. For instance, sets of offers and objectives can be attached to a given time period.

Another source of inconsistencies is evolution of knowledge over time. Having new information may lead to new offers and objectives. But these can contradict existing offers and objectives. How should these sets be updated ?

Now, the abstract concept of time may be difficult to grasp. For human beings the evolutionary development of the sense of time has certainly followed a path analogous to that followed by children. But nowadays it is clear that it very much depends on cultural environments. One will not bargain the same way in an African market as in a Chinese one. The outcome might not be as important as the path taken to reach it. One can prove, for instance, that in a simple bargaining situation, a limited number of offers and counteroffers is theoretically sufficient to get an optimized result, but if this number of exchanges is not "culturally" acceptable, no bargaining is possible.

Words like "yesterday," "today," or "tomorrow," point out notions that do not necessarily refer to precise metrics, as, for example, in the expression "the world of tomorrow."

Some researchers have suggested distinguishing five temporalities [6]:

- *Nootemporality* or *noetic time*: temporal reality of the mature human mind.
- *Biotemporality* or biological time: temporal reality of living organism as far as its biological functions are concerned.
- *Eotemporality* or the physicist's time: the simplest form of continuous time.
- *Prototemporality* or the world of elementary particles where precise location of instants has no meaning.
- *Atemporality*: absence of time.

In international negotiations, objectives are generally temporally marked. The first difficulty is to choose the right time abstraction and then see how to deal with the fact that it should be the same for all parties, given that differentiation at this level would introduce unsolvable theoretical problems.

These remarks show that any kind of model should be elaborated under close cooperation with practitioners. Basic concepts should be clear, easily understandable, and known limitations of the model clearly explained in advance to avoid disappointment.

3 A Limited Rationality Approach

3.1 Trading optimization for threshold

Optimization is a process to obtain the "best" solution to a given problem. The exact meaning of "best" depends on the particular problem but also on who is optimizing.

Now, in "real life," situations may occur where optimization might not always be appropriate (see for instance, [17], [18]).

First, because mathematical intricacies arise as soon as optimization deals with several variables or when functions exhibit nonsmoothness features. These intricacies may be such that one cannot solve the optimization problem and must be satisfied by stating some properties of a potential solution.

Second, because the time required for getting to the solution may be unacceptable.

Two examples may illustrate this fact.

First, during the Vietnam peace talks, President Nixon noted that when the North Vietnamese party made a proposal, United States (US) specialists asked for an amount of time which he found unreasonable to analyze the proposal, as during that time, the military situation on the battlefield was quickly evolving, making the proposal obsolete. That is what Richard Nixon called "crisis by analysis."

The second deals with an early warning against a nuclear attack. On the one hand, space contains a lot of debris, while an enemy missile carrying a nuclear load may protect itself against a defensive counter–missile strike through deploying a vast number of decoys. On the other hand, the defender needs to anticipate the attacker's trajectory, to know in which direction to send its missiles. Because of electromagnetic rays, false images may appear on the defender's radar. Add to that the more sophisticated the early warning system is, the higher the probability of an error, and you might get an idea of the computing complexity needed to deal with the situation in a matter of minutes. It may happen that optimization requires more time than is available, and yet a decision has to be taken on whether to do nothing while under the risk of a nuclear attack, or to launch a counterstrike that might be unjustified and lead to dramatic consequences for both sides.

More generally, information required for proper optimization is not always available to the decision-making process [9].

Simon [20] precisely considers that, given these difficulties, optimization is out of reach. He therefore suggests a bounded rationality, which trades optimization for "satisficing" results.

Other difficulties may arise that are not directly linked with computational problems.

Decision makers may not be absolutely sure about which variable they want to optimize, or optimization cannot be reduced to a variable or variables set optimization.

Thus, an optimization problem may deal with variables that are by essence not quantifiable (for instance, when deciding which of several pieces of art is the most beautiful). Similarly, a simple observation of everyday life shows that decision makers do not always think in terms of optimization. More often than not they base their decisions on a good-or-bad type of appreciation of the decision consequences [2]. Very deeply rooted and widespread cultural factors may lead decision makers to dichotomize: in most human cultures, people distinguish between good and evil, right and wrong (and here the rigor of mathematics seems to add value to that attitude). Personal positioning may also impact the assessment system.

Dichotomy is particularly important in international relations, when a decision deals with whether to go to war or not, to sign an agreement or not, etc. In most cases optimization could clearly provide an answer, but for various reasons its use seems to be unnatural. It seems appropriate therefore to develop the "good-or-bad" approach used by real decision makers, knowing that this "rough" way of looking at the world might later be refined.

More precisely we shall first consider that each state of the world is seen by a decision maker as either "acceptable" or "not acceptable," the limit of acceptability being defined by some threshold value. To decide whether a state of the world is acceptable requires only rough information. A decision maker will be considered rational if he devotes all his efforts to reaching a state of the world that is acceptable to him.

Trading optimization for thresholds may alleviate some constraints established by standard game theory on decision makers, assumed to be "superrational." If it does not eliminate problems of common knowledge, it requires less information about rationality: at a first level, all I need to know about the other guy's rationality is that he prefers good to bad.

Of course, the question arises: is what I consider good *for* him, considered good *by* him? But by then we have already left the banks of common knowledge. The matter here is the quality of my information. No doubt it is an important question, but mistakes are less costly than in standard games: all that I have to know is on which side of the threshold he is, not whether he is far from or near to that threshold.

3.2 An introductive example: The Icelandic fisheries dispute

When the dispute between Iceland and the United Kingdom (UK) began in 1958, each party had a claim:

- Iceland wanted to extend its territorial limits from 4 to 12 miles offshore.
- The UK wanted to be able to fish in the 4–12-miles zone.

A solution satisfying both parties was obvious: the UK recognizes the new Icelandic territorial waters, which in turn lets British fishermen fish these waters. Very soon it was clear that what was unacceptable to the parties was a unilateral advantage, while both could do with a situation where no demand would be satisfied.

So interaction can here be modeled by the very simple game shown in *Table 1*, where "accepts" and "refuses" refer to the demand made by the other side.

Table 1. The Icelandic fisheries dispute, version 1

		ICELAND	
		b_1 = accepts	b_2 = refuses
UK	a_1 = accepts	(1,1)	(0,1)
	a_2 = refuses	(1,0)	(1,1)

One can see that by refusing the other party's demand, each player guarantees himself an acceptable outcome. Therefore, refusal strategies will be termed *safe*.

On the other hand, acceptance strategies bear the risk of giving the player who selects them an unacceptable outcome. Therefore, they will be termed *dangerous*. Since all that players want is an acceptable outcome, one expects that they will choose a safe strategy. Indeed, suppose the UK accepts the Icelandic claim, in order to remain on friendly terms with Iceland. If Iceland does not in turn let the British continue to fish, the UK will be in an unacceptable position. There is no rationale for the UK to take a bet that Iceland will accept its demand. In other words, refusal strategies seem to be the only *playable* ones.

That is exactly what happened. Each refused the other side's demands.[4]

If *playable* is here equivalent to *safe*, and *no-playable* to *dangerous*, it is not the case in general. Assume for instance that refusal of the other party's claim meant military escalation, and furthermore that for the UK such an escalation was acceptable because of its assumed naval superiority, while it was not for Iceland. The game is then represented by the matrix of *Table 2*.

Table 2. The Icelandic fisheries dispute, version 2

		ICELAND	
		b_1 = accepts	b_2 = refuses
UK	a_1 = accepts	(1,1)	(0,1)
	a_2 = refuses	(1,0)	(1,0)

[4] The game is a binary version of the Prisoner's Dilemma, and the simple reasoning developed suffices to reach the solution of the Prisoner's Dilemma. Indeed, interpreting the above results in terms of standard games, shows that the selected strategic pair is one of the two Nash equilibria (i.e., no player has an interest in unilaterally changing his strategy). In other words, resorting to the simple reasoning developed here above, leads to a solution that can be considered as "refine" with regard to the set of standard solutions. We shall come back to that later.

The two strategies for Iceland are equivalent, and lead to an unacceptable outcome if the UK refuses the Icelandic claim. As in a game a player *must* play, Iceland could choose any strategy, despite the fact it would possibly lead to an unacceptable outcome. This would, in turn, reinforce the UK's decision to refuse Iceland's claim.

The consequence was the occurrence of several incidents involving gunfire between UK trawlers and destroyers on the one side and Icelandic coastguards on the other.

3.3 Games of deterrence: Basic properties

The example above paves the way for definitions and properties of *games of deterrence*.

Thus, in the first version of the dispute, "playable" means "good," while in the second, it means "I had to do it because I had no better choice." Thus, a distinction needs to be made between:

- A *positively playable* strategy that guarantees the player who selects it an acceptable outcome, provided the other player is rational.[5]
- A strategy *playable by default*, chosen by a player who has no positively playable strategy.

A strategy will be considered *playable* if it is either positively playable or playable by default.

In the second version, accepting the Icelandic claim was not a playable strategy for the UK. Suppose Iceland refuses the presence of British fishermen. Iceland could do this, as refusal is a playable strategy (although by default). By accepting an extension of Icelandic territorial waters, the UK would get an unacceptable outcome, except for the case (not considered here) where a binding agreement would take place. Thus, determination of a strategy's playability is a recursive process, as soon as these strategies are dangerous: acceptance by the UK of an extension of Icelandic territorial waters is not playable, because Icelandic refusal of British claims is playable.

To get a clearer criterion for deciding whether a dangerous strategy is playable or not, let us introduce the concept of deterrence as follows. Given a strategic pair, we shall say that a strategy for the UK is *deterrent vis-à-vis* a strategy for Iceland if the three following conditions apply:

1. The British strategy is playable.
2. Implementation of both strategies leads to an unacceptable outcome for Iceland.
3. Iceland has an alternative strategy that is positively playable.

Condition 3 means that when you try to put pressure on the other party, you should always leave it with an acceptable alternative, for you can expect everything from someone who has nothing to lose.

[5] In the sense that his behavior is consistent with his aim of getting an acceptable outcome.

It has been shown, that playability and deterrence are two sides of the same coin. More precisely, a strategy for Iceland is playable if and only if there is no British strategy deterrent vis-à-vis it [14].

An *equilibrium* of the game is any playable strategic pair. If both strategies of the pair are positively playable, we shall speak of a *positively playable* equilibrium. In the first version of the game, there is only one equilibrium, for which each party refuses the other side's demand. This equilibrium is positively playable.

We define in a similar way a *playable by default* equilibrium.

It follows that every matrix game of deterrence has one equilibrium at least.

It has been shown [14] that a positively playable equilibrium is a Nash equilibrium: thus, in binary games, positively playable equilibria are a "refinement" of Nash equilibria. In its first version, the game has two Nash equilibria (both parties accept the other side's demand; both refuse) while there was only one equilibrium in the game of deterrence (both parties refuse).

3.4 Assessing strategies' playability

In the above example, developing the recursive process to assess the strategies' playability can be done straightforwardly. Now, two different, albeit related, kinds of problems can arise.

First, by developing a purely intuitive reasoning, one can make mistakes, for instance, find one *solution*[6] where there are two. Consider a version of the Icelandic fisheries dispute in which there is no military escalation unless both parties refuse the other side's demands, the reason being that in the case of an asymmetric situation, a side payment is made by the party that refuses to compensate for the other party's acceptance. Consider furthermore that for both parties, the only unacceptable situation is military escalation. The new matrix game then is the one shown in *Table 3*.

Table 3. The Icelandic fisheries dispute, version 3

		ICELAND	
		b_1 = accepts	b_2 = refuses
UK	a_1 = accepts	(1,1)	(1,1)
	a_2 = refuses	(1,1)	(0,0)

It is safe for each party to accept the other party's demand. Now, with the alternative strategies—leading to military conflict—things are a little different. Assume that refusal is not playable for Iceland. Then obviously it is playable for the UK, and vice versa. That means that the game has two solutions:

1. Both UK strategies are positively playable; only acceptance is positively playable for Iceland (which means that refusal is not playable).

[6] By solution we mean an assessment of each strategy in terms of playability.

2. Only acceptance is positively playable for the UK, while both strategies for Iceland are playable.

If the UK thinks that the first solution will be adopted, while Iceland thinks that the second will be adopted, both might refuse and hence begin a military escalation. Thus, a multiplicity of solutions may lead to misperception with serious consequences.

Another reason why multiplicity of solutions should be analyzed is that sophisticated players may use it to maintain ambiguity as long as possible during the negotiation process. As long as no strategy implementation has been undertaken, a player can use ambiguity to exert pressure on the other. Thus, the UK, by letting Iceland know that refusal is a playable strategy, might deter the latter from refusing British fishing boats into its "new territorial waters," which, in turn, will enable the UK to keep on fishing in these waters[7] while refusing an extension of Icelandic territorial waters.

The second kind of problem arising with assessment of strategies playability is intricacies.

As soon as one leaves the banks of 2x2 games, the recursive process may develop exponentially, with the consequence that after a while, even the best minds can get lost.[8]

A general method is needed. Therefore, we shall associate with each strategy s a binary *positive playability* index J(s). For instance, in the last game, the two solutions can be rewritten as follows:

1. $J(a_1) = 1$; $J(a_2) = 1$; $J(b_1) = 1$; $J(b_2) = 0$
2. $J(a_1) = 1$; $J(a_2) = 0$; $J(b_1) = 1$; $J(b_2) = 1$

We shall also associate with each player P a binary *by default playability index* j_P.

Positive playability indices and, by default, playability indices are linked at two levels:

1. $j_p = 1$, only if all player P's strategies, positive playability indices equal 0

[7] *Non-dits* might be central in a negotiation process. Sometimes an agreement can be reached only because of ambiguity. Each party not only gives a different meaning to the terms of an agreement, but knows that this ambiguity exists. A good example is the Akaba meeting, when Israeli Prime Minister Ariel Sharon agreed that Israel should withdraw from "illegal settlements," while on the other hand Prime Minister Mahmoud Abbas promised to "fight" terrorist attacks on Israeli civilians. No precision was given as to what "illegal settlements" or "fight" meant. It is most probable that these precisions could not be given because those terms mean different things for each of the parties; moreover, both parties knew that insisting at this stage on giving precise definitions would have certainly led to a failure and hence damaged the dynamics of negotiation.

[8] At first sight, such an observation seems to speak against the method proposed here because if game theory is to bring something to negotiations, it is certainly not by complicating practitioners' lives. But complication here does not stem from the approach followed (games of deterrence) but from the intrinsic number of possibilities that can occur in "real life," which, in turn, requires a general method to be dealt with.

2. Given a strategic pair (a,b) associated with outcome pair (0,x) where x is some binary number, J(a) equals 0 as soon as j_P equals 1.

Combining these two conditions leads to a binary equations system called the *playability system* [14].

3.5 Graphs of deterrence

When the number of strategies is large, directly solving the playability system may be lengthy. Therefore, it may be interesting to look for a categorization of games of deterrence, such that games of a given group display the same structural properties, associated with some specific solution set.

Let us associate with each matrix game a *graph of deterrence*, characterized as follows:

1. The graph is oriented: each arc has a definite origin and a definite extremity.
2. The origin and the extremity of an arc are strategies of different players.
3. A strategic pair (a,b) can be associated with an arc of origin a (respectively, b) and extremity b (respectively, a) if and only if (a,b) can be associated with the outcome vector (x,0) (respectively [0,y]) where x (respectively, y) is a binary number.

For instance, in the first version of the game, strategic pair (a_2, b_1) defined by the UK refuses while Iceland accepts, is associated with outcome pair (1,0). Thus, there exists an arc of origin a_2 and extremity b_1. Similarly, we can define an arc of origin b_2 and extremity a_1. No other arc can be built. So the graph associated with the game is:

UK does not accept $\rightarrow$ Iceland accepts
Iceland does not accept $\rightarrow$ UK accepts[9]

This graph is comprised of two paths that we shall respectively call:

1. An A path, that is, a path the root of which is a strategy of player A (in this case the UK).
2. A B path, that is, a path the root of which is a strategy of player B (in this case Iceland).

In the third version of the game, the graph comprises a loop:

UK does not accept $\leftrightarrow$ Iceland does not accept

A loop is the simplest example of graphs with no root, which we shall call G graphs.

Thus, there are three types of elementary graphs. It can be shown [14] that each of these elementary graphs displays some structural properties associated with the solutions of the corresponding matrix game of deterrence:

[9] Remember that the arrow here does not have the meaning of an implication but represents an arc in a graph of deterrence. Consider, for instance, the first arrow. It means that if Iceland does not accept British boats in its new territorial waters, the UK will get an unacceptable outcome if it accepts the extension of Icelandic territorial waters.

1. In a game of deterrence, the graph of which is an A path, the only positively playable strategy is the root (strategy of A), while all strategies of B are playable by default.
2. Games associated with B paths display similar properties.
3. In a game of deterrence associated with a G graph, there is at least one solution for which all strategies are playable by default (this is the only solution in the case of a loop).

It has been shown that any graph of deterrence can be broken down into connected parts, each one being an A path, a B path or a G graph. Depending on the presence of these components in the graph, one can distinguish seven types of games: A, B, G, A–B, A–G, B–G, A–B–G. This typology leads to a classification theorem which states that the game type defines the solution set: the first version of the game is of type A–B, while the third is of type A–B–G.

4 Extending the Concept of Deterrence

4.1 Sequential games

Simultaneous matrix games[10] squeeze time. But in real life, time may be a core element in the negotiation process, either to put pressure on one party or to get more information. There are several ways of reintroducing time in the decision process. Among these are sequential games, which let players exchange information and change their minds if they are not happy with the choices they have made so far. To let players make a *move*, we shall introduce rules based on *preference orders*.

A player might care for the other party. In standard games, this concern is fully taken into account by the players' utilities, and there is no interpersonal utility comparison. In games of deterrence, things are a little different, since it may be difficult to integrate into a 1 or a 0 what a player thinks simultaneously about his situation and the situation of the other party, more precisely, since:

1. Each binary pair can take four different values.
2. One can assume that in a negotiation, players are not masochists (i.e., a player will always prefer for himself a 1 to a 0).

It can easily be established that each player can have only four preference orders. If we consider the UK, for instance, these preference orders are [14]:

- $\mathrm{P}^1_{\mathrm{UK}}$: $(1,0) \geq (1,1) \geq (0,0) \geq (0,1)$ (*prudent malevolence*)
- $\mathrm{P}^2_{\mathrm{UK}}$: $(1,0) \geq (1,1) \geq (0,1) \geq (0,0)$ (*partial malevolence*)
- $\mathrm{P}^3_{\mathrm{UK}}$: $(1,1) \geq (1,0) \geq (0,0) \geq (0,1)$ (*partial benevolence*)
- $\mathrm{P}^4_{\mathrm{UK}}$: $(1,1) \geq (1,0) \geq (0,1) \geq (0,0)$ (*prudent benevolence*)

[10] That is, games in which, when selecting a strategy, a player has no information about the other player's choice.

Let us consider, for example, prudent malevolence. The UK first wants to be in a good situation and then to put Iceland in a corner. If the UK is in a bad situation, it wants Iceland to feel equally bad. Using these preference orders amounts to replacing the game of deterrence by some standard game, where each player's utility function is precisely given by his preference order. Let us consider, for instance, the first version of the game and assume that the two players are malevolent (either prudent or partial). There is only one equilibrium: each party refuses the other side's demand.

Things are more complicated with preference orders characterized by benevolence. The states of the game characterized by symmetric strategic pairs are stable: no player has an interest in moving.

Consider, for instance, the strategic pair (refuses, accepts) associated with the outcome pair (1,0). On the one hand the situation is unacceptable for Iceland, while on the other it is clearly not the best for the UK, given its order of preference. So both players have an interest in moving.

Now, what if they move simultaneously? The game is then in the state associated with a strategic pair (accepts, refuses) and an outcome pair (0,1). This state is symmetrical to the previous state, which means that both players will have an interest in moving, and so on.

If nothing else is done, the game will continually oscillate between the two above states. To avoid such a situation, we introduce an *order of precedence* which, in each state where both players have an interest in moving, gives priority to one of them. The consequences appear in the array of *Table 4*, which crosses the order of precedence with the initial state of the game.

Table 4. Precedence order and initial state of the game

Precedence/ Initial state of the game	UK gets priority	Iceland gets priority
UK accepts and Iceland refuses	UK and Iceland refuse	UK and Iceland accept
Iceland accepts and UK refuses	UK and Iceland accept	UK and Iceland refuse

The conclusion is obvious: to satisfy both party's claims, the party that initially refuses the other party's demand should change its position.

Let us finally consider the case where the UK is malevolent and Iceland is benevolent. All demands are rejected, as is the case when both parties are malevolent. The fact that one party is benevolent does not help to reach an agreement based on demands satisfaction: you need two to tango.

The process used here is the one usually developed to obtain a Nash equilibrium. One starts from an *initial strategic pair* and sees whether a player has an interest in moving from his initial choice. If the same operation is repeated for every strategic pair, all solutions are eventually found. These solutions shall be termed *final strategic pairs*.

Consider now the second version of the dispute, where both parties are malevolent. Suppose furthermore that there is no *precedence order*. If initially both players accept the other party's demand, both should move, and finally refuse.[11] If initially, only the UK accepts the other party's demand, it has an interest in moving, and finally both players refuse. For the other two possible initial strategic pairs, no player has an interest to move.

More generally, for each pair of preference orders, the players must analyze the relationship between the initial strategic pair[12] and the final one, in order to properly select the former.[13]

With each pair of preference orders, a subset of the set of final strategic pairs can be associated (in many cases this subset reduces to a single element). It is then theoretically possible to select pairs of preference orders that will form equilibria.

To summarize, we introduce here three games of deterrence:

- The game of preferences;
- The game of initial strategies;
- The game of sequential solutions.

The game of preferences acts here as a meta-game. As an illustration let us go back to the second version of the dispute, assuming there is no precedence order and the two parties are benevolent. Except for the initial situation where both parties reject the other side's demands, all initial strategic pairs lead to mutual acceptance as a final strategic pair. The game of initial strategies is then represented in *Table 5*.

Table 5. Game of initial strategies

		ICELAND	
		b_1 = accepts	b_2 = refuses
UK	a_1 = accepts	(1,1)	(1,1)
	a_2 = refuses	(1,1)	(1,0)

The game has two equilibria: in the first, both parties accept; in the second, the UK refuses while Iceland accepts.[14]

One could similarly establish that when the two parties are malevolent, the game of initial strategies matrix comprises only pairs (1,0). There are four equilibria.

The same thing happens if one party is benevolent and the other malevolent.

We can now build the game of preferences matrix (see *Table 6*).

Any strategic pair is an equilibrium in the game of deterrence, which means that preference orders do not matter in selecting initial strategies. One could argue that

[11] It could be argued that if Iceland was a *nonmyopic* player (i.e., that looks further than the next stage of the game), it should not try to move from the initial state, since the series of moves leads to an unacceptable state. But, in fact, myopia is irrelevant, since the UK has also an interest in moving from the initial state, with the consequence that in any case, the final state will be (refuses, refuses).

[12] Analogous to the conflict point in the negotiation games of Nash.

[13] Similar in a way to the selection of the pair of optimal threats in the Nash games of threat.

[14] There are three Nash equilibria in the standard game associated with the same matrix, which shows once more that positively playable equilibria are refinements of Nash equilibria.

Table 6. Game of preferences

	P_I^1	P_I^2	P_I^3	P_I^4
P_{UK}^1	(1,0)	(1,0)	(1,0)	(1,0)
P_{UK}^2	(1,0)	(1,0)	(1,0)	(1,0)
P_{UK}^3	(1,0)	(1,0)	(1,1)	(1,1)
P_{UK}^4	(1,0)	(1,0)	(1,1)	(1,1)

if the game of preferences had been considered as a standard game, the conclusions would have been different: the four strategic pairs for which the UK is malevolent while Iceland is benevolent, are not Nash equilibria.

We shall return to that after introducing the game of sequential solutions, the matrix of which will be built as follows. Given an initial strategic pair (α,β), if α is a component of the final strategic pair (α,x), then the outcome for party A is the outcome associated with strategic pair (α,x). If not, that outcome equals 0.[15] The same applies to strategies of party B.

To illustrate this, let us assume that in the second version of the Icelandic fisheries dispute both players are malevolent and there is no precedence order. If the UK initially accepts the Icelandic demand, both parties will move, which means that in the game of sequential solutions, we can associate the same outcome pair (0,0) with mutual acceptance and the case where only the UK accepts the other party's demand.

On the other hand, if the UK initially refuses the Icelandic claim, no party has an interest in moving. As the situation in which only Iceland accepts the UK's demand and the situation where the two sides refuse, are both associated with the outcome pair (1,0) in the original game, they will also be associated in the game of sequential solutions, which is represented by the matrix of *Table 7*.

Table 7. Game of sequential solutions

		ICELAND	
		b_1 = accepts	b_2 = refuses
UK	a_1 = accepts	(0,0)	(0,0)
	a_2 = refuses	(1,0)	(1,0)

It is safe for the UK to refuse, while both strategies of Iceland are playable by default. Thus, the game has two equilibria, for which the UK refuses the extension of Icelandic territorial waters.

Let us suppose first that in the level 2 game of sequential solutions built on the level 1 game of preferences, the UK's order of preference is malevolence. Then in the latter, all Icelandic strategies are playable by default, while only two strategies of the UK are playable (and safe): the two malevolent orders of preference. Then,

[15] Where outcome 0 means that the initial strategy a could not do as a final strategy for party A.

in the meta-game of level 1, the UK will necessarily choose one of these strategies, which makes irrelevant the status of strategic pairs comprised of a benevolent order of preference for the UK and a malevolent order of preferences for Iceland.

Suppose then, that in the game of sequential solutions, the UK's order of preference is benevolence. Then, only benevolent strategies are playable (and safe) for the UK and for Iceland (safe strategies of one party are deterrent vis-à-vis the nonsafe strategies of the other side). From this, we conclude that:

1. Level 1 orders of preference are the same as in level 2.
2. Although at first sight the deterrence approach looks coarser than the standard one, in reality it is just the opposite, as soon as metareasoning is involved.

Sequential games may be of the utmost importance, not only during the negotiation process, but also once an agreement has been reached, in order to check the parties' compliance [14].

Consider for instance that an agreement has just been signed, which gives the UK permission to fish during some periods of the year. Several months after the signature, the UK wants to be sure that Iceland will fully comply. Verification can be modeled by the matrix of *Table 8*, where:

- A means that Iceland accepts British fishermen in its new waters, and $\underline{A}$ that it does not.
- B means that the UK believes what Iceland says, and $\underline{B}$ that it does not.

Table 8. The verification game

		ICELAND	
		A	$\underline{A}$
UK	B	(1,1)	(0,1)
	$\underline{B}$	(1,0)	(1,1)

Interpretation of the matrix is as follows:

- All cases are acceptable for Iceland except when it accepts and the UK does not believe it.
- All cases are acceptable for the UK except if Iceland does not accept and the UK thinks it does.
- On the whole there is only one equilibrium: ($\underline{B}$,$\underline{A}$).[16]

The sequential process may help in analyzing the parties' sincerity. Assume for instance that:

- The parties' initial declaration corresponds to the pair (B,A).

[16] There are two Nash equilibria, (B,A) and ($\underline{B}$,$\underline{A}$), associated with the same outcome pairs. Deterrence analysis shows that Iceland and the UK will choose the second one because they do not trust each other: the agreement is then dead.

- Iceland is not sincere (i.e., its real choice was not to accept).
- Iceland's preference order is prudent malevolence.
- The UK's preference order is partial benevolence.
- Each party knows the other party's preference order.

Since Iceland's initial choice was the best for her, she has no reason to change. The UK, thinking that the initial strategic pair is (B,A), has no reason to change either, while she is expecting Iceland to change. As this change does not occur, the UK realizes that Iceland did not tell the truth.

4.2 Fuzzy games: Beyond good and evil

Until now, to the question, "Is this strategy playable (positively or by default)?", the answer was either "Yes" or "No." Now there are gray areas in life, for which answers such as "more or less" would be more appropriate than simply "Yes" or "No."

Such answers enable us to refine the strategies' ordering, which is especially interesting when strategies are playable by default. Perhaps allowing the playability system to have *fuzzy* solutions (i.e., solutions such that playability indices may take any value between 0 and 1 and not just 0 or 1) will facilitate the players' choices.

Consider that in the Iceland fisheries dispute each party can at the same time make demands and accept or refuse the other side's demands. Suppose the game matrix is the one of *Table 9*, where D stands for demands and A for accepts.

Table 9. The Icelandic fisheries dispute, version 4

		ICELAND		
		b_1 = D and A	b_2 = D and ¬A	b_3 = ¬D and A
	a_1 = D and A	(1,1)	(1,0)	(1,1)
UK	a_2 = D and ¬A	(0,1)	(1,1)	(1,0)
	a_3 = ¬D and A	(1,0)	(0,1)	(1,1)

The graph of deterrence is the A path: $a_1 \rightarrow b_2 \rightarrow a_3 \rightarrow b_1 \rightarrow a_2 \rightarrow b_3$.

In the nonfuzzy game, this implies a unique solution characterized as follows:

- The safe strategy a_1 (accept extension of Icelandic territorial waters and claim fishing rights in these waters) is the only playable strategy for the UK.
- All strategies of Iceland are playable by default.

It has been shown [16] that fuzzyfication[17] produces a "furthering effect":

- Playability of the UK's strategies decreases along the path.
- Playability of Iceland's strategies increases along the path.[18]

[17] That is, letting playability indices possibly take any values between 0 and 1.

[18] It has also been shown that both sequences of playabilities would converge toward 0.5 if the path length tended toward the infinite, which is of course not the case here.

Here: $J(a_1) = 1$; $J(a_3) = .48$; $J(a_2) = .36$; $J(b_2) = 0$; $J(b_1) = .25$; $J(b_3) = .31$

The UK will accept an extension of Icelandic waters and claim access to these new waters, while Iceland will select the furthest strategy from the root, that is, it will have no claims but will accept British ones. Both players will thus guarantee themselves an acceptable outcome.

The conclusion may look surprising, since it means a surrender of Iceland to the UK, while the UK was ready to accept extension of Icelandic waters. The reason stems from the fact that, for Iceland, the status quo is acceptable. If not, $J(b_3)$ vanishes and the strategy which maximizes the positive playability index is b_1 (Iceland demands an extension and accepts the British claim).

That is exactly what happened: the issue was raised at the United Nations Law of the Sea Conference in 1960, before being directly discussed between the parties. An agreement was reached in 1961, by which the UK accepted an extension of Icelandic territorial waters, while Iceland provided special waivers for the UK to fish in these new waters for several months each year.

4.3 Dealing with more than two parties

Bilateral negotiations represent the simplest type of negotiations. Furthermore, when the players' strategies can be numbered,[19] using matrices is quite handy. But using matrices in N player games is not possible in general.[20] One has to resort to a file describing the correspondence between the strategic and outcome N-tuples. The various playabilities introduced in the two-player case can be extended to the N-player case.

In noncooperative games, the playability properties of player K's strategies can be obtained by considering that K plays with a super-player comprising the N-1 other players, whose strategies are playable if and only if all strategies of the N-1 players are playable.

One must also consider the case where some parties form a coalition (i.e., not only form a single player in the eyes of the other parties but also in their own eyes).

We can reasonably assume that a state of the world is unacceptable to a coalition if it is unacceptable to at least one party in the coalition. In other words, the coalition's outcome is the product of the outcomes of all parties that are members of the coalition.

Let us now apply these concepts to the example of the Spratly archipelago conflict, already analyzed by Brams and Taylor [4].

---

[19] That is, when each strategy can be associated with a number, using matrix formalism is quite handy.

[20] It is done sometimes with three simple player games where the number of strategies of at least one player is limited, such that a particular matrix is associated with each strategy of that player.

4.4 The context of the Spratly conflict

Disputants are Brunei, China, Malaysia, the Philippines, Taiwan, and Vietnam. After several skirmishes with Vietnam, whether in the Spratly or the Paracels islands, and above all its takeover of Mischief Reefs, China does not want to be subject to hostile reactions that would make it the black sheep of the region.

Furthermore, China is presently not sure that it has the military strength to support a protracted conflict in the archipelago, the northern border of which is over 600 miles south of continental China, and that encompasses more than 170 islands, reefs, or rocks, scattered over an area of nearly 800,000 square kilometers. All-out military operations can be discarded. For obvious reasons the same can be said of the other parties.

Moreover, a conflict escalation would be prejudicial to the freedom of navigation on one of the world's most important maritime routes. It would trigger intervention of third parties, such as the USA, which China wants to keep out of the zone. Nevertheless, China is not prepared to step down from its claim of total sovereignty over the South China Sea, for nationalist, economic, and strategic reasons. The region is rich in natural resources, namely oil and gas, which are of utmost importance for China's fast-growing economy. Taking control of these resources would obviously be beneficial to China because it would be less dependent for its energy supply, and also from a financial point of view. This would also trigger as a positive feedback the allocation of more financial resources to military expenditures, giving China in the long run the capacity to project military forces in the region, sustain a protracted conflict, and seize more natural resources.

China has tried until now to make these apparently contradictory objectives consistent, by paying lip service to cooperation with the Association of Southeast Asian Nations (ASEAN), while simultaneously refusing to engage in a binding code of conduct, downplaying the conflict, and postponing its settlement.

Other disputants such as Vietnam and the Philippines, considering the dispute as important for them, and probably understanding the reasons behind the Chinese attitude, would like a quick settlement. The Declaration on the Code of Conduct signed in 2002 seems to exclude military skirmishes in the near future, but does not consider other possible moves.

4.5 A simplified model

After having broken down the archipelago into five regions, Brams and Taylor [4] use the Adjusted Winner's approach to reach a share of these regions between two parties, China and ASEAN. The explicit assumption behind the model is that both parties consider that keeping the status quo is not in their interest, and therefore wish to settle the dispute by a fair division.

In the sequel we shall relax this assumption and suppose in the light of previous skirmishes that parties may take unilateral steps, based on the present situation and the power ratio.

Moreover, we shall consider that ASEAN should not necessarily be regarded as a single party, as divergences of interest that occurred in the past between its members show that the latter may, a priori, adopt different behaviors vis-à-vis China.

Nevertheless, to focus on principles and to avoid computational developments, we shall consider a simplified interaction, in which Brunei having few claims, except the respect of its continental shelf near Louisa Reef, and no real naval force, can be discarded as a player. The same can be done with Malaysia, which controls no island, reef, or rock, while its naval force does not have the capacity to engage in important skirmishes.

Taiwan and China, even though opposed on several issues concerning the dispute (like the Macclesfield Banks), adopt more or less the same positions in the name of Chinese history. Therefore, we may in a first approximation also discard Taiwan.

On the whole we shall restrict the game to three players: China (C), Vietnam (V), and the Philippines (P). Extending the three-player game to five or six players presents no theoretical difficulties, but requires more computation.

To build the game, we need to describe each player's strategic set, and the correspondences between the set of strategic triples and the set of outcome triples. This is in fact quite a difficult problem as a change in wording or interpretation of strategies can lead to very different conclusions, as we shall exemplify by considering successively two examples.

4.6 Example 1

Each party has the three following generic strategies available:

- Nibble an unoccupied island, a reef, or a rock here and there.
- Consolidate occupation of places already occupied.
- Concede unoccupied places nibbled by third parties.

In a first version of the game, we shall assume that:

- Nibbling by all parties, as it seems to legitimize a fait accompli, is unacceptable for all parties.
- Nibbling by one party is unacceptable by parties which do not do the same.
- Consolidation of occupation by a party is acceptable by the other parties.
- One party can consider conceding is acceptable as long as there is reciprocity.
- When two parties concede, they concede to each other, and the third party is then left in an unacceptable situation unless it nibbles.

On the whole, the three generic strategies here above form an escalation process. The above strategies will be denoted by the following indices:

- n for nibble.
- o for consolidate occupation.
- c for concede.

Table 10. The Spratly dispute version 1

(C_n,P_n,V_n) --- (0,0,0)	(C_n,P_c,V_o) --- (1,0,0)	(C_o,P_o,V_c) --- (1,1,0)	(C_c,P_o,V_n) --- (0,0,1)
(C_n,P_n,V_o) --- (1,1,0)	(C_n,P_c,V_c) --- (1,0,0)	(C_o,P_c,V_n) --- (0,0,1)	(C_c,P_o,V_o) --- (0,1,1)
(C_n,P_n,V_c) --- (1,1,0)	(C_o,P_n,V_n) --- (0,1,1)	(C_o,P_c,V_o) --- (1,0,1)	(C_c,P_o,V_c) --- (1,0,1)
(C_n,P_o,V_n) --- (1,0,1)	(C_o,P_n,V_o) --- (0,1,0)	(C_o,P_c,V_c) --- (0,1,1)	(C_c,P_c,V_n) --- (0,0,1)
(C_n,P_o,V_o) --- (1,0,0)	(C_o,P_n,V_c) --- (0,1,0)	(C_c,P_n,V_n) --- (0,1,1)	(C_c,P_c,V_o) --- (1,1,0)
(C_n,P_o,V_c) --- (1,0,0)	(C_o,P_o,V_n) --- (0,0,1)	(C_c,P_n,V_o) --- (0,1,0)	(C_c,P_c,V_c) --- (1,1,1)
(C_n,P_c,V_n) --- (1,0,1)	(C_o,P_o,V_o) --- (1,1,1)	(C_c,P_n,V_c) --- (0,1,0)	

It follows that if C, P, and V denote China, the Philippines and Vietnam, respectively, the Spratly dispute is represented by the correspondence list of *Table 10.*

Consider for instance Chinese strategies. Comparison of the three columns in the list shows that $J(C_n) \geq J(C_O) and J(C_n) \geq J(Cc)$.

The game being symmetric:
$J(P_n) \geq J(P_O)$; $J(P_n) \geq J(P_c)$ and $J(V_n) \geq J(V_O)$; $J(V_n) \geq J(V_c)$. Given these inequalities, one expects that each country will select the strategy with the highest positive playability, whence the equilibrium (C_n, P_n, V_n): China, the Philippines, and Vietnam will start a "nibbling" race, with unacceptable consequences for each one, which means that nibbling strategies are not positively playable. The above inequalities then imply the same for the strategies of occupation, consolidation, and concessions.

On the whole, all strategic triples are by default equilibria.[21]

An ordering of the set of equilibria can be achieved by extending the analysis to the case where playabilities are fuzzy. We then find a unique equilibrium (C_n, P_n, V_n).

This equilibrium exhibits properties that resemble those of the Nash equilibrium in the Prisoner's Dilemma. Indeed, this equilibrium is unacceptable for the three countries, whereas other strategic triples like (C_c, P_c, V_c) are acceptable for all parties.

Let us now consider that nibbling by a party is unacceptable to other parties, whatever their strategies. Conclusions remain unchanged. This robustness seems to proceed in part from the symmetry assumed between the players. Now, if we look at the three parties, we can question this symmetry. Before examining asymmetric games, one should nevertheless be aware that symmetry in the game does not stem from the symmetry between the players' capacities but is the result of their perceptions.

For instance, nibbling by China may be unacceptable to the Philippines because the islands that are susceptible to be seized by the Chinese are located near the Philippines coast and their seizure seems a direct threat to Philippines' security. On the other hand, nibbling by the Philippines may be unacceptable for China, not because of security but because it puts into question the Chinese claim of a total sovereignty over the South China Sea.

Let us then introduce an asymmetry between the parties by assuming that:

[21] It should be noted that the game symmetry implies that of the set of equilibria is relative to the three players, but not symmetry of the equilibria themselves.

- For China, nibbling by another party is unacceptable, whatever the Chinese strategy.
- For the Philippines and Vietnam, nibbling can be accepted on the basis of reciprocity.

It can be shown that inequalities between positive playability indices remain unchanged, the only difference being that nibbling strategies are now safe for the Philippines and Vietnam, while all Chinese strategies are playable by default. It follows that consolidating occupation and conceding are not playable strategies for either the Philippines or Vietnam. There are three equilibria in the nonfuzzy case and one in the fuzzy case (the same as before).

Finally, consider the case that differs from the previous one by the fact that general nibbling is unacceptable to all parties because of a risk of escalation. The conclusions are the same as before. On the whole, the model exhibits real robustness, and little hope for a peaceful settlement.

Thus, let us consider the case where Vietnam and the Philippines, while being rivals, are nevertheless members of a coalition (i.e., ASEAN). The game now becomes a matrix one (see *Table 11*).

Table 11. The Vietnam/Philippines coalition, version 1

	P_cV_c	P_cV_o	P_cV_n	P_oV_c	P_oV_o	P_oV_n	P_nV_c	P_nV_o	P_nV_n
C_n	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(0,0)	(0,0)	(0,0)
C_o	(0,1)	(1,0)	(0,0)	(1,0)	(1,1)	(0,0)	(0,0)	(0,0)	(0,1)
C_c	(1,1)	(1,0)	(0,0)	(1,0)	(0,1)	(0,0)	(0,0)	(0,0)	(0,1)

The game exhibits no safe strategy. Inequalities between positive playability indices for the Chinese strategies are the same as before. For the coalition these inequalities are:

$$J(P_nV_n) = J(P_oV_o) = J(P_cV_c) \geq J(P_cV_o) = J(P_cV_n) = J(P_oV_c) = J(P_oV_n) = J(P_nV_c) = J(P_nV_o)$$

Assume for instance that C_n is positively playable. It follows that all coalition strategies are playable by default, which contradicts the assumption. Consequently, all Chinese strategies are playable by default, which implies that all the coalition's strategies are as well.

In the fuzzy game, China will choose C_n because it has the highest positive playability index. All the coalition strategies are then equivalent.

In general, the noncooperative game induces the players to escalate. Putting Vietnam and the Philippines in a coalition does not change these conclusions. Moreover, it can be shown that changing assumptions does not change the conclusions. The results contradict the observations made on the evolution of the conflict until now.

If one cannot exclude these conclusions being verified in the future, their dramatic implications command that some second thought be given to the formulation of the strategic sets and the corresponding assumptions.

4.7 Example 2

Let us change the problem formulation and consider the following generic strategies:

- Offensive;
- Defensive;
- Passive.

We can reasonably assume that:

- An offensive strategy includes a defensive one in the sense that if party A develops a strategy of an offensive nature, then it will defend itself against another party's offensive.
- *Passive* means "does nothing," even in the face of another party's offensive.

These strategies may be implemented differently. We need, therefore, to give specific interpretations of what is called offensive and defensive. For party A, interpretations appear in *Table 12* (the other parties being B and C).

Table 12. Types of offensive and defensive

		Offensive		
		Against B and C	Against B	Against C
Defensive	Nonspecific			
	Specific			

By specific, we mean a defensive strategy implemented by a party which has been the direct subject of an offensive. In contrast, nonspecific means a defense against any offensive lead by one party against any one of the other two. We also need to make a distinction between the players. Two of them, China and Vietnam, claim sovereignty over the whole of the Spratly archipelago, while the third (i.e., the Philippines), claims sovereignty over only a part of it.

Therefore, any offensive in the archipelago is considered by China and Vietnam as an offensive against them, while some offensives are of no concern to the Philippines.

Of course, this is a simplification of what really happens, since all parts of the Spratly do not bear the same importance for China or for Vietnam.

Given the context, we can reasonably assume that no party would risk launching a global offensive because of either the risk of escalation, or, more simply, the inability to do so. Then:

- The Philippines has only one offensive strategy, resented as an aggression by both China and Vietnam, while the two other players have two (against the Philippines and the third party and against the third party only).
- China and Vietnam have only one defensive strategy (nonspecific), while the Philippines has two.

Adding the passive strategy means that each player has four strategies available.

The above strategies will be denoted by the following indices:

- O_x for offensive against x (x being China (c), Philippines (p), Vietnam (v), all (a)).
- D_s or D_{ns} for defensive specific or nonspecific.
- N for no reaction.

We shall here assume that:

1. Offensive by two parties or more is unacceptable for all parties because it carries the risk of an uncontrolled escalation.
2. One offensive to which the party concerned replies with a defensive strategy is acceptable both for the offender and the defender.
3. Absence of reaction to an offensive concerning a party is unacceptable to that party.
4. Defense against an offensive to which the Philippines are not subject is unacceptable to it, given the risks of escalation and the relative absence of interest to the country.
5. Defensive strategy by two parties as a response to a single offensive strategy of the third party is acceptable for the two defenders but unacceptable for the offender.
6. Except for the assumptions above, all three parties play a symmetric role in the game.

The Spratly dispute can then be represented by the correspondence list of *Table 13*, where the first term refers to China, the second to the Philippines, and the third to Vietnam.

Table 13. The Spratly dispute, version 2

(O_a,O_a,O_a) ---- (0,0,0)	(O_v,O_a,O_a) ---- (0,0,0)	(D_{ns},O_a,O_a) ---- (0,0,0)	(N,O_a,O_a) ---- (0,0,0)
(O_a,O_a,O_c) ---- (0,0,0)	(O_v,O_a,O_c) ---- (0,0,0)	(D_{ns},O_a,O_c) ---- (0,0,0)	(N,O_a,O_c) ---- (0,0,0)
(O_a,O_a,D_{ns}) --- (0,0,0)	(O_v,O_a,D_{ns}) --- (0,0,0)	(D_{ns},O_a,D_{ns}) --- (1,0,1)	(N,O_a,D_{ns}) --- (0,1,1)
(O_a,O_a,N) ----- (0,0,0)	(O_v,O_a,N) ----- (0,0,0)	(D_{ns},O_a,N) ----- (1,1,0)	(N,O_a,N) ----- (0,1,0)
(O_a,D_{ns},O_a) --- (0,0,0)	(O_v,D_{ns},O_a) --- (0,0,0)	(D_{ns},D_{ns},O_a) --- (1,1,0)	(N,D_{ns},O_a) --- (0,1,1)
(O_a,D_{ns},O_c) --- (0,0,0)	(O_v,D_{ns},O_c) --- (0,0,0)	(D_{ns},D_{ns},O_c) --- (1,0,0)	(N,D_{ns},O_c) --- (0,0,1)
(O_a,D_{ns},D_{ns}) -- (0,1,1)	(O_v,D_{ns},D_{ns}) -- (0,0,1)	(D_{ns},D_{ns},D_{ns}) -- (1,1,1)	(N,D_{ns},D_{ns}) -- (1,1,1)
(O_a,D_{ns},N) ---- (1,1,0)	(O_v,D_{ns},N) ---- (1,0,0)	(D_{ns},O_{ns},N) ---- (1,1,1)	(N,D_{ns},N) ---- (1,1,1)
(O_a,D_s,O_a) ---- (0,0,0)	(O_v,D_s,O_a) ---- (0,0,0)	(D_{ns},D_s,O_a) ---- (1,1,0)	(N,D_s,O_a) ---- (0,1,1)
(O_a,D_s,O_c) ---- (0,0,0)	(O_v,D_s,O_c) ---- (0,0,0)	(D_{ns},D_s,O_c) ---- (1,1,1)	(N,D_s,O_c) ---- (0,1,1)
(O_a,D_s,D_{ns}) --- (0,1,1)	(O_v,D_s,D_{ns}) --- (1,1,1)	(D_{ns},D_s,O_{ns}) --- (1,1,1)	(N,D_s,O_{ns}) --- (1,1,1)
(O_a,D_s,N) ----- (1,1,0)	(O_v,D_s,N) ----- (1,1,0)	(D_{ns},D_s,N) ----- (1,1,1)	(N,D_s,N) ----- (1,1,1)
(O_a,N,O_a) ----- (0,0,0)	(O_v,N,O_a) ----- (0,0,0)	(D_{ns},N,O_a) ----- (1,0,1)	(N,N,O_a) ----- (0,0,1)
(O_a,N,O_c) ----- (0,0,0)	(O_v,N,O_c) ----- (0,0,0)	(D_{ns},N,O_c) ----- (1,1,1)	(N,N,O_c) ----- (0,1,1)
(O_a,N,D_{ns}) ---- (1,0,1)	(O_v,N,D_{ns}) ---- (1,1,1)	(D_{ns},N,D_{ns}) ---- (1,1,1)	(N,N,D_{ns}) ---- (1,1,1)
(O_a,N,N) ------ (1,0,0)	(O_v,N,N) ------ (1,1,0)	(D_{ns},N,N) ------ (1,1,1)	(N,N,N) ------ (1,1,1)

Let us consider first the Chinese strategies: $J(D_{ns}) \geq J(N) \geq J(O_v) \geq J(O_a)$. As Vietnam and China play symmetric roles, for Vietnam: $J(D_{ns}) \geq J(N) \geq J(O_c) \geq J(O_a)$.

Lastly, for the Philippines:
$J(D_s) \geq J(D_{ns}) \geq J(O_a)$ and $J(D_s) \geq J(N) \geq J(O_a)$.

No player has a safe strategy. There are several solutions, depending on the dominant strategies' positive playability. But one can expect the parties to play their dominant strategies, in which case there is no escalation.

Consider now a second version of the game that differs from the first only in that hypothesis 1 is modified in that it assumes that two offensives led by other parties would be considered unacceptable by a player only if this player is directly concerned with these two offensives and that three offensives are unacceptable to all players.

Just as in the first version, offensive strategies are never positively playable, while passive strategies may be playable for some solutions, and defensive strategies are always playable. Therefore, rational players should choose defensive strategies.

Consider now a game derived from game 1 by changing hypothesis 3, and assuming instead that given the balance of power between the Philippines or Vietnam on the one hand and China on the other, the former may simply have no other possibility than to consider acceptable not to react *alone* to a Chinese offensive. Thus, hypothesis 3 now reads: "Absence of reaction to the Philippines or a Vietnamese offensive that directly concerns a party is unacceptable to that party. Absence of reaction to a Chinese offensive that directly concerns a party is unacceptable to that party, only if the party concerned by the offensive decides to select a nonspecific defensive strategy." Once again the conclusions are the same.

If we now consider a coalition between the Philippines and Vietnam, the game is as described by the matrix of *Table 14*.

Table 14. The Vietnam/Philippines coalition, version 2

	O_A	O_V	D_{ns}	N
(O_A,O_A)	(0,0)	(0,0)	(0,0)	(0,0)
(O_A,O_c)	(0,0)	(0,0)	(0,0)	(0,0)
(O_A,D_{ns})	(0,0)	(0,0)	(0,1)	(1,0)
(O_A,N)	(0,0)	(0,0)	(0,1)	(0,0)
(D_{ns},O_A)	(0,0)	(0,0)	(0,1)	(1,0)
(D_{ns},O_c)	(0,0)	(0,0)	(0,1)	(0,0)
(D_{ns},D_{ns})	(1,0)	(0,0)	(1,1)	(1,1)
(D_{ns},N)	(0,1)	(0,1)	(1,1)	(1,1)
(D_s,O_A)	(0,0)	(0,0)	(0,1)	(1,0)
(D_s,O_c)	(0,0)	(0,0)	(1,1)	(1,0)
(D_s,D_{ns})	(1,0)	(1,1)	(1,1)	(1,1)
(D_s,N)	(0,1)	(0,1)	(1,1)	(1,1)
(N,O_A)	(0,0)	(0,0)	(0,1)	(0,0)
(N,O_c)	(0,0)	(0,0)	(1,1)	(1,0)
(N,D_{ns})	(0,1)	(1,1)	(1,1)	(1,1)
(N,N)	(0,1)	(0,1)	(1,1)	(1,1)

The coalition has one safe strategy (D_s, D_{NS}), while China has none. The existence of these strategies implies that Chinese strategy O_A is not positively playable. Moreover, for China: $J(D_{NS}) \geq J(N) \geq J(O_V) \geq J(O_A)$.

D_{NS} being dominant for China is playable. Then coalition strategies (O_A, O_A) and (O_A, O_C) are not playable, which implies that Chinese strategy D_{NS} is positively playable. Therefore, one expects that coalition members will adopt a (safe) defensive attitude. Similarly, since China has a dominant positively playable strategy, it has no reason to choose an alternative. Varying the assumptions as in example 1 does not change the conclusions.

We can see that examples 1 and 2 lead to very different conclusions. First, while in example 1, offensive strategies are dominant and never positively playable, this is not the case in example 2, where dominant strategies are always defensive and may even be positively playable.

Second, while in example 1 building a coalition does not introduce safe or even positively playable strategies, this is not the case in example 2, where the coalition has a safe strategy and China has a positively playable strategy. If building a coalition is useless in example 1, it helps bringing stability in example 2.

4.8 Toward a semantic theory of the Spratly issue

The above discussions show how interaction and, in particular, negotiation processes may result in entirely different scenarios, simply because of a slight change in the wording and the meaning of the strategies: in the first case, no matter what the assumptions about acceptability, interaction inevitably leads to escalation; in the second case, players would always refrain from offensive actions, and coalition would only reinforce that conclusion.

This takes us back to the problem of managing ambiguity, which is at the core of negotiation processes. Ambiguity may prove fruitful in some cases when it is a necessary condition for an agreement to be reached. In other cases, ambiguity might lead to disaster for some if not all the parties. In any case, exploration of the various possible meanings of propositions or strategies must be done in order for the negotiators to be sure that the words they use have the anticipated effect on the negotiation process.[22]

Developing a full theory about ambiguities is outside their scope. Nevertheless, we can pave the way for such a theory by proposing to translate assumptions into a set of logical formulas, with ambiguities leading naturally to different sets of formulas. The extensive research done in the field of logic and natural language can then be fully capitalized in order to use all the formal tools already developed, namely knowledge-deduction systems.

For instance, the six assumptions taken in the first version of example 2 can be translated using the following categories:

[22] In a way that issue recalls the Adjusted Winner, where one party can make deceptive assessments in order to maximize its outcome. Albeit different in nature, wording may also be sometimes considered as a decision problem including a deception dimension.

- Constants or domain of interpretation: China, Philippines, Vietnam;
- Variables: x, y, z, which domain is defined above;
- Defined predicates: see below;
- Logical operators: *not, and, or, imply* (denoted by $\Rightarrow$);
- Logical quantifiers: $\forall$.

More precisely, the six assumptions can be described using the following three predicates:

- Acc (X): state of the world acceptable for X;
- Off (X,Y): X develops an offensive strategy against Y;
- Def (X,Y,Z): X develops a defensive strategy to respond to an offensive led by Y against Z.

A specific defensive strategy led by X considers only the situation where Z = X. It stems from the preceding that the six assumptions above can be restated as follows:

1. $\forall$ X, $\forall$ Y, $\forall$ Z (Off(X,Y) and Off(Y,Z)) or Off(X,Y) and Off(Z,X)) or Off(X,Y) and Off(Z,Y)) or Off(X,Y) and Off(Y,X)) and Y $\neg$ X and Z $\neq$ Y and Z $\neq$ X $\Rightarrow$ not ACC(X) and (not ACC(Y)) and (not ACC(Z))
2. $\forall$ X, $\forall$ Y, $\forall$ Z Off(Y,X) and (not Off(Z,X)) and Def(X,Y,X) and Y $\neq$ X and Z $\neq$ Y and Z $\neq$ X $\Rightarrow$ ACC(X) and ACC(Y)
3. $\forall$ X, $\forall$ Y Off(Y,X) and (not Def(X,Y,X)) and Y $\neq$ X $\Rightarrow$ not ACC(X)
4. $\forall$ X, $\forall$ Y $\forall$ Z Off(Y,Z) and Z $\neq$ Phil and Y $\neq$ Phil and Z $\neq$ Y and Def(Phil,Y,Z) $\Rightarrow$ not ACC(Phil)
5. $\forall$ X, $\forall$ Y, $\forall$ Z Y $\neq$ X and Z $\neq$ Y and Z $\neq$ X, Off (X,Y) and (not [Off(Y,Z) or Off(Z,X) or Off(Z,Y) or Off(Y,X)] and Def (Y,X,Y) and Def (Z,X,Y)] $\Rightarrow$ ACC(Y) and ACC(Z) and (not ACC (X))
6. The symmetry assumption is already taken into account by the previous ones.

Looking at the set of possible predicate combinations, and associating each combination with a truth value, paves the way for automatic development of specific games based on these predicates.

4.9 From multistage games to diplomacy

In real life, negotiations are processes, the dynamics of which can be represented by a sequence of games [16]. The game at a given stage depends on the outcomes of the game at the previous stage(s). For instance, the Philippines and Vietnam may decide to form a coalition if the previous game was characterized by the assumptions of the first version of example 2. Assume that conditions change again, such that the Philippines may decide to leave the coalition and try to negotiate on a bilateral basis, etc. Such decisions, taken by each party at each stage, define meta-games. Of course, a party cannot define the game that will be played at the next stage, only through its own meta-decision. The next game is defined by the set of meta-decisions taken by each party at the present stage.

Moreover, with each meta-decisions tripling, we can associate not only a specific game but the solution set corresponding to this game, and accordingly the game outcomes. This means that at each stage, and for each set of assumptions, a meta-game of deterrence can be built up. The sequence of such meta-games defines a multistage meta-game, in which a player (meta) strategy is a sequence of meta-decisions. If these meta-decisions concern only the buildup of alliances and coalitions, we speak of the corresponding strategies as *diplomacies*.

5 Game Theorists and Negotiators: A Path toward Cooperation

The relationship between quantitative methods in social sciences and "real life" has been a subject of debate for decades. Its source lies in problems found on both sides. For a quantitative problem to be correctly stated, a set of assumptions is required, as well as a set of rules to show how these assumptions should be combined. Therefore, one can always pretend that the results are as good (reliable) as the worst (quality) of the set of assumptions and rules.

If one wants to keep the problem tractable, technical intricacies quickly limit the volume of assumptions and the complexity of the set of rules. It can then be argued that quantitative methods are by nature incomplete tools for representing real-life issues: they do not embrace an entire reality because either they do not encompass the nuances involved or they fail to fully take those elements into account when developing technical processes for reaching conclusions.

Critics point out that some assumptions, however seductive they appear, remain purely theoretical. The debate on rationality epitomizes this. Let us just recall a sample of pending issues. The radicals question a tool pretending to model social interaction, while it is centrally based on rationality. Are not "real" players most often irrational? Thus, what is the point of developing a battery of appropriate tools if players are not always rational? Would it not be more interesting to focus on decision traps and biases, as, for instance, the American School of Experimental Psychology [19] does?[23]

Some do not question the existence of rational players, but the type of rationality on which they consider game theory to be based (i.e., instrumental rationality), or they ask, "Does game theory simply repeat what everyone already knows in a language that no-one understands?" [8]. Others consider that "The trouble with game theory is that it can explain anything. If a bank president was standing in the street and setting his pants on fire, some game theorist would explain it as rational" [11].

"Real-life" decision makers often either do not understand the tools developed by game theory or do not want these tools to be implemented because it would put their status in question. This is particularly true when quantitative methods deal with high-level decision making. In that case the social status of the decision maker lies in his capacity to let his environment believe he will take the appropriate decision.

[23] Even if game theory takes into account excursions from rationality, through for instance a Trembling Hand-like approach.

In other words, decision *is* power: on the one hand, decision impacts environment; on the other, letting others know that he is the one who decides, gives the decision maker leverage and hence determines his social status [15].

Therefore, the decision maker may feel he is losing power if a decision is taken by some mechanical device entirely developed through quantitative methods, and even worse, if these methods appear to the decision maker as some kind of black box that he cannot control.

With this flow of mutual critics, is it reasonable to envisage some kind of reconciliation between game theory and the "real world"? Perhaps, through excessive optimism, we think that yes, a bridge can be built between the two sides, along the following lines:

- Some sociological analysis of critics and criticism should be undertaken ("criticism of criticism" [15] beyond scientific analysis of critical content.
- The dispute over rationality should be stopped by combining a memorandum of understanding, recognizing that many issues will still be debatable for quite some time, with an agreement over the use of simple tools with limited scope.
- A constant dialog should be organized between the developers and potential users of decision tools, so that the latter can participate in the development (at least in the nontechnical stages, which are often of utmost importance in the definition, significance, and orientations of the model), while the former do not try to impose ready-made solutions, which might not fit the actual problem under consideration.

An appropriate catalyst for this dialog may be the prior development of a lexicon, aiming at an agreement by both sides on the meaning of each term. Such a development carries a good chance that not only misperceptions but also mistrust can be avoided. In particular, it can help prevent the decision maker from fearing exclusion from the decision process, and to commit himself to appling a solution coming from a black box.

A good start toward reconciling game theory and practical negotiations would be to:

1. Resort to dynamic games in order to embrace the preoccupation of practitioners eager to build an interaction process that will enable them to refine their information about the other side.
2. Consider games in which each player has a limited rationality that requires only rough information to be determined.

6 Conclusions

The model presented here has been validated at each of the two levels of interactions by two "real-life" international negotiations: the Icelandic fisheries and the Spratly disputes.

In the first case the negotiation outcome predicted by the model is exactly the one that happened at the end of the real negotiating process.

In the second case, negotiations are still going on; hence their outcome is not yet known, but the model highlights the fact that small semantic variations can generate big divergences in the process, something already known by real-life negotiators for whom the way a sentence is phrased or a simple word is used may be of the utmost importance. In the example considered, the results of the model show potential and dangerous oscillations between war and peace and confirm the nonquantitative studies undertaken by specialists of the region, according to which the Spratly dispute bears a high level of uncertainty.

More generally, games of deterrence are not only fitted for describing situations for which only qualitative information is available, but their binary features also make them directly usable by negotiators unfamiliar with the complex tools of standard game theory: anybody can fill a binary matrix, the entries of which only indicate whether a given state of the world is acceptable or not to a given party. Moreover, games describing elementary go–no-go situations can be solved without resorting to sophisticated mathematical tools. In the case of more complex interactions—that is, when the number of actions that can be potentially taken by the parties is high—systems of automatic resolution can be made available, such that negotiators willing to explore the consequences of a given set of interactions have only to answer yes or no to the question of acceptability of a given situation for a given party.

Conversely, the graphs associated with games of deterrence may be used to explore the change needed in the conditions of negotiation to make them evolve in a direction more favorable to a party.

Last, but not least, the model defines a framework for dealing with semantic issues by analyzing all possible meanings of a given sentence and the consequences in terms of negotiation outcomes of selecting one particular meaning among the many that are possible.

References

1. Bell DE, Raiffa H, Tversky A (1988) Decision making. Cambridge University Press, Cambridge, UK
2. Bestougeff H, Rudnianski MM (1998) Games of deterrence and satisficing models applied to business process modeling. Proceedings AAAI Spring Symposium, Stanford, CA, USA
3. Brams SJ (1990) Negotiation games. Routledge, London, UK
4. Brams SJ, Taylor AD (2000) The win-win solution: guaranteeing fair share to everybody. WW Norton and Company, New York, USA
5. Fitting M (1983) Proofs methods or models on intuitionistic topics. Synthese Library, Vol 169, Reidel Publishing, Boston, MA, USA
6. Fraser JT (1987) Time: the familiar stranger. Tempus Press, Hertfordshire, UK
7. Garey MR, Johnson DS (1979) Computers and intractability: a guide of theory of NP completeness. Freeman, San Francisco, USA
8. Heargrave Heaps SP, Varoufakis Y (1995) Game theory, a critical introduction. Routledge, London, UK

9. Kahneman D, Slovic P, Tversky A (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge, UK
10. Lyons J, (1977) Semantics. Cambridge University Press, Cambridge, MA, USA
11. Postrel S (1995) Burning your britches behind you: can policy scholars bank on game theory. In: Rumelt RP, Schendel DE, Teece DJ (eds) Fundamental issues in strategy, a research agenda. Harvard Business School Press, Boston, MA, USA, 221–228
12. Pearl J (1988) Probabilistic reasoning in intelligent systems: network of plausible inferences. Morgan Kaufmann Publishers, San Mateo, CA, USA
13. Rescher N, Brandom R (1980) The logic of uncertainty. Basil Blackwell, London, UK
14. Rudnianski M (1992) Deterrence typology and nuclear stability: a game theoretical approach. In: Avenhaus R, Karkar H, Rudnianski M (eds) Defense decision making: analytical support and crisis management. Springer Verlag, Berlin, Germany, 137–168
15. Rudnianski M (1993) La théorie des jeux comme outil de la décision internationale. In: Le Trimestre du Monde, 4éme Trimestre, Paris, France, 59–77
16. Rudnianski M (1997) Deterrence, fuzzyness and causality. Proceedings ISAS 97, World Multiconference on Systemics, Cybernetics and Informatics, Caracas, Venezuela
17. Rudnianski M, Bestougeff H (2001) Modeling tasks and teams through game theoretical agents, formal approaches to agent based systems 2000. Lecture Notes in Artificial Intelligence, Springer, Berlin, Germany
18. Rudnianski M, d'Assignies A (1995) From MAD to MAD. In: Best ML, Hughes-Wilson J, Piontkowsky A (eds) Strategic stability in the post-cold war world and the future of nuclear disarmament. NATO ASI series, Disarmament Technologies, Kluwer Academic Publishers, Vol 3, 229–259
19. Russo JE, Schoemaker P (1990) Decision traps: ten barriers to brilliant decision-making and how to overcome them. Fireside, New York, USA
20. Simon H (1983) Models of bounded rationality. Vol 1 & 2, MIT Press, Cambridge, MA, USA

Greek–Turkish Territorial Waters Game*

Serdar Ş. Güner

Bilkent University, Department of International Relations, Ankara, Turkey

1 Introduction

Unlike the previous contributions to this section on Models of International Negotiations this chapter deals with a concrete and as yet unsolved international conflict, namely, that between Greece and Turkey over the breadth of territorial waters in the Aegean Sea. Greece claims that it has the freedom to extend its territorial waters to 12 miles, whereas Turkey has indicated that a Greek extension of territorial waters constitutes a casus belli. Currently, both countries apply the six-mile limit; several crises have occurred over this issue in the past, but the status quo has remained unchanged.

The conflict is modeled as a noncooperative game of asymmetric incomplete information in extensive form, where Greece may misperceive Turkish war costs. Therefore, following Harsanyi [6] we assume there are three rational players, namely Greece, Hard Turkey, and Soft Turkey, the latter two being characterized by Greek perceptions of Turkish war costs. Depending on the game parameters, computed Nash equilibria imply alternative explanations of the status quo, war, and the revised status quo in the Aegean Sea. While the status quo can depend upon Greek misperceptions, war and the revised status quo are found to be independent of them.

Practitioners or diplomats might be interested in explaining the continued observation of the six-mile limit, the periodic crises, and the likelihood of a war between Greece and Turkey. As there are no open negotiations between the littorals (Greece refuses to bring the issue to the negotiation table), intermittent declarations of positions constitute a tacit bargaining process. We can explain war and the endurance or the revision of the Aegean status quo under misperceptions using a simple model.

It is also possible to analyze the issue without requiring rationality assumptions. For example, Güner [4] proposes an evolutionary game by considering foreign policies as templates that reflect standard operating procedures and fixed strategic

* I am grateful to the International Institute for Applied Systems Analysis for its invitation to join the research project on Formal Models and International Negotiations and for financial assistance. I particularly thank Dan Druckman, Rudolf Avenhaus, and I. William Zartman for their generous help, guidance, and suggestions.

prescriptions. The game indicates a periodicity of hawkish and dovish foreign policies that generate recurrent crises over the Aegean Sea. When Greek hawks face a lesser number of Turkish hawks they fare better than Greek doves. Accordingly, Greek hawks are imitated, and the number of Greek hawks expands. Greek foreign policies mainly challenge the status quo. Matched with a greater number of Greek hawks, however, Turkish no-reaction policies become less successful, unlike those that demand a forceful response. Successful Turkish hawks are then imitated and Turkish foreign policy reacts to the challenges. Greek doves then again become more successful and they in turn multiply, as Greek hawks begin to encounter a greater number of Turkish hawks. Back to square one: the whole evolutionary process restarts. It is impossible, however, to investigate conscious strategic signals that players can give in an evolutionary game. As a result, the model I propose relies on the sophisticated rationality of unitary players.

In the sense of the definition of models *of* international negotiations, which is used throughout this publication, negotiations are not explicitly modeled here. In addition, the concrete model developed does not reflect an impartial view of the conflict, but rather that of a Turkish analyst. However, as outlined in this volume's general introduction it is demonstrated that the model could serve at least two policy purposes, which could become important for ongoing and forthcoming negotiations. First, policymakers could be encouraged to modify the present model according to their ideas of perceptions, moves, costs, etc., and second, models of the kind presented here could guide official policies in appropriate directions, thus contributing to the planning of promising negotiation strategies.

2 The Conflict

A major issue confronting Greece and Turkey is the breadth of territorial waters in the Aegean Sea. Greece claims that it has the freedom to extend its territorial waters to 12 miles, while Turkey has indicated that a Greek move to extend territorial waters constitutes a casus belli. Currently, both countries apply six-mile limits. Several crises have already occurred over the issue, yet the status quo has not been disturbed.

Greece and Turkey are the only littoral states in the Aegean. More than 3,000 islands, islets, and rocks cover the sea. All, apart from İmroz (Imyros), Bozcaada (Tenedos), and Tavşan (Rabbit) islands, belong to Greece with some rocks and islets forming contested sovereignty zones.[2] The 1923 Treaty of Lausanne fixed the extension of the littoral states' territorial waters at three miles. Greece unilaterally declared territorial waters of six miles in 1936 during a détente period between the two countries. Turkey responded in 1964 with a similar move, and the current status quo formed. Greece and Turkey each maintain six miles of territorial sea. In the Aegean, this delimitation produces Greek territorial waters of approximately 44 percent, Turkish territorial waters of 7.5 percent, and high seas of about 49 percent.

[2] Turkey and Greece reached the brink of war in 1996 over the status of Imia/Kardak islets situated 5.5 miles from the nearest Greek island Kalymnos, 2.2 miles from the Turkish Çavuş (Kato) island, and 3.8 miles from the Turkish mainland.

In accordance with the United Nations Convention on the Law of the Sea (UNCLOS), signed in 1982, which entered into force in 1994, signatory states have the right to establish territorial waters of up to 12 miles. The percentages of Greek, Turkish, and international waters would then become 71.5 percent, 8.8 percent, and 19.7 percent ([1], [11]). Greece, as a signatory state, considers the determination of the breadth of its territorial waters to be a sovereign right. It claims it will extend its territorial waters to 12 miles in the future. A revised status quo, if both littoral states extended their territorial waters to 12 miles, implies the resolution of the continental shelf issue in favor of Greece and the undersea connection of the Greek mainland with thousands of islands scattered around the Aegean. This constitutes a considerable continental gain of shelf [1]. While Greece defends the rule of territorial integrity, that the islands and the mainland form an unbreakable whole and cannot be separated from the mainland, Turkey insists that the continental shelf delimitation should be established by drawing an equidistant line between the Greek and Turkish continental land masses and that the Greek islands clustering along the Turkish coast cannot have their own continental shelves.

Following the ratification of UNCLOS by the Greek parliament in June 1994, the Turkish parliament approved a resolution authorizing the Government to use all necessary measures to protect the rights of Turkey should the need arise. The Turkish position stems from Article 300 of UNCLOS, according to which: "Parties shall fulfill in good faith the obligations assumed under this Convention and shall exercise the rights, jurisdiction and freedoms recognized in this Convention in a manner which would not constitute an *abuse of right*."[3] The Aegean, according to Turkey, is a semi-enclosed sea and therefore requires the application of particular rules. Turkey insists that a Greek extension of its territorial waters to 12 miles will imply that even maritime transport between Turkish ports would require Greek permission. Turkey considers this to be an abuse of right. Greece argues exactly the opposite (i.e., the Aegean is not a semi-enclosed sea, and that the Turkish declaration of casus belli is against international norms). Greece believes that according to the UN Charter, Article 2, Paragraph 4, its territorial integrity is under threat.

Overall, the issue of the continental shelf in the Aegean is the only matter that Greece considers as controversial and negotiable. Greece insists on the one hand that it can extend its territorial seas to 12 miles at any time and sticks to the territorial integrity thesis, claiming that the Aegean is not a semi-enclosed sea and that the Turkish position of casus belli is contrary to the UN Charter. On the other hand, Turkey maintains that there is no rule applicable to all seas; specific rules should be applied like "the right of geographically disadvantaged states" in accordance with Article 70 of UNCLOS, and that no unilateral actions but coordination of actions should be permitted in semi-enclosed seas as recommended in Article 123 of UNCLOS.

Both countries seriously considered the option of war in 1976, 1987, and 1996. A crisis in 1976 saw both countries claiming rights to drill for oil in international waters. Greece and Turkey claimed that these areas belonged to their continental shelves. The genesis of this crisis was in 1972 when Greece granted oil exploration

[3] Emphasis added.

licenses on the high seas. Greece claimed it was granting these licenses for oil exploration on its own continental shelf. Turkey reacted in 1973 by acting similarly and permitting its national oil company to drill on the high seas. It argued that these areas are on the extension of the Anatolian Peninsula and form the Turkish continental shelf. A crisis erupted in 1976 after Turkey sent an exploration ship to search for oil in the Aegean, and the UN Security Council issued Resolution 395 in August of the same year calling upon Greece and Turkey to do everything in their power to reduce the tension in the Aegean and asking them to resume direct negotiations. The Greek reaction and claims compelled Turkey to declare any move by Greece to extend its territorial waters as a casus belli. The tension subsequently eased, and both sides later agreed in Bern to inform each other of their exploration activities and to respect the status quo. A crisis in 1987 occurred almost in the same fashion, with Greece starting to drill for oil and Turkey reacting by sending a ship into the contested waters. The tension eased within days. The prime ministers of both countries met in Davos, Switzerland, and reached an understanding. Neither the Davos process nor the Bern Agreement settled the issue of the continental shelf. Both crises indicated a serious likelihood of war but ended in agreements (explicit or tacit) that did not last long. There was a further crisis in 1996 over Imia/Kardak Rocks with both countries claiming sovereignty. The crisis de-escalated with active United States mediation and the status quo ante was finally restored. We agree with Bahcheli [3]:

> While most Aegean quarreling has centered on the continental shelf, the territorial sea issue is the one that is most vital for Turkey. The two issues are not unrelated, since all of the shelf claimed by Greece would accrue to it automatically, were it able to implement a 12-mile territorial claim.

Either claiming 12-mile territorial waters or the continental shelf finally amounts to the same sovereignty demand.

3 The Model

We begin by assuming that Greece is unsure about the costs of war for Turkey. Turkey is assumed to have two distinct types of war costs: war is costly for one type of Turkey but not that costly for the other type. Furthermore, it is assumed that Turkey knows Greek uncertainty, Greece knows that Turkey knows its uncertainty, Greece knows that Turkey knows that Greece knows its uncertainty and so on, ad infinitum. We further assume that war is costly for both sides, but Greek gain is almost doubled in a new status quo.

Greece can decide between six and 12 miles of territorial waters without exactly knowing what type of Turkey it faces. Once Greece makes its choice of 12 miles, Turkey could respond by either declaring war or accepting a new status quo in the Aegean Sea. The game ends if Greece chooses to remain at six miles of territorial waters or Turkey responds either by fighting back to Greek extension of its territorial waters or not fighting back. These assumptions imply the model represented graphically in *Figure 1*.

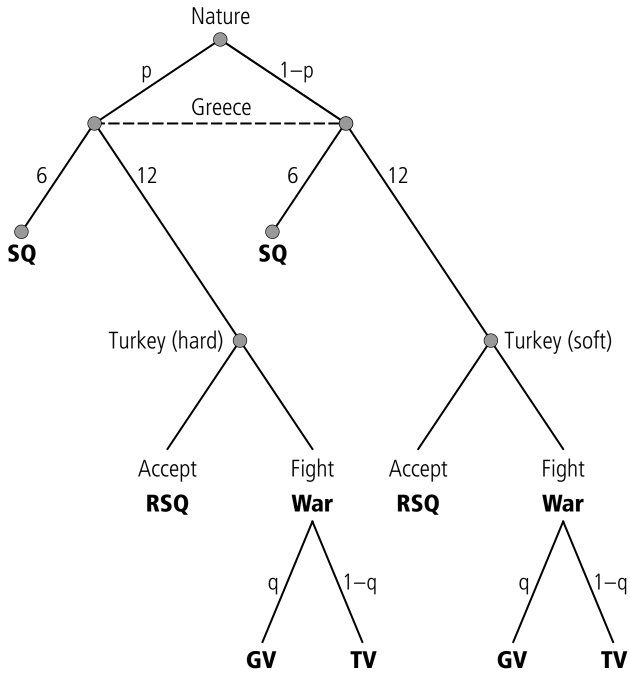

Fig. 1. Greek–Turkish territorial-waters conflict game (general)

In *Figure 1* the status quo and the revised status quo are abbreviated as SQ and RSQ, Greek victory and Turkish victory as GV and TV respectively. The tree greatly simplifies the dispute: there are only two players, and each player has two actions. Nature is the chance move that determines the type of Turkey. Turkey can either be hard or soft type with p and $1 - p$ probabilities, respectively. The Soft is presumed to suffer more than the Hard at the outbreak of war. Nature's move determines Greek misperception. Greece does not know which type is going to react when it decides to extend its territorial waters to 12 miles. The broken line tying nodes where Greece takes a decision indicates Greek suspicion.

After Nature determines its misperception, Greece moves by choosing either six or 12 miles. If it chooses to remain at a six-mile limit, the current status quo is not disturbed. The outcome becomes the status quo (indicated as SQ in the tree). If it chooses 12 miles, then Turkey reacts by either accepting 12 miles or fighting. If Turkey accepts the 12-mile delimitation, then the outcome is the revised status quo (RSQ). If Turkey fights after a Greek extension, the outcome is war. Either Greece (GV) or Turkey prevails (TV) respectively with probabilities q and $1 - q$ in the case of war.

Players obtain no territorial gain and suffer no war costs in the current status quo with respect to other outcomes in the game. Thus, for convenience, status quo payoffs for Greece and Turkey may be normalized to (0, 0) where the first component on the left is the Greek and that on the right is Turkish. If Greece extends and Turkey accepts, the status quo is revised. Greece almost doubles its Aegean territory, which is territory lost to Turkey. It can thus be assumed that Greece obtains 2 and Turkey -2 in the revised status quo. If Greece chooses to extend its territorial waters to 12 miles and Turkey responds by fighting, war occurs. If Greece wins the war, a new status quo occurs. Greece obtains its utility of the new status quo but suffers war costs that are denoted by the parameter g. Greece receives $2-g$ as a result of victory. If Greece loses, the old status quo continues under war costs. As Greece obtains 0 in the status quo but only suffers costs of war, a Greek defeat payoff is $-g$.

Both types of Turkey obtain 0 in the current status quo. If they accept a 12-mile norm in the Aegean Sea after Greece extends its territorial waters, both types of Turkey lose what Greece gains. Thus, the Hard and the Soft both obtain -2 in the revised status quo without war. If they respond to Greek extension by a military reaction, the status quo is revised through war, given that Greece becomes victorious. In this case, the Hard's payoff is assumed to be $-2 - t/2$ and the Soft's payoff is $-2-t$, where the parameter t denotes Turkish war costs. The distinction between the Hard and the Soft is that the Hard suffers half of what the Soft does in case of war. If they respond by attack and they become victorious, the status quo is not revised. The Hard's payoff is measured by its costs of war that is $-t/2$, and the Soft's payoff is measured by $-t$ in this instance.

In essence, two principal variables define Greek and Turkish preferences over the possible outcomes of the dispute: Greek and Turkish values in the revised status quo and both countries' war costs. These values are defined relative to the value of the status quo normalized to zero for both Greece and Turkey, see *Figure 2*.

Wars are prospective events that include either a Greek or a Turkish victory with specified probabilities. Given that Greece prevails with q and Turkey prevails with $1-q$ probabilities, the Greek expected war payoff becomes: $q(2-g)+(1-q)(-g) = 2q-g$. For Turkey, the Hard type similarly obtains: $q(-2-t/2)+(1-q)(-t/2) = -2q-t/2$. The Soft gets: $q(-2-t)+(1-q)(-t) = -2q-t$. Now our game tree is complete (see *Figure 3*).

4 Equilibrium Analysis

The Hard fights rather than accept 12 miles if its expected war payoff strictly exceeds its value in the revised status quo. This implies: $-2 < -2q-t/2$; therefore, $1-t/4 > q$. Hence, if the likelihood of Greek victory in the case of war satisfies this condition, the Hard fights Greek extension of territorial waters. If it is greater than $1-t/4$, the Hard accepts. If it is equal to $1-t/4$, the Hard is indifferent between responding by attack and accepts 12 miles as the new norm. Similarly, the Soft fights, given that Greece extends its territorial waters if $-2 < -2q-t$, that is, if $1-t/2 > q$. Otherwise, if $1-t/2 < q$, the Soft accepts. If $q = 1-t/2$, then the Soft is indifferent to fighting and accepts 12 miles as a new norm of territorial sea delimitation.

There are five different equilibria of the game. The equilibrium strategies are characterized by a triplet, where the first component denotes hard Turkey's decision,

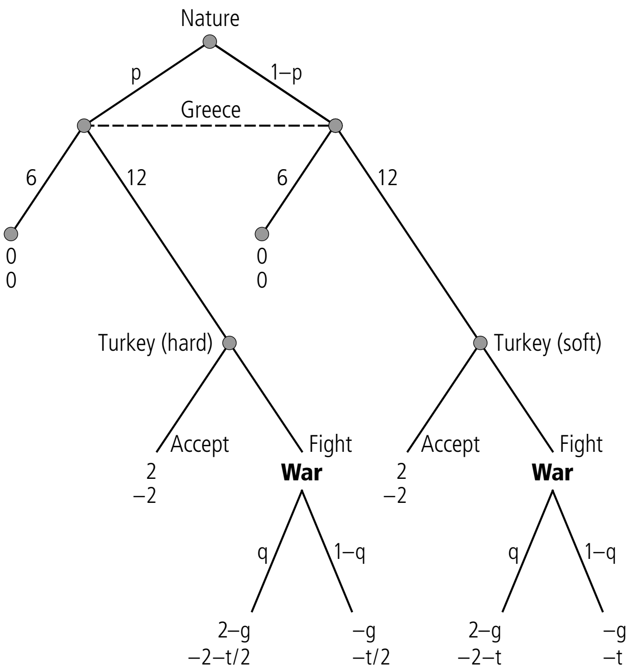

Fig. 2. Greek–Turkish territorial-waters conflict game (detailed)

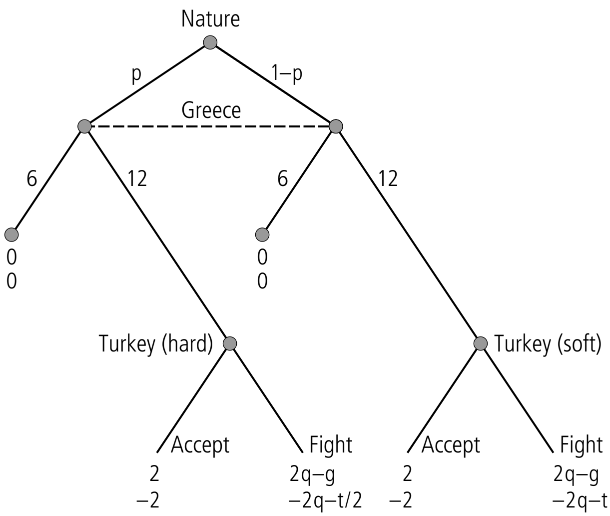

Fig. 3. Greek–Turkish territorial-waters conflict game (reduced)

the second soft Turkey's decision, and the third Greece's decision. The equilibria are classified into three cases according to the likelihood of Greek victory as compared to Turkish losses in case of war.

Case 1: $1 - \frac{t}{2} < q < 1 - \frac{t}{4}$
Case 1.1: $q < \frac{g}{2}$
Case 1.1.1: $p > \frac{2}{2(1-q)+g}$
Equilibrium 1: $\{$*Fight, Accept, Remain at six miles*$\}$
Case 1.1.2: $p < \frac{2}{2(1-q)+g}$
Equilibrium 2: $\{$*Fight, Accept, Extend to 12 miles*$\}$
Case 1.2: $q > \frac{g}{2}$
Equilibrium 2: $\{$*Fight, Accept, Extend to 12 miles*$\}$
Case 2: $q < 1 - \frac{t}{2}$
Case 2.1: $q < \frac{g}{2}$
Equilibrium 3: $\{$*Fight, Fight, Remain at six miles*$\}$
Case 2.2: $q > \frac{g}{2}$
Equilibrium 4: $\{$*Fight, Fight, Extend to 12 miles*$\}$
Case 3: $q > 1 - \frac{t}{4}$
Equilibrium 5: $\{$*Accept, Accept, Extend to 12 miles*$\}$

Figure 4 represents equilibrium regions under different parametric conditions.

5 Interpretation of Results

To assess how the model improves and disciplines intuitions or established viewpoints on the Aegean issue, we need to interpret the equilibria. The first two equilibria involve distinct actions for the two types of Turkey: if Greece extends its territorial waters to 12 miles, Hard Turkey fights and Soft Turkey accepts the new status quo. Greece, on the other hand, remains at six miles or extends its territorial-waters limit if its belief of facing Hard Turkey is above or below a threshold, respectively. Greek beliefs about Turkey's type matter only in these equilibria: in the remainder, both types of Turkey take the same action, and Greek misperceptions do not matter. We find that it is not necessary to think about Greek misperceptions to explain the current status quo, a future war, or a revised status quo. A practitioner could produce the same arguments without any difficulty if he or she is interested in such modeling efforts.

There are two ways of searching for implications of the first two equilibria.[4] One is to assume that the Greek probability of prevailing in war q is fixed and g varies. The second is to assume that g is fixed and q varies. The threshold gets progressively smaller (bigger) when Greek war costs increase (decrease), and it gets larger (smaller) when the Greek probability of prevailing increases (decreases). Suppose

[4] We can also interpret the first two equilibria by fixing the threshold and varying the Greek belief of facing the Hard. However, these interpretations require extra-game concerns to justify variations in that belief.

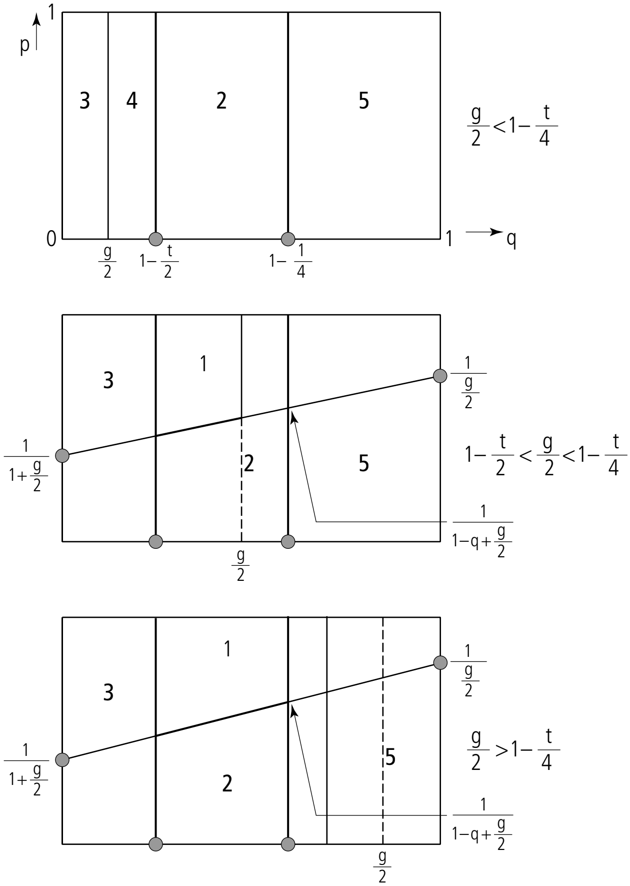

Fig. 4. Graphical representation of the parameter regions determining the five equilibria for the Greek–Turkish Territorial-Waters Conflict game

that the Greek belief of facing the Hard is fixed as relatively small or big. We obtain the following implications:

1. If Greek war costs increase, it is likely that a small Greek belief of facing the Hard will exceed the threshold that gets smaller. Greece will then decide to remain at six miles.
2. If Greek probability of victory decreases, it is again likely that a small Greek belief of facing the Hard will exceed a threshold that gets smaller. Greece will again remain at a six-mile limit.
3. If Greek war costs decrease, it is likely that a small Greek belief of facing the Hard will not exceed the threshold that gets bigger. Greece will then extend its territorial waters to 12 miles.

4. If Greek probability of victory increases, it is likely that a small Greek belief of facing the Hard will not exceed the threshold that gets bigger. Greece will then extend its territorial waters to 12 miles.
5. If Greek war costs increase, it is likely that a high Greek belief of facing the Hard will exceed the threshold that gets smaller. Greece will then remain at a six-mile limit.
6. If Greek probability of victory decreases, it is likely that a high Greek belief of facing the Hard will exceed a threshold that gets smaller. Greece will then remain at a six-mile limit.
7. If Greek war costs decrease, it is likely that a high Greek belief of facing the Hard will not exceed a threshold that gets bigger. Greece will then extend its territorial waters to 12 miles.
8. If Greek probability of victory increases, it is likely that a high Greek belief of facing the Hard will not exceed a threshold that gets bigger. Greece will then extend to 12 miles.

The special feature of the remaining equilibria is that both types of Turkey take the same action. Either they both accept or fight against the Greek extension. In the third equilibrium, Greece has a relatively low likelihood of prevailing in a war and both types of Turkey fight including the Soft. The equilibrium puts two conditions upon a Greek likelihood of prevailing: not only must q be smaller than a threshold but it also has to be smaller than half of Greek war costs. Letting for example $q = 1/3$ one obtains $1/3 < g/2$, and this implies that $g > 2/3$. Under these conditions, Greece keeps the extent of its territorial waters at six miles regardless of its belief about Turkish war costs. Hence, it is not Greek misperceptions but mainly Greek costs of war and the likelihood of prevailing that drive Greek behavior.

The Greek action changes when the likelihood of prevailing exceeds half of Greek war costs, even though it is still smaller than the threshold. Greece extends its territorial waters, given that both types of Turkey respond by fighting. This is the fourth equilibrium. There is no deterrent effect of the Turkish declaration of a casus belli. The failure of the deterrence stems solely from the relationship between Greek war costs and a Greek likelihood of victory. In the last equilibrium, Greece is almost certain that it will prevail in a war and extends its territorial waters. As the Greek likelihood of prevailing is high, the Turkish likelihood becomes quite small. Regardless of its belief about Turkish war costs, Greece extends and Turkey does not militarily react to Greek action. Both types of Turkey accept the new status quo. The status quo is then peacefully revised. Hence, it is possible that a new status quo emerges in the Aegean Sea without war.

The balance of military capabilities is an important indicator of the probability of prevailing in a conflict and war costs. Huth and Russett [8] take the balance of active manpower as an indicator of the balance in general. According to Smith *et al.* [9] the balance of military capability indicates the probability of prevailing in a conflict. Huth *et al.* [7] maintain that the country enjoying superiority in terms of the balance of conventional capabilities has a higher probability of victory and would also suffer lesser war costs. Turkey possesses the second largest army in the North

Atlantic Treaty Organization (NATO), surpassed only by the United States. Turkey's population and size of total armed forces constantly surpass those of Greece by a wide margin.[5]

As the balance of forces favors Turkey, it can be assumed that the Turkish probability of prevailing in a conflict is high and Turkish war costs are low. Accordingly, as the first and the second cases rely upon a relatively high Greek likelihood of prevailing, the third and the fourth equilibria become plausible. The balance of power in favor of Turkey also implies a high Greek perception of facing the Hard. However, even if a Greek likelihood of prevailing is small and both types of Turkey fight against the extension, Greece can still decide to accept 12 miles, as the fourth equilibrium demonstrates.

6 Policy Implications

How could the game model be applied to be useful for policymakers and analysts? Tanter [10] indicates that the relevance of a model (that is, its feasibility and applicability to current policy needs) is a necessary condition for use by decision makers and policy analysts. Accordingly, these individuals should judge that the game models are an important issue and that assumed preferences, information conditions, and policy options are realistic. For example, if both countries live in a particularly cooperative period, policymakers, either Greek or Turkish, may hold that the territorial waters issue is not vital. Further, policymakers could argue that both countries possess a far larger range of options than specified by the model. The challenge is first to convey an understanding and then an appreciation of the abstract feature of formal models. When this has been appreciated, policymakers can be encouraged to construct their own models that would enable them to generate alternative preferences and policy options. In this sense the model is a tool for generating competing explanations of deterrence and its failure in the Aegean issue and, where possible, for guiding official policies in desired directions.

The model could also serve organizational purposes [10]. If, for example, the ministry of foreign affairs is interested in informational asymmetries in securing the status quo, then the model may well contribute to the organization's planning. The same model could also contribute to the planning done by military and civilians in the ministry of defense. Foreign policy is often a result of deliberations among a variety of agency representatives or political leaders [2]. The appreciation of the model at a practical level would then depend upon the intricate organizational and bureaucratic conditions over which the modeler has limited control. Nevertheless, it is still possible to deal with such complications by using evolutionary frameworks [4].

[5] For example, Greece and Turkey's population in 2000–2001 totaled 10.7 million and 66 million, respectively. The figures for 1969–1970 were 8.8 million and 34 million, respectively. In 1969–1970 the size of Greek and Turkish armed forces (on active duty) were, respectively, 159,000 and 483,000. In 2000–2001 they became 159,170 and 609,700, respectively. (*Source: The Military Balance, Institute for Strategic Studies, London.*)

Finally, Greek and Turkish policymakers and analysts may construct different games for the Aegean issue. The game at hand models the Turkish reaction to a possible Greek extension of territorial waters and subsequent outcomes, including maintaining the status quo, revising the status quo, and war. A Greek version might differ. Unlike the game which models the type of Turkish reaction to possible Greek moves, Greek foreign policymakers may be interested in how to react to possible Turkish moves. For example, Greek practitioners and analysts might be interested in the timing of an extension decision or in linking the issue to Turkey's potential entry into the European Union. They could try to forecast Turkish responses to these moves by constructing simple games. In general, these models would reflect both similarities and differences between Greek and Turkish perceptions of relevant problems with regard to this issue.

7 Conclusion

Game models discipline strategic calculations. The game, though simple, goes beyond disciplined and rigorous deductions: it demonstrates the possibility of taking such subjective elements as misperceptions into account in a formal model. Intuitively, practitioners might think that Greek beliefs about Turkish war costs are quite important in bilateral relations. Yet we found that Greek beliefs do not matter in the continuation of the status quo. Greece is deterred from extending its territorial waters as long as its war costs reduce its probability of winning a military confrontation with Turkey. If Greek war costs get smaller, then deterrence fails and war occurs. Hence, unlike Greek misperceptions, Greek war costs determine to a great extent the state of Greek–Turkish dispute over the Aegean Sea.

It is possible to extend the model by introducing two-sided misperceptions, so that Turkey may also misperceive Greek war costs. Such a model could imply different equilibria that are dependent on or independent of Greek and Turkish misperceptions. Güner and Druckman [5] offer such a game that models a bronze-age interaction. The extensions are numerous; in fact, they could add additional states into the interaction, they could modify information conditions, and options available to the decision makers. Of course, more parameters have to be estimated, and the results cannot be interpreted as easily as before. Therefore, it is reasonable that one builds a simple model initially and relaxes after making some assumptions pertaining to more complex interactions. One would possibly find that new assumptions do or do not have an impact upon equilibria. When comparing models, practitioners may satisfy their curiosity about the consequences of different presumptions. Thus, game models do not only imply disciplined thoughts but may also function as research appetizers.

References

1. Ahnish F (1993) The international law of maritime boundaries and the practice of states in the Mediterranean Sea. Clarendon Press, Oxford, UK

2. Allison G T (1971) Essence of decision: explaining the Cuban missile crisis. Scott, Foresman and Company, Glenview, IL, USA
3. Bahcheli T (2000) Turkish policy toward Greece. In: Makowsky A, Sayari S (eds) Turkey's new world. 131–152, Washington Institute for Near East Policy, Washington DC, USA
4. Güner SŞ (2004) Aegean territorial waters conflict: an evolutionary narrative. Conflict Management and Peace Science 21:297–312
5. Güner SŞ and Druckman D (2000) Identification of a princess under incomplete information: an Amarna story. Theory and Decision 48:383–407
6. Harsanyi JC (1967–68) Games with incomplete information played by Bayesian players. Management Science 14 (Series A):159–182, 320–334, 486–502
7. Huth P, Bennett S, Gelpi C (1993) The escalation of great power militarized disputes: testing rational deterrence theory and structural realism. American Political Science Review 87:609–623
8. Huth P, Russett B (1993) General deterrence between enduring rivals: testing three competing models. American Political Science Review 87:61–73
9. Smith R, Sola M, Spagnolo F (2000). The prisoner's dilemma and regime-switching in the Greek–Turkish arms race. Journal of Peace Research 37:737–750
10. Tanter R (1972) Policy relevance of models in world politics. Journal of Conflict Resolution 16:555–583
11. Wilson A (1979) The Aegean dispute. Adelphi Papers no. 195, International Institute for Strategic Studies, London, UK

Part III

Models in International Negotiations

These models are heuristic and dynamic by their very nature; they are established with the help of external data and they can be used to guide the process of negotiations. They can be seen as making important contributions to the diagnosis and formulation of negotiation positions. Insofar as they are more decision analytical than game-theoretical, they help to answer questions such as: What are the alternatives? What are the uncertainties? What are the multiple conflicting objectives?

Two chapters are devoted to this type of formal model, and they could not be more different: IIASA's RAINS model is a prominent example, as is the contribution of the MIT Model to the Law of the Sea negotiations.

Some Lessons from the Use of the RAINS Model in International Negotiations

Markus Amann

International Institute for Applied Systems Analysis (IIASA) Laxenburg, Austria

1 The RAINS Model

The Regional Air Pollution INformation and Simulation (RAINS) model provides a consistent framework for the analysis of strategies to reduce emissions of air pollutants focusing on acidification, eutrophication, and ground-level ozone [9]. The RAINS model combines information on the anthropogenic driving forces of emissions (with databases on current and future economic activities, energy consumption levels, etc.) on the sources of emissions, on emission control options and costs, on the atmospheric dispersion of pollutants, and on environmental sensitivities (i.e., databases on critical loads). In order to create a consistent and comprehensive picture of the options for simultaneously addressing the three environmental problems, the model considers emissions of sulfur dioxide (SO_2), nitrogen oxides (NO_x), ammonia (NH_3), and volatile organic compounds (VOCs). A schematic diagram of the RAINS model is displayed in *Figure 1*.

The RAINS model follows air pollutants from their origin to their environmental effects. To capture the specifics of emission-generating activities, the European implementation of the RAINS model incorporates databases on energy consumption for 42 regions in Europe, distinguishing 22 categories of fuel use in six economic sectors. Industrial activities distinguish about 40 sectors, and agriculture is modeled by 20 types of activities. The time horizon extends from 1990 until 2020. With this statistical information about emission-generating activities, RAINS estimates sectoral emissions of SO_2, NO_x, NH_3, and VOCs based on the CORINAIR inventory of the European Environmental Agency [5] and on national information. RAINS includes a large number of options for controlling emissions, represented by their characteristic technical and economic features.

Atmospheric dispersion processes over Europe for sulfur and nitrogen compounds are modeled on source–receptor relationships derived from the European Monitoring and Evaluation Program (EMEP) model developed at the Norwegian Meteorological Institute [3]. The formation and long-range transport of tropospheric ozone is described by nonlinear source–receptor relationships between the emissions

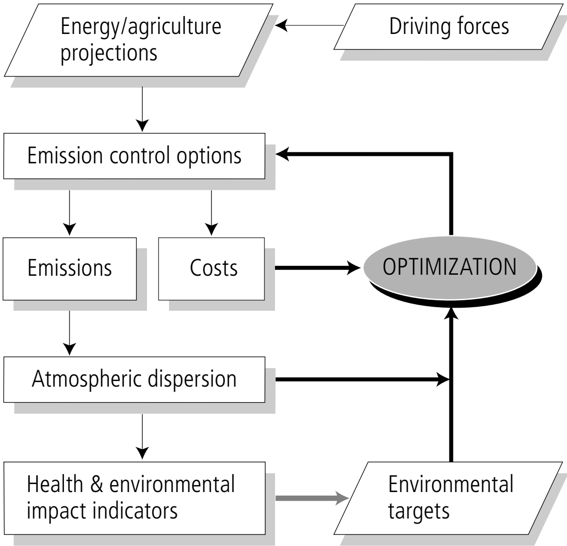

Fig. 1. Schematic flowchart of the RAINS model framework

of NO_x and VOCs and regional long-term ozone concentrations that were statistically derived from the EMEP photo-oxidants model [6].

Fields of deposition and ambient concentrations of pollutants resulting from a particular emission scenario are subsequently compared against thresholds (so-called critical loads and levels (see [8]), that are considered according to current knowledge not to cause negative impacts to human health and the environment. The RAINS model incorporates databases on critical loads and critical levels compiled at the Coordination Centre for Effects (CCE) at the National Institute for Public Health and Environmental Protection (RIVM) in the Netherlands [7].

The RAINS model can be operated in a "scenario analysis" mode (i.e., following the emission pathways from their sources to their environmental impacts). In this case the model provides estimates of regional costs and environmental benefits of alternative emission control strategies. Alternatively, an "optimization mode" was developed to identify cost-optimal allocations of emission reductions in order to achieve specified environmental targets [2]. These two modes correspond to the familiar phases of the negotiation process into diagnosis and formulation [13]. In the policy applications of RAINS, this optimization or formulation mode turned out to be a powerful tool to identify cost-effective solutions, even though it put a high demand on an efficient interaction between modelers and decision makers to reach a full understanding of the optimization rationale that is required for the acceptance of its outcome.

2 Use of RAINS in International Negotiations

2.1 The first round of model application: The negotiations on the Second Sulphur Protocol

The need for an integrated assessment model to provide a scientific basis for emission reductions under the Convention on Long-range Transboundary Air Pollution was apparent by 1985. For various reasons the executive body of the Convention decided to have the RAINS model perform most of the analyses underlying the negotiations over the Second Sulphur Protocol. The results of two other models (i.e., the Abatement Strategies Assessment Model (ASAM) developed by Imperial College, London, and the Coordinated Abatement Strategy Model (CASM) of the Stockholm Environment Institute) were used for comparison. RAINS was selected because it was more fully developed than the other two models at the start of the negotiations; its results would be credible to countries both in Eastern and Western Europe because of the international constituency of the International Institute for Applied Systems Analysis (IIASA), the organization developing the model and because a 1991 workshop on the model's usability had already given potential users some familiarity with it [11].

To give the negotiations over the Second Sulphur Protocol a firm scientific basis, the executive body formed a special Task Force on Integrated Assessment Modeling to oversee the modeling of the optimal amount of emission reductions (and their distribution over space) as well as the costs and benefits associated with those reductions. The Task Force reviewed the RAINS model, analyzed and interpreted results of the RAINS model for specific scenarios, and compared them with results from the other models. The simulations themselves were conducted in close collaboration with the Working Group on Strategies, where the negotiations between the Parties took place. The emission reductions of the ultimate scenario, aiming at closing the gap between the 1990 deposition and the critical loads by 60 percent, were accepted by most countries in accordance with their obligation laid down in the Protocol in 1994.

At the beginning of the negotiations over the Second Sulphur Protocol there was a great deal of concern over the uncertainties in the simulations with RAINS and the other models. Concerns were raised about (i) the usual uncertainties in the data, including gaps in observable data and the necessary imprecise predictions of future development, (ii) the steady-state nature of the critical loads concept, which ignores the dynamic nature of biological and chemical processes in the nature, and (iii) the fact that RAINS did not include structural changes in the energy sector as a means of controlling emissions. As time went by, however, the negotiators shifted their attention to the political assumptions and other constraints adopted in the optimization exercises. One reason for this shift may have been the growing realization that the uncertainties in such inputs were greater than those in the models themselves. It is also possible that some delegates were simply raising concern about uncertainty as a means of delaying a vote on the Protocol until there was more political support at home [11].

2.2 The second round of model application: The Acidification Strategy of the European Union and the Gothenburg Protocol to Abate Acidification, Eutrophication, and Ground-level Ozone

After the signature of the Second Sulphur Protocol in 1994, the Convention on Long-range Transboundary Air Pollution focused on a revision of its First NO_x Protocol, which had entered into force in 1991. By that time, scientific evidence had demonstrated that emissions of nitrogen oxides have multiple effects on the environment. Notably, they contribute to acidification and eutrophication (excess fertilization) of terrestrial and aquatic ecosystems and play a central role in the formation of ground-level ozone. It was found that, under certain conditions, in ozone plumes, reductions of NO_x emissions in urban areas could lead to further increases in ozone.

The multipollutant/multieffect concept of the RAINS model offered a clear concept for addressing the multifaceted nature of NO_x controls: RAINS offered an operational method for designing emission-control strategies that simultaneously addressed acidification, eutrophication, and ground-level ozone. The model facilitated the development of coordinated emission controls that yielded positive-sum benefits to all parties for all three environmental effects, despite the potentially counterproductive response of ozone formation to reductions in NO_x emissions.

A workshop organized by IIASA for the negotiators of the Convention on Long-range Transboundary Air Pollution clearly demonstrated the strong need for considering further controls of sulfur dioxide emissions if such multipollutant/multieffect strategies were to be cost-effective. Although the recent signing of the Second Sulphur Protocol was not on the immediate agenda of the negotiations under the Convention, the Parties accepted this requirement and agreed to strive for a multipollutant/multieffect protocol by including further measures for SO_2 emissions.

It is important to realize that the evidence provided by integrated assessment models about the potential gains of an integrated approach in a complex situation, where imbalanced emission controls could lead to a deterioration in environmental conditions, convinced decision makers to revise their political agenda and to use integrated assessment models for exploring cost-effective solutions. It is also clear that this decision was taken in full awareness—and acceptance—of the uncertainties of integrated assessment models.

While the Convention was preparing for a multipollutant/multieffect protocol, the European Union (EU), after the accession of Austria, Finland, and Sweden, embarked on parallel discussions about the orientation of its further clean air policy. At the European Council meeting in March 1995, Sweden requested a report from the EU Commission that would include an assessment of the impact of current and proposed EU legislation on acidification and, as a follow-up to that report, an EU acidification strategy. Until then, key personnel in the Commission and the Environment Directorate had almost exclusively favored a "best available technology" approach [12]. This approach, primarily developed by Germany, is more interested in the emissions side of the issue and its related technological options than in the environmental side, as manifested in the thinking on critical loads. For its communication on an EU acidification strategy [4], the Commission used the RAINS integrated assessment model to build an analytical bridge between the techno-economic aspects of emission control strategies and their environmental impacts. This bridging concept was appreciated by the European Parliament and the Council, and in 1997 the Commission began to prepare a proposal on a Directive on National Emission Ceilings using the RAINS model as the central analytical tool for deriving quantitative emission caps for the member states [1].

From 1997 to 1999, analysts at IIASA used the RAINS model for an iterative analysis of more than 100 emission-control scenarios. This was done in close interaction with negotiators at the Working Group on Strategies of the Convention on Long-range Transboundary Air Pollution and, in parallel, with the staff at the European Commission and the various working groups established by the Commission involving representatives of member states and other stakeholders. In total, the analysis ran through 11 iterations, where decision makers requested model calculations, analyzed their results, and suggested modified scenario runs. All reports produced by IIASA for these negotiations are freely available on the Internet (http://www.iiasa.ac.at/rains).

The shared use of IIASA's RAINS model for both activities helped to maintain consistency between the work of the Convention on Long-range Transboundary Air Pollution and the European Commission. Both bodies used similar assumptions and environmental objectives, although the analysis for the Convention obviously included countries that were not members of the EU. Finally, both processes resulted in similar emission ceilings, with the Emission Ceilings Directive of the EU being slightly more demanding than the Gothenburg Protocol to Abate Acidification, Eutrophication, and Ground-level Ozone. The interplay of these two policy processes is analyzed in Wettestad (2002), which also provides detailed comparisons on the obligations for individual countries.

3 Some Lessons from Using a Formal Model in Negotiations

With these applications, the RAINS model is a prominent example of a formal model that was repeatedly used in international negotiation processes. The following discussion will focus on some aspects that contributed to the success of the model developers in having their model accepted by the Parties and used to guide international negotiations. By no means should this list be seen as exhaustive. It solely represents the views of the model developers.

3.1 The focus/setting of the formal model

The RAINS model began receiving the interest of negotiators after its capabilities were extended from the initial "scenario analysis" to an "optimization" mode. While the scenario analysis mode was useful in prenegotiation diagnosis to illustrate the economic and environmental consequences of an exogenously assumed pattern of emission controls, the optimization features provide the basis for thinking about formulas by systematically identifying the least-cost allocation of emission controls that meet exogenously determined environmental targets.

Within the scenario or diagnosis mode, the number of "what-if" scenarios that could be explored with the RAINS model was limited, which made it impossible to examine fully the consequences of even the most important permutations of emission-control measures in all economic sectors of close to 48 Parties. Thus, in practice the scenarios addressed a limited number of technology-related emission-control rationales, but could not add to a systematic analysis of environmentally driven emission-control strategies that were in the focus of the Convention after the NO_x Protocol. Although the main feature of the scenario mode was the assessment of the environmental effects of emission controls, the quantification of the latter was hampered by methodological problems in the spatial downscaling of the impact assessment, which did not allow effects to be predicted for specific ecosystems. Consequently, the pure scenario analysis provided only limited insight to negotiators who had to find distributions of emission-control obligations across countries that were acceptable to all the Parties.

The situation changed as soon as the optimization feature of the RAINS model was developed, which made it possible to identify distributions of abatement burdens across Parties that were most "efficient" according to a selected rationale. The RAINS optimization provided the ideal complement to the "critical loads" concept that was accepted by the Convention in the First NO_x Protocol as a rationale for future emission-control agreements because it allowed the least-cost allocation of measures to be determined that would achieve environmental targets established in terms of critical loads. Thus, the optimization concept became an important element of a "science-based" rationale that was required as a basis for the coming emission-reduction accords. By calculating country- and sector-specific reduction requirements for any exogenously specified environmental target, the RAINS optimization provided results that were of immediate relevance to the negotiators because they met the spatial and temporal scales that were relevant for decision makers. The optimization was also attractive because, while striving for a common target (equal environmental improvement for all parties), it considered the environmental and economic differences among Parties that led to objectively justifiable differences in abatement efforts. Resulting inequities in abatement burdens were based on scientifically determined differences in environmental sensitivities, atmospheric dispersion characteristics, or emission-source structures. The negotiators could focus their negotiations on the ambition level of their environmental objectives, the political acceptability of the implied costs, and their distribution across Parties, while they could leave technicalities (the quantification of objective differences among countries) to the formal model. The model was seen as a common knowledge base that allowed negotiators to focus on the policy issues. ("Let us put the facts on the table; we will fight about politics later.")

It was also important that the optimization problem as set up in the RAINS model did not provide an absolute and unique answer to the air pollution problem. The actual results of an optimization run depended on the environmental objectives (e.g., the acceptable environmental risk) as established by the negotiators, the goal function (minimization of total emission control costs), and the problem framing (e.g., the exclusion of changes in the energy systems, which could not be directly influenced by environmental policies in Europe). All these settings were subject to negotiations, and the optimization results were critically influenced by the policy choices on these issues. Thus, the RAINS model did not determine policy choices but deliberately left room for the decisions of negotiators.

To find its place within the negotiation process, it was crucial that the RAINS model did not attempt to impose the choice of the optimum abatement level (i.e., the level of environmental ambition or of acceptable risk) and thus refrained from a full-fledged cost–benefit analysis in which the optimum abatement would be determined by the marginal (monetized) benefits. Instead, the choice of the appropriate control level was left to decision makers, while the model was used to search for the least-cost solution to meet the environmental targets of the negotiators.

3.2 The use of the RAINS model in the negotiations

The RAINS model was considered as a tool that provided guidance to the negotiators in searching for acceptable compromise solutions. The modeling team of IIASA actively participated in the negotiation sessions of the Working Group on Strategies on a routine basis and communicated model results to the negotiators, responding to questions and proposals made by the negotiators. This direct interaction between modelers and negotiators emerged in several iterations, where the Working Group on Strategies formulated proposals for environmental targets, which were then subsequently evaluated by the RAINS model along with their implied emission controls and their distributional aspects across the Parties. These results were then reported back to the negotiating group, which, if it found them unacceptable, proposed alternative environmental targets. The documentation of all model runs was made available to the negotiators before the sessions. In addition, IIASA implemented free and direct access to an online version of the RAINS model via the Internet, which enabled interested Parties and stakeholders to evaluate their own positions as well as the consequences of proposals made by other Parties. In 1999, during the negotiations for the Gothenburg Protocol and the Emission Ceilings Directive, more than 8,000 model runs were registered by IIASA.

It was important for the ultimate acceptance of model results that the modeling team at IIASA responded to concerns of decision makers about certain aspects of the model analysis, and implemented a number of improvements during the negotiation process. For instance, in response to concerns on the robustness of model results expressed by the European Council of Environment Ministers in its negotiations of the acidification strategy, the modeling team developed a “compensation mechanism” that relaxed the spatial specificity of the environmental improvements that drive the optimization, while safeguarding the overall level of improvements in each country.

3.3 Elements contributing to the acceptance of the RAINS model for use in negotiations

The following elements could be considered as critical for reaching acceptance of the model for use in negotiations:

- The reliance on peer-reviewed disciplinary science as the basis for the RAINS integrated assessment model.
- The peer review of how RAINS performed the integration of the multiple aspects that are relevant for the design of cost-effective emission control strategies, and how RAINS treats uncertainties.
- The review and formal acceptance of all data in the RAINS model by all the Parties.
- Open access to the model.
- IIASA's role as an impartial broker with an interest in the integrity of the process, but not in the outcome per se.
- The large network of IIASA alumni who participated in the development of the RAINS model at different stages and who returned to their home countries, taking expertise on the model back with them.
- Many of the Parties to the Convention on Long-range Transboundary Air Pollution are also members of IIASA and therefore feel involved in the Institute's modeling activities.

3.4 Who are the actors in the negotiations?

The interaction between modelers and decision makers is often seen simplistically from the outside as a group of disciplinary scientific specialists advising top-level politicians (e.g., ministers) on actions to be taken. If formal models are used in the negotiations, a group of politicians is seen sitting behind computers.

At least in the European negotiations on air pollution, however, the reality was somewhat different. To understand the role of formal models in negotiations it is necessary to understand who the actors are, their background, and their role in the negotiations. One could usually distinguish at least four groups of actors:

- Disciplinary scientists, who are specialists in their own disciplines and who undertake research on specific aspects of air pollution.
- Integrated assessment modelers, who distill the findings from disciplinary scientists and put them into consistent relations with each other.
- National representatives at the negotiating meetings (e.g., civil servants) who are familiar with the substance (often with an academic background in a related field) and who represent the interests of countries in the negotiations. These people regularly attend, for example, the Working Group on Strategies of the Convention and prepare/negotiate the draft agreements. They also communicate progress made at these meetings (including new insights from the modeling exercise) to their home countries and are responsible for the understanding that model results that influence negotiating positions will be shared by all relevant actors in the country.
- High-level decision makers (politicians, sometimes ministers) who enter the negotiations only in critical phases, negotiate critical compromises, and sign protocols.

From the perspective of model development, the integrated assessment modelers had to maintain close relationships with the disciplinary scientists, whose findings they are using in their model, as well as with the national representatives, who need to be confident about the model and its results and to be able to communicate these back home. There was less direct interaction between disciplinary scientists and negotiators/politicians and between modelers and politicians—an important gap also present in the Law of the Sea negotiations analyzed in the chapter by Antrim in this publication.

One should not underestimate the role of industrial stakeholders. Industrial stakeholders lobby with decision makers to prevent unfavorable outcomes on a regular basis, although in most cases not in official meetings. For reaching widespread acceptance of model results, IIASA kept industrial stakeholders involved in the modeling activities, either through consultations on critical issues or through active participation in model reviews.

A critical role was also played by the Secretariat of the Convention, where the members of the Secretariat, many of them with significant personal scientific and modeling experience, actively supported the use of the RAINS model in the negotiation process. This helped to anchor the RAINS modeling activities within the formal structure of the working groups of the Convention, so that over time the entire working structure of the Convention developed into an integrated assessment process that is supported by a formal integrated assessment model, the RAINS model. *Figure 2* outlines the recent organizational structure of the Convention that was implemented after the signature of the Gothenburg Protocol. At that time it was decided to bring the Task Force on Integrated Assessment Modeling, whose principal role is to review the RAINS model, and the Centre for Integrated Assessment Modelling (CIAM), which is hosted by IIASA, under the EMEP umbrella, thereby extending the modeling mandate of EMEP from atmospheric dispersion to integrated assessment modeling. Thus, after 15 years of integrated assessment activities, the Convention officially recognized the value of integrated assessment modeling and incorporated this activity into its institutional arrangements.

In the context of the European Union, the European Commission had an active interest in using the RAINS model for its policy analyses. The Commission provided substantial funding to IIASA's modeling team to carry out the required analyses. The Commission staff maintained close contact with IIASA in defining scenarios and presented them to the Council and to the European Parliament.

While the institutional setting is important for the use of formal models in negotiation processes, one should not underestimate the importance of personal relationships between all players in the process. Building personal trust between modelers, negotiators, Secretariat, and stakeholders is an indispensable prerequisite for having model results accepted to guide negotiations. Such trust can be established only over the long term and requires the collective experience shared by all actors going through iterative phases of model applications.

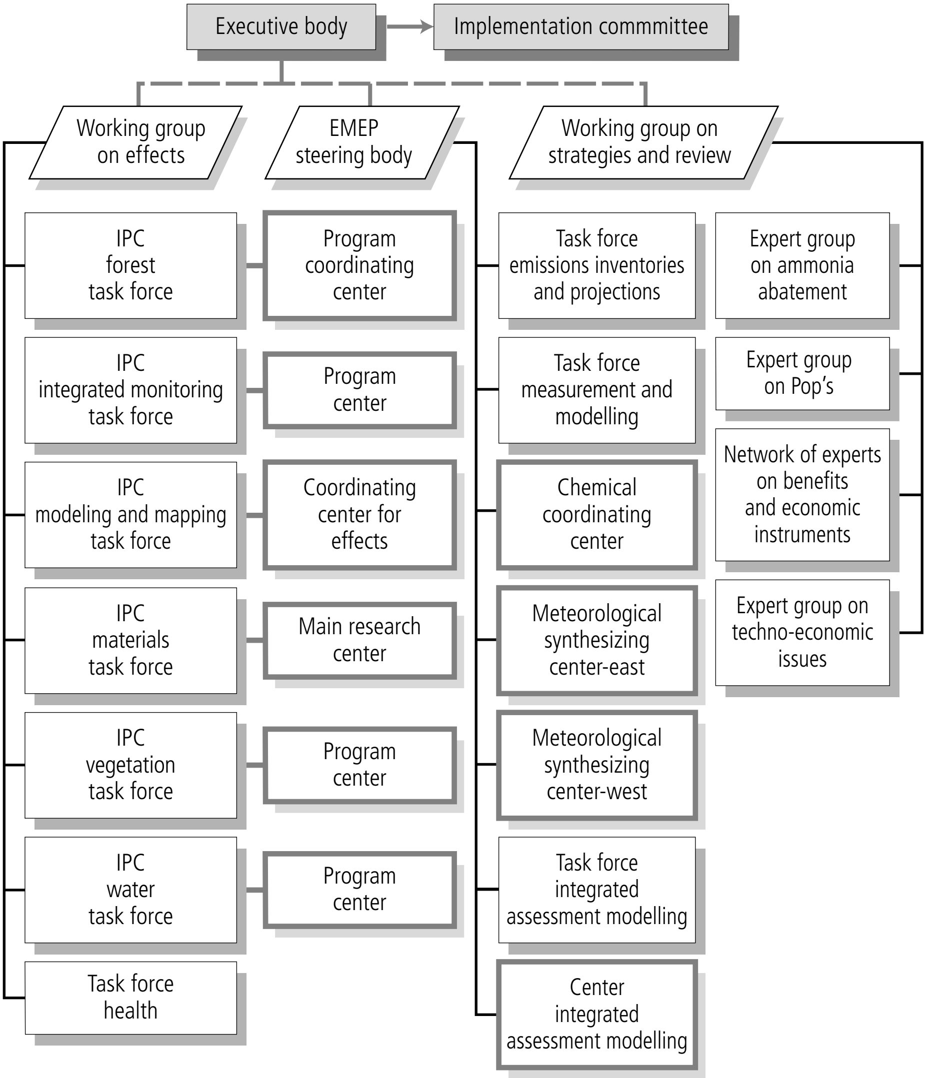

Fig. 2. The organizational structure of the work under the Convention on Long-range Transboundary Air Pollution (as of 2003)

3.5 The role of uncertainties

Uncertainties and robustness are crucial aspects of the development and application of integrated assessment models in a policy process. One can distinguish six steps toward uncertainty management in an interactive learning process, with good interaction among all actors involved:

- Denial of uncertainties;
- Acknowledgement of uncertainties;
- Specification of types of uncertainty;
- Quantitative assessment of uncertainties;
- Specification of policy relevance;
- Uncertainty management through negotiation.

There are different types of uncertainty. Some can be handled with statistical and other methods; others cannot be dealt with in the same manner. It is important to follow a systematic approach to uncertainties in order to gain confidence. Such an approach should differentiate between the reducible and the irreducible uncertainties. For the most significant sources of reducible uncertainties, it should determine by how much further scientific effort could increase the robustness of the models. For irreducible uncertainties, it should be made specific how they could influence the model outcome, for example, through alternative scenarios.

Policymakers are aware of uncertainty. They are interested in the sources of uncertainty and whether/how they can be reduced. Decision makers are looking for a rational basis for decisions, but various driving forces often dominate specific decisions. In many cases, uncertainties are not the most important argument. The reliance on model results will be higher (independent of uncertainties) if model results fit the political driving forces. But decision makers might opt for more uncertain solutions, if they match their policy objectives (as was the case, for example, with acceptance of carbon sinks for the Kyoto accounting scheme). There is a risk that uncertainty can be misused as an argument for delay, when there are opposing scientific and/or political positions.

Policymakers, in contrast to scientists, are not interested in the detailed statistics about uncertainties. They are interested in robust strategies. Robustness implies that strategies (control needs and priorities between countries, sectors, pollutants) do not significantly change because of changes in the uncertain model elements. Robust strategies should avoid regret investments (no-regret approach) and/or the risk of serious damage (precautionary approach).

A good process can help (and has helped) to deal with uncertainty: transparency, participation, and consensus building around scenarios. Much of this was done in preparing the Gothenburg Protocol.

3.6 The treatment of uncertainties in the RAINS model

For the use of the RAINS model for the Gothenburg Protocol negotiations, the model developers considered uncertainty management as an important guiding principle during the model development phase and adopted a variety of measures in model design and scenario planning to systematically minimize the potential influence of uncertainties on policy-relevant model output. For instance, at all phases of model development and use, explicit confidence intervals (for emission-control potentials, deposition ranges, ozone levels, ecosystems sensitivities, etc.) defined the range within which the model was proven to work with sufficient accuracy. Potential reliance on optimized solutions for single point estimates were avoided through integral measures for environmental sensitivities. Specially designed compensation mechanisms allowed controlled violation of environmental targets for single ecosystems with potentially uncertain sensitivities. Wherever possible, preference was given to relative model outcomes (comparing two model outputs) rather than to absolute values. For ground-level ozone, less weight was given to extreme meteorological situations because their representativeness was questionable and the performance of the meteorological model for such rare situations was less certain.

In addition to the consideration of uncertainties within the framework of the model itself, attempts were made to limit the effect of uncertainties on the model optimization outcome by selecting an appropriate method of setting the optimization targets. In practice, the potential influence of uncertainties was minimized by using “gap closure” targets (relative improvements), by developing a compensation mechanism for targets, using explicit model confidence intervals, and excluding extreme situations.

Sensitivity analysis attempted to identify systematic biases and showed that with large probability the emission reductions resulting from the model calculations could be considered as minimum requirements, suggesting that there is only little chance that policy measures suggested by the model needed to be revised in the future in the light of new information.

In 2001 the RAINS model was expanded to examine how errors (quantified uncertainties) in the input parameters propagated through the model for ecosystem protection against acidification [10]. One of the most important findings is that integrated assessment model results tend to be more robust than their input parameters. This is because of a compensation of statistically independent errors. However, it was also found that, because of limited information, one of the most uncertain elements of the uncertainty analysis is the quantification of the uncertainties themselves.

Some questions about the usefulness of such an approach remain:

- Are the Parties ready to put increased effort into providing and subsequently agreeing upon the data needed for such an analysis?
- Would the Parties be prepared to follow abatement strategies derived with such a method (i.e., pay more for strategies that yield the same environmental improvements but with a higher probability of attainment?).

4 Conclusions

A number of lessons can be learned from the use of the RAINS integrated assessment model in international negotiations:

1. Integrated assessment is a long participatory process, not a one-off event. Integrated assessment models can support such a process (and they can become crucial elements of such a process), but they cannot replace the process. Thus, it is unlikely that a single model developed and applied in isolation from a policy process would be accepted as a common tool in international negotiations.
2. It is important for modelers to recognize the potential role of a formal model in a negotiation process. At least in the case of the RAINS model, the formal model was seen as a tool, that could be used by the negotiators to facilitate and quantify their deliberations. It is questionable if the interest of negotiators would have been as high if the model had delivered ready solutions and where nothing was left for negotiations.
3. This also underlines the importance for modelers of acknowledging the specific roles of the various actors in the negotiation process and of designing the model in such a way that it provides practical assistance to the actual questions of these actors. In the case of the RAINS model, the choice of the appropriate spatial and temporal scales and the extension of the model to the optimization approach beyond scenario analysis were critical.
4. Finally, the treatment of uncertainties is an important aspect when a formal model is applied in negotiations. It is vital that the meaning of uncertainties of the formal model and its data is seen in the negotiating context and that solutions are developed that accommodate the main concerns of negotiators. Such solutions might be different from conventional scientific approaches for uncertainty analysis, but they might transmit more relevant information to negotiators.

It is the opinion of the developers of the RAINS model that for many of the aspects mentioned above, the success of the RAINS model in international negotiations is related more to the permanent and close dialog between model developers and negotiators than to the existence of perfect model machinery.

References

1. Amann M, Lutz M (2000) The revision of the air quality legislation in the European Union related to ground-level ozone. Journal of Hazardous Materials 78:41–62
2. Amann M, Makowski M (2001) Assessment of air quality. In: Wierzbicki A, Makowski M, Wessels J (eds) Model-based decision support methodology with environmental applications. Kluwer Academic Publishers, Dordrecht, The Netherlands, Boston, MA, USA, London, UK
3. Barrett K, Sandnes H (1996) Transboundary acidifying air pollution calculated transport and exchange across Europe 1985–1995. In: Berge E (ed) Transboundary air pollution in Europe. EMEP/MSC-W Report 1/96, Meteorological Synthesizing Centre - West, Norwegian Meteorological Institute, Oslo, Norway

4. CEC (1997) Communication of the Council and the European Parliament on a Community strategy to combat acidification. COM(97) 88 final, Commission of the European Communities, Brussels, Belgium
5. EEA (1996) Joint EMEP/CORINAIR'90 Atmospheric Emission Inventory Guidebook. First edition, Vol. 1–2, European Environmental Agency (EEA), Copenhagen, Denmark
6. Heyes C, Schöpp W, Amann M, Unger S (1996) A reduced-form model to predict long-term ozone concentrations in Europe. WP-96-12, International Institute for Applied Systems Analysis, Laxenburg, Austria
7. Posch M, de Smet PAM, Hettelingh JP, Downing RJ (1999) Calculation and mapping of critical thresholds in Europe. Status Report 1999, Coordination Center for Effects, RIVM, Bilthoven, The Netherlands
8. Nilsson J, Grennfelt P (1988) Critical loads for sulphur and nitrogen. Nord 1988:97, Nordic Council of Ministers, Copenhagen, Denmark
9. Schöpp W, Amann M, Cofala J, Heyes C, Klimont Z (1999) Integrated assessment of European air pollution emission control strategies. Environmental Modeling and Software 14(1):1–9
10. Suutari R, Amann M, Cofala J, Klimont Z, Posch M, Schöpp W (2001) From economic activities to ecosystem protection in Europe. An uncertainty analysis of two scenarios of the RAINS Integrated Assessment Model. EMEP CIAM/CCE Report 1/2001, International Institute for Applied Systems Analysis, Laxenburg, Austria
11. Tuinstra W, Hordijk L, Amann M (1999) Using computer models in international negotiations. The case of acidification in Europe. Environment 41(9):33–42
12. Wettestad J (2002) Clearing the air. Europe tackles transboundary pollution. Environment 44(2):32–40
13. Zartman IW, Berman M (1982) The practical negotiator. Yale University Press, Newhaven, CT, USA

Converting Competition to Collaboration: Creative Applications of Models in the Law of the Sea Negotiations

Caitlyn L. Antrim

Center for Leadership in Global Diplomacy, Arlington, VA, USA

1 Introduction

A scientific curiosity for almost a century, the mineral resources of the deep ocean floor took on a sudden importance in the 1960s and early 1970s when it was claimed there were hundreds of billions of dollars worth of minerals lying on the seabed beyond the limits of national jurisdiction (see [3]). Suddenly this curiosity became the center of attention in efforts to develop a new and comprehensive law of the sea to replace the already outdated and insufficient Geneva Conventions on the Law of the Sea, adopted in 1958.

For developed countries, the mineral resources of the deep ocean floor remained more of a curiosity than a top line issue, but developing countries considered them seriously, either as a potential new source of revenue or as a potential competitor to those developing countries dependent upon mineral-export earnings for their economies. When the Third United Nations Conference on the Law of the Sea (UNCLOS) convened in 1973, three negotiating committees were established. One addressed traditional uses of the sea—naval and commercial transport, fishing, and continental shelf resources. Another, spurred on by the 1972 Stockholm Conference on the Human Environment, focused on marine science and the environment. A third committee was established to address the management of the resources of the seabed beyond national jurisdiction. At the beginning of the conference there was little expectation that this issue, and this committee, would become the keystone for the completion of the desired comprehensive Law of the Sea Convention, but as progress was made on other issues, the disputes between developed and developing countries over seabed resources became more acute and threatened the success of the entire Conference.

The negotiation of the regime for deep ocean resources was aided greatly by the contribution of a technical and economic model of the exploitation of the mineral resources of the deep ocean floor. The importance of the model was attested to by the chairman of the negotiation group and President of the Conference, Ambassador Tommy T.B. Koh, who commented:

> In the group of financial experts we were immediately confronted with the need to agree on a set of assumptions. Without an agreed framework of assumptions it would not have been possible for us to carry on with our discussions. We agreed that the best study to date was that undertaken by the Massachusetts Institute of Technology (MIT), entitled "A Cost Model of Deep Ocean Mining and Associated Regulatory Issues," hereinafter referred to as the MIT Study.

These sentiments were not universally shared, however. Some observers, who opposed the final Convention, felt that the United States (USA) had willingly, if unwittingly, given away valuable knowledge that could have better been used as a partisan tool to obtain better terms from the developing countries. Peter Leitner [2], an observer on the United States (US) delegation, later wrote:

> While a slow, evolutionary process of sophistication within Committee One was underway prior to Geneva, it was possible to influence the conference gradually toward the US position both by waiting for a gradual weakening of the Group of 77 as a dominating political bloc and by selectively releasing information to the conference.

He went on to write:

> It is interesting to note that the motivation behind releasing the model was to prevent, through education, the expression of unrealistic proposals, which would confuse the negotiating process. However, this well-intentioned action brought about exactly what it sought to eliminate.

These two perspectives illustrate the natural tension between collaboration and competition present in negotiations. Leitner's comments, in particular, also display a belief in the adage "knowledge is power" and that such power should be used in pursuit of partisan advantage, while Koh believed that agreement needed to be reached through shared information and mutual compromise.

The MIT model was a successful example of the use of a formal model as a collaborative tool in a complex multiparty negotiation. This success was due to factors related to the issue under negotiation, the model and its creators, and the negotiators and negotiation process that put the model to use.

2 The Issue: Sharing the "Common Heritage of Mankind"

Advances in technology in the 1960s for the exploitation of the resources of the ocean made the mining of minerals on the ocean floor beyond national jurisdiction a real possibility. Almost a century after their discovery, polymetallic nodules (accumulations of manganese and other metals that cover parts of the deep ocean floor) became more than scientific curiosities—they became potentially valuable commodities. This potential, and the rising interest among developing countries in improving their economic conditions, led the United Nations to deem these resources to be the "common heritage of mankind."

It was not until the third UNCLOS was convened that it became necessary to define what was meant by "common heritage." Would seabed resources be developed by private companies or by a single multinational organization? How would revenues be collected and distributed? These and other questions took on increasing importance as ocean mining became the keystone for developing state approval of the rest of the Convention, for which the United States sought near universal acceptance to ensure that its other provisions would become accepted international law regulating the behavior of all countries.

There were many issues to be resolved within the negotiations on seabed mineral resources. Some were approached through the normal give and take of conventional negotiations, but one issue proved to be too stubborn for conventional means. On the subject of the financial issues related to deep-ocean mining, not only did industrialized and developing countries differ in their interests, but they also differed in their basic assumptions as to the profitability of seabed mining. These differences were too wide to be bridged by simple trading of concessions. Instead, technical and economic analyses were used; first, to develop a common set of assumptions; second, to identify possible areas of compromise; and third, to express the costs and benefits of possible compromises in terms of the interests of the participants. In the end, the financial provisions proved to be one of the most successful of the many components of the seabed negotiations; and, in the view of key participants in the process, the computer model that produced the technical and economic analyses was critical to that success (see [5]).

3 The Law of the Sea Negotiations

In the late 1960s and early 1970s, a confluence of interests of the superpowers (interested in restricting the trend toward an extension of national jurisdiction seaward), developing coastal states (interested in maximizing the contributions of ocean resources toward their national development), small industrialized countries (who wanted to establish fishing, pollution, and other regulations that would be recognized and respected by all other countries), and a variety of other countries with varied interests led to a third attempt to establish a universally recognized Law of the Sea. The third UNCLOS was arguably the most complicated negotiation ever undertaken, comprising over 150 countries and over 1,000 delegates and spanning eight years of formal negotiations and several more years of preparation.

The objective of obtaining a universal and comprehensive legal regime for the oceans set some conditions on the negotiations. First, no sovereign state could be forced to accept the outcome of the negotiations. Second, the range of issues that had to be addressed in order to draw all states into the negotiations was extensive. The result was agreement that the Convention would be negotiated as a "package deal" and that every effort would be made to ensure that the agreement was adopted by consensus and that votes on matters of substance would be avoided. On the positive side this allowed for concessions in one issue to be offset by gains in another. However, the complexity of the negotiations was far greater than any previous UN conference, leading Ambassador Elliot Richardson [7] to describe it as being "like playing no-limit poker and three-dimensional chess at the same time."

The Conference began its substantive work in 1974. By 1977 many of the critical issues were resolved, in principle if not in detail. However, commitment by developing countries on these issues was contingent on the negotiation of an agreement regulating the exploitation of the resources of the deep ocean floor.

Polymetallic nodules, which at the time of the Conference were the only seabed resources that had shown potential for economic development, are small, lumpy accumulations of manganese and iron oxides. Such deposits are found in many parts of the ocean, but in one region of the Pacific the accumulations are enriched with nickel, copper, and cobalt. Spurred on by thoughts of the potential value of the metals contained in these nodules, the developing countries, acting through the Group of 77 (a negotiating group established in the UN Conference on Trade and Development (UNCTAD) that had grown to over 100 members by the time of UNCLOS), made the resolution of deep seabed resource issue their quid pro quo for agreement on the rest of the Convention.

By late 1976 the basic framework of the regime for managing the mineral resources of the deep ocean floor was established. An "International Seabed Authority" would be established to implement the agreement, with powers and responsibilities enumerated in the Convention. In recognition of the differing economic systems prevailing among the Conference participants, there would be a "parallel system" in which one system would manage the exploration and exploitation activities of states and their nationals and another system where a new organization, dubbed the "Enterprise" would conduct mining operations under the direction of the Seabed Authority as a means of giving states without their own ocean mining capability the opportunity to participate in the exploitation of the "Common Heritage."

The task facing the negotiators charged with writing the regime into the text of a convention was complicated by two factors. First, as the navigation, fishing, and continental shelf issues were resolved, industrialized country mining companies became more vocal in defense of commercial ocean mining interests and in opposition to the Group of 77. Second, the aftermath of the decision in 1976 of the Chairman of Committee I (responsible for seabed issues) to substitute his own draft text in place of a compromise text negotiated between industrialized and developing countries left a legacy of distrust that was difficult to overcome. In response, the rules of the conference were modified to limit the role of the Committee I chairman by involving the senior officers of the Conference in decisions relating to the text and by assigning responsibility for the resolution of outstanding issues to other conference leaders. The Conference established three special negotiating groups on seabed matters: one on the system of exploration and exploitation, another on the financial arrangements for exploitation, and a third on the structure of the Assembly and Council that would govern the application of the Convention to seabed resources.

Ambassador Tommy Koh, from Singapore, was selected as chairman of Negotiating Group 2 (NG 2) (the negotiating group on financial arrangements). The work of this negotiating group appeared to be particularly difficult for two reasons: first, the results would be explicit—no amount of diplomatic language could hide the cost that countries and companies would have to pay in the event of an agreement, and, second, to many observers it appeared to be a zero-sum game, where the gains of one side would be at the expense of the other.

The group spent its first year laying the groundwork for the negotiation. Earlier, the Conference secretariat had reviewed possible elements of a revenue-sharing regime, but efforts to negotiate a formula to implement them had failed. The United States, India, and other countries presented opening positions that were based on differing assumptions about the economics of deep-ocean mining. Thus, Ambassador Koh began a process of education for himself and for the other members of the negotiating group on the various aspects of ocean mining, with the hope of narrowing the range of disagreement about the potential profitability of mining and identifying areas that could lead to compromise.

Koh addressed economic and political matters in great detail. He faced the need to obtain consensus in the committee of the whole (all of the participating countries) while increasing the involvement of experts in the analysis of issues and the development of proposed solutions. To achieve this he developed a system that constantly moved from inclusion to exclusion and back again. Each morning he would convene Negotiating Group 2 to review the current state of negotiations and to give delegates the opportunity to respond with statements and questions. In the afternoon he would convene a "Group of Financial Experts" (the GFE) to discuss the details of proposals in depth. While the GFE was open to all countries, Koh's focus on technical detail and his selection of a smaller and less comfortable meeting room led many delegates to defer to trusted colleagues on other delegations to participate in this group (knowing that every morning they would learn of the work of the GFE). Even the GFE proved to be limited in its ability to create new approaches, so Koh established his own group of technical advisors, drawing upon individuals from the UN and UNCTAD secretariats and a few respected analysts from national delegations. These people formed Koh's creative team, which developed proposals for Koh to introduce in the GFE, where they could be discussed in depth. When proposals gained support in the GFE, Koh would present them for discussion in the morning sessions of NG 2. Eventually, when a new proposal gathered support, Koh would introduce it into his proposed compromise agreement and then begin the process again. In addition, Koh established a group of senior national representatives and secretariat advisors who represented the major interest groups and who, having the respect of their colleagues, could help him move his proposals toward agreement.

In early 1978, disagreements about the likely profit from ocean mining resulted in continued support of proposals at the extreme high and low levels of revenue sharing. The negotiations were difficult because there were no accepted tools to analyze and compare alternative proposals. Further, by identifying the many components that could be negotiated and providing no means of comparison or analysis, the negotiation appeared to have become even more formidable.

Without a means of reducing the gulf between the assumptions of the members of the negotiating group, the financial negotiations were in danger of becoming a rigid and unproductive contest between developing and industrialized countries. It was in this setting that a new source of information on deep-ocean mining was introduced. A report of a research project at MIT that was released and distributed to conference participants (see [5]) provided a means of moving away from the adversary situation that had been developing and toward a more cooperative effort to seek agreement.

4 The MIT Model

The model of deep-ocean mining presented in the MIT report was simple in concept. It consisted of a technical description of the elements of a representative deep-ocean mining system, an analysis of the costs of building and operating the system, and an analysis of costs and revenues over time, including a measure of profitability.

The MIT model was a computer simulation of a hypothetical deep-ocean mining system designed to collect polymetallic nodules from the deep ocean floor and to process them to produce nickel, copper, and cobalt. The design was based on company proposals, patent descriptions, analogies to similar systems and equipment, and engineering analyses conducted by the MIT investigators. Conventional procedures for updating equipment prices, adapting estimates to different scales of operation and estimating the engineering fees, and contingencies and other costs were incorporated into the model. Four major technical sectors were described: preproduction activities of research, development, prospecting and exploration; mining; sea transportation; and metal processing. In turn, each sector was divided further into major systems and items of equipment. The estimates of preparatory, capital, and operating costs were aggregated and scheduled for use in the financial analysis section of the model.

The financial analysis section allocated the costs and expenses over the predevelopment, development and operational periods of the project. It also calculated the annual revenues based on assumptions of metal content of the nodules, metal recovery efficiency of the processing plant, and market prices for the recovered metals. Based on the annual expenses and revenues, the model calculated payments to the International Seabed Authority and corporate income taxes. The net cash flow for each year was calculated from revenues, capital investment, operating expenses, payments to the Authority, and federal tax payments.

The final step, and one of the most important from the user's perspective, was the calculation of the simulated project's overall profitability. Many measures are used by business analysts to measure the financial performance of business firms. Two of these measures were calculated by the MIT model: net present value (NPV) and internal rate of return (IROR). These measures are similar in that they both recognize the "time value of money." That is, the idea that a dollar of benefits received some time in the future is of less value than a dollar received today. In calculating either measure of profitability, future benefits are reduced by a factor known as the "discount rate," which is similar to an interest rate. For example, for a discount rate of 10 percent, the value of a dollar received one year from now would be equal to

$1/(1+0.1) or 91 cents. In calculating the NPV, a discount rate is specified and the sum of discounted benefits is calculated. The IROR calculation is somewhat similar, except that the computer makes repeated calculations of the discount rate in order to determine the discount rate at which the sum of discount costs and benefits are equal to zero. Although both measures were calculated by the model, the IROR was generally preferred because it did not require the arbitrary specification of a particular discount rate (which some observers feared could be taken to represent a fair, allowable, or sufficient profit).

The strengths of the MIT model were in the depth of its technical detail and its adaptable structure. Creating a model of an industry that had yet to build a full working prototype, let alone conduct full commercial operations, presented a challenging task. However, by drawing from the public descriptions provided by the ocean mining companies in the technical press and professional conferences, from patent descriptions, and from academic analyses of nodule composition, distribution, and other features, it was possible to break the system into components that were comparable to technology in use in offshore oil production, marine transportation, mineral processing, and other disciplines. Through the use of established cost-engineering techniques, costs of existing processes and equipment were scaled and modified to estimate the costs of the components of the ocean mining system. The financial analysis was a conventional discounted cash flow calculation, but it was structured to allow for changes in the scheduling of operations, of the national and international systems of taxation, and of the assumptions about the behavior of metal prices.

The final result of these complexities was a "black box." Into one end, assumptions as to the technical design, mineral resource distribution, metal prices, taxation, scheduling, and other factors were entered. Out of the other end came a measure of profitability, the internal rate of return (IROR).

4.1 Development of the model

The MIT model began with rather modest expectations. Growing out of an experimental course in Law and Engineering sponsored by MIT and Harvard Law School, it was further developed to support graduate research into the effects of regulatory constraints on the investment decisions of companies regarding deep-ocean mining. In the early stages of research it was found that the generalized cost estimates prepared by industry and by the US government were of little use in evaluating corporate investment decisions and of even less use in evaluating the effects of regulatory actions on those decisions. As a result, a significant amount of work was devoted to a detailed description of the technology of a deep-ocean mining operation, evaluation of the costs resulting from the particular choice of technologies chosen for development, and a financial analysis of the costs and potential revenues over time in order to produce a measure of profitability.

Initial research on the MIT model was undertaken without outside financial support. As a result, calculations were made without the use of a computer and took over a day to complete. This proved to be a serious limitation on the model's usefulness.

While the research was in progress, the National Oceanic and Atmospheric Administration (NOAA) took note of the model as a possible tool to assist it in environmental and economic assessments for domestic oversight of ocean mining. NOAA contracted MIT to expand the model and convert it to run on the MIT computer facility so that the effects of changes in the model's assumptions, both technical and policy-related, could be tested quickly.

A draft report on the MIT model, with a detailed description (including its structure and all of its variables) was delivered to NOAA in February of 1977. The model was circulated to government agencies and outside experts for comment. Encouraged by the initial round of comments, NOAA asked MIT to participate in a workshop in which industry experts were asked to critique in depth the technical and financial aspects of the model. In addition, the draft report was circulated to a growing circle of reviewers. With the results of the workshop and the outside reviews in hand, MIT modified the model to take the comments into account. In addition, the results of other NOAA contracts that evaluated the land-based processing plant and the nodule-transportation system were made available to MIT. The new information was carefully evaluated by MIT and, when it offered improvements over earlier assumptions, the model was modified.

In addition to the critiques of the model's assumptions and structure, NOAA sponsored several presentations of the MIT model to representatives of the Departments of State, Commerce, Interior, Defense, and the Treasury and to several Congressional committees to identify areas that could serve as topics for timely application of the model to public policy issues. The model was evaluated for its sensitivity to changes in technical and political assumptions and in March 1978, was published and released to the public.

4.2 Use of the MIT model in the financial negotiations

The MIT model was originally formulated to be a planning aid to the US government in the development of regulations for the management of US ocean-mining companies. Its principal purpose was to identify the regulatory actions that might significantly reduce the attractiveness of the investment from the corporate standpoint and to compare those effects with those resulting from alternative types of regulation. However, in mid-1977 a draft report on the model was sent to a Commerce Department economist who served on the delegation to the Conference. Deeply involved in the financial negotiations at UNCLOS, he recognized the value of the model as a tool for illustrating the details of complicated systems of taxation and comparing the effects of alternative proposals. After the model was brought to the attention of the leader of the US delegation, Ambassador Elliot Richardson, it was used to analyze the effect of variations of royalty rates on project profitability. The ties between MIT and the UNCLOS delegation became closer when the economist left the Commerce Department to attend Harvard where he became a consultant to the MIT group. Soon afterward, one of the MIT staff moved to the Commerce Department and joined the US delegation. This intermingling of staff ensured that the MIT model was kept in tune with the needs of the government, particularly those of UNCLOS delegation.

5 The MIT Model as a Negotiating Tool

The role of the MIT model in UNCLOS evolved as the negotiations moved through three phases: diagnosis and understanding of the issues, development of formulas of possible agreements, and definition of the details of the final formulation.

5.1 Diagnosis: Improving the understanding of ocean mining

In the late summer of 1978 the role of the MIT model began to change from a tool of the US government to a servant of UNCLOS as a whole. As developed and developing countries made incompatible proposals for the financial terms for exploitation, Ambassador Koh sought a means to move the negotiations forward. The MIT model appeared to offer an opportunity to give the push he needed. At Ambassador Richardson's request, NOAA made the model available for use at UNCLOS.

Before the model could be put to use it had to be accepted by the Conference participants. A workshop was held for this purpose at which the MIT team described the model in detail, illustrated its use, and responded to questions and comments. To avoid the appearance of premature acceptance, the workshop was not held at the UN but was conducted under the auspices of the Neptune Group (a group of Church organizations that facilitated the negotiations) (see [3]). Key representatives of developed and developing countries and of the ocean-mining industry were invited. After a full day of detailed presentations by the MIT staff, including a long question-and-answer period, most participants were favorably impressed by the MIT model and its developers. There was some criticism voiced by representatives of the European ocean-mining industry, in part reflecting concern about an independent source of information about their operations. However, the industry comments and the responses by the MIT team helped convince the developing country leaders that, in spite of its US origins, the MIT model would not be biased against their interests.

The workshop also provided the first opportunity for the application of the model to proposals at the Conference when the Norwegian representative, Ambassador Jens Evensen, asked the MIT team to evaluate his proposed financial provisions. Although not planned as part of the workshop agenda, the team had already tested the Evensen proposal. Upon hearing the results of the analysis, Ambassador Evensen thanked MIT for the work and indicated that he would use the results in reformulating his proposal. Acceptance by Ambassador Evensen, who was highly respected by developing country representatives, for his serious efforts to reach a compromise, sealed the approval of the MIT model for use as a tool in the negotiations.

5.2 Formula building: The model as an analytic tool in the negotiations

Participants in the negotiations had different expectations about the profitability of deep-ocean mining as well as different objectives for the financial provisions. Developing states had confidence in the high profitability scenarios, while the ocean-mining countries placed their emphasis on the cases of marginal profitability. The interests of developing countries lay in ensuring high cumulative payment from an ocean-mining project, without regard to the timing of the payments, and the industrialized countries worried more about the profitability of the operations, which put emphasis on maintaining high income during the earlier years of the project. Their greatest concern was for low profitability scenarios, where financial arrangements might cause companies to decide against going forward with mining operations.

The model was adapted to test formulas that would exploit the differences in expectations and objectives. First, a variety of scenarios were developed to cover the range of projections of future profitability (see *Table 1*). Second, a subroutine was incorporated to model a general system of taxation that included the components identified earlier in the discussions in NG 2, including profit sharing and royalties, application fees and minimum payments, depreciation allowance for capital goods, and other financial factors. This subroutine allowed the model to analyze complicated systems that included progressive taxation based on an annual return on investment attributable to the mining operation alone (separate from the sea transportation and land-based processing components), and taxation rates that would change as a result of time, gross revenues, or profitability. A compact version of the model that incorporated these features was developed for use on an Apple II computer (state of the art for personal computers at that time) that was maintained by members of the MIT team serving on the US delegation. The results of analyses conducted on the compact version were made available to sponsors of proposals submitted to NG 2.

Table 1. Economic scenarios for analyses of deep ocean mining (in million US dollars, 1978 value)

	Variables Used in Test Scenarios				
Scenario	Research and development	Capital cost	Operating cost	Initial revenue	Annual price inflation
A	150	604.9	123.00	226.5	0.0%
B	150	604.9	123.00	226.5	1.0%
C	50	493.1	100.35	258.2	0.0%
D	50	616.3	138.75	348.5	0.0%
E	50	616.3	138.75	348.5	2.5%
F	50	493.1	100.55	348.5	2.5%

All scenarios are based on a modified baseline case of the 1978 MIT model. The modifications are that the project is 100 percent equity financed and the operating life is extended form 20 to 25 years. Initial revenues are based on metal prices during the first year of the project, with prices increasing as specified by the annual price inflation rate.

During the 1977–1979 period there were more than a dozen significant proposals on the financial payments for seabed mining. The progress of these proposals toward a compromise is illustrated in *Figure 1*. As can be seen, the initial positions were far apart in 1977. In early 1978, partly because of the growing understanding of the economics of ocean mining that resulted from cooperative work between the MIT team and the US delegation, the industrialized countries began to move away from their extreme positions. By late 1978, through the use of the model, ambassadors Evensen and Koh were each able to develop innovative proposals whose effect on profitability was much lower than earlier proposals made by India and whose total contributions to the Seabed Authority were greater than the proposals by the industrialized countries. Developing country resistance to these proposals began to diminish as their delegates had an opportunity to study the MIT report and to talk with the authors.

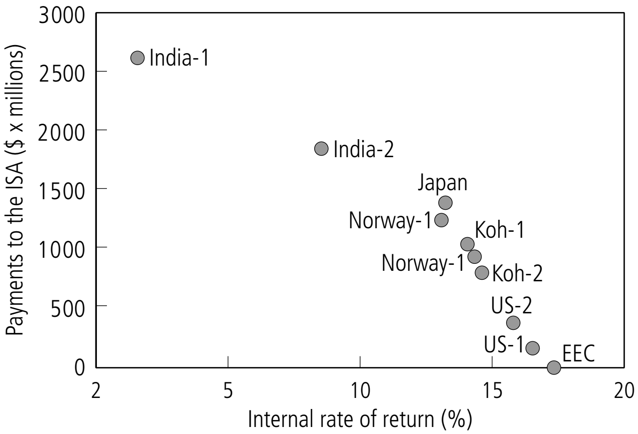

Fig. 1. Financial proposals in negotiation space
Note: Compromise proposals by Norway, Japan, and Koh filled the gap between the proposals by India and the United States (USA) and EEC, but failed to reach the minimally acceptable levels for those two groups and thus were not successful as compromise agreements. Additional effort was needed to develop new formulas for agreement that would produce results to the "northeast" of the Koh–Norway–Japan proposals.

The model was put to use as an analytic tool in support of formula building during the resumed session in 1979 of the Conference in New York. The United States presented a detailed analysis of the financial provisions developed at the preceding session of the Conference. Using the MIT model to examine 14 scenarios based on a variety of assumptions of costs, revenues, and other factors, the analysis illustrated that the provisions put a relatively higher financial burden on marginally profitable mining operations than on highly profitable cases. Based on this analysis, the USA suggested that the system be modified to be more progressive by reducing the burden on low-profit cases while increasing the payments required in high-profit cases.

The compact model was also used to conduct an analysis of the Norwegian proposal made during the August 1979 New York session (see [9]). This analysis was distributed to conference delegates by the members of the MIT team in their personal capacity rather than as members of the US delegation, and it refrained from an interpretation of the results. However, the analysis clearly showed the proposal to be more progressive than the earlier Geneva proposal, thereby meeting, at least in part, the US recommendations.

5.3 Specifying the details

Discussions in NG 2 then moved to the detail phase of negotiations to address the specifics of what would become the final compromise. Koh's expert staff evaluated scores of possible compromise formulations on the compact model. When analyses of the model narrowed the selection down to several possible formulations, MIT was asked to conduct the final tests on the formulas of the full model on the MIT campus for a group of six scenarios selected to illustrate the effects of the proposals on cases of interest to both industrialized and developing countries (see [10]). Ambassador Koh and his group of experts used the results of these analyses to select their recommended compromise. The MIT results were also included in the report of the Chairman of Committee I to explain the effects of the proposal to the Conference participants.

Koh's compromise was the culmination of a process that resulted in increasingly complicated proposals designed to resolve specific problems identified by various countries. The structure of the proposal included an initial application fee, royalties on gross sales of metals from nodules, a progressive profit-sharing system whose marginal payment increased for higher profit levels on mining (but not on sea transportation or processing), a two-stage schedule of rates for both royalties and profit sharing that was triggered by the recovery of investment with interest, and minimum rates for annual payments and for the attribution of profit to the mining operations. The details of this proposal are shown in *Table 2*.

Table 2. Structure of the mixed system of financial arrangement

Application fee	\$500,000			
Minimum annual payment	\$1,000,000			
Royalties and revenue sharing based on two-period structure				
	Royalty rates Total net revenues	Revenue-sharing rates (Mining income only)		
		Step 1	Step 2	Step 3
First period	2%	35%	42.5%	50%
Second period	4%	40%	50%	70%
Switchover to second period:	Recovery of 110% of investment (constant dollars)			
Return on investment:	Annual net revenues divided by development costs			
Income subject to sharing:	Gross revenues less operating costs, depreciation, royalties, and fees to the Authority			
Depreciation:	10-year period, straight-line method			
Income attributable to mining:	Ratio of mining costs to total costs, with minimum ratio limited to 25%			

Source: Third United Nations Conference on the Law of the Sea, Report of Negotiations Held by the Chairman and the Coordinators of the Working Group of 21, (A/CONF.62/C.1/L/26, 21 August 1979)

Koh's proposal was presented as a compromise between the proposal of Ambassador Evensen and a combined USA–European Economic Community (EEC)–Japan proposal. The MIT model was used to assess the impact of the proposal on profitability and cumulative payments to the Authority (see *Table 3*).

Table 3. Results of analyses of the final compromise proposal

Scenario	Internal rate of return		Cumulative payments to the Seabed Authority (millions of dollars)
	Without payments	With payments	
A	7.0%	6.1%	258
B	9.8%	8.6%	429
C	15.4%	13.8%	574
D	21.7%	19.5%	1,015
E	22.3%	20.2%	1,792
F	26.6%	23.9%	1,964

Source: Third United Nations Conference on the Law of the Sea, Report of Negotiations Held by the Chairman and the Coordinators of the Working Group of 21 (A/CONF.62/C.1/L/26, 21 August 1979)

A key factor in the identification of a mutually acceptable compromise was the recognition that the two major groups differed both in their measure of self-interest and in their consideration of risk. Developing countries were concerned with the sharing of financial benefits of exploitation with an emphasis on the principle of sharing of benefits over the specific timing of payments. They were also influenced by their belief that seabed mining would be an extremely profitable activity and so were most concerned with the terms for sharing benefits from high-profit scenarios. Industrialized countries were concerned with the impact of revenue sharing on profitability and, while they hoped such operations would be profitable, they were most concerned that initial operations that might operate at lower profit levels not be taxed out of business. These considerations went into the development of Koh's final compromise proposal, which is presented in *Figure 2*, relative to the competing proposals of Norway and the industrialized country proposal of the USA, EEC, and Japan. In this comparison the interests of the developing countries are represented by the average cumulative payment of the three high-profit scenarios while the industrialized country interest is presented as the average IROR of the three least profitable of the six scenarios. Koh's final proposal was intentionally prepared to split the difference between the final Norwegian and USA–EEC–Japan proposal in both interests for the scenarios of greatest concern to each group.

Koh's final compromise proposal culminated two years of effort to achieve the most efficient compromise possible, one that made maximum use of differences in values and risk assessment as well as creative development of a formula and values for revenue sharing. When it proved to fall just short of developing country needs, Koh was able to link the formula with separate discussions on the financing of the Enterprise, the mining venture to be operated by the Authority. Having demonstrated that the financial terms of mining contracts had achieved their best possible outcome, Koh was able to convince the industrialized states to make a final concession

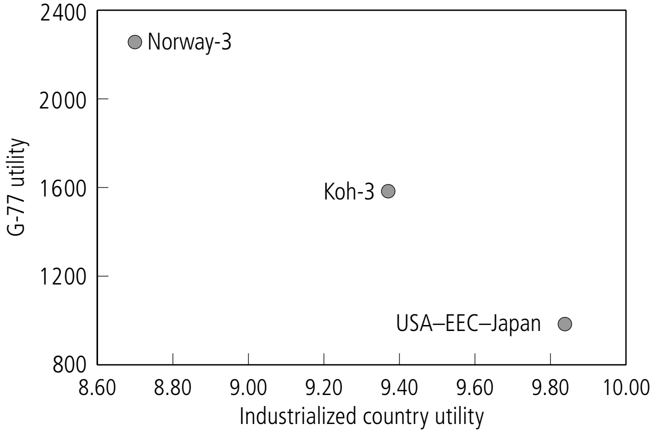

Fig. 2. Final financial proposals in modified utility space
Note: G77 utility is the average of the payments to the authority calculated under the three high-return cases. Industrialized country utility is the average of the three low-return cases. The use of different cases reflects the different groups assessment of the scenarios for which they were most concerned.

to increase the funds to be provided to the Enterprise in the form of zero-interest loans. *Figure 3* illustrates the effect of linking the financing of the Enterprise to the formula for the financial terms of contracts. The linkage created a new formula that included a solution that met or exceeded the minimum utility for both developing and industrialized countries (expressed as cumulative payments for developing countries and impact on profitability for industrialized countries).

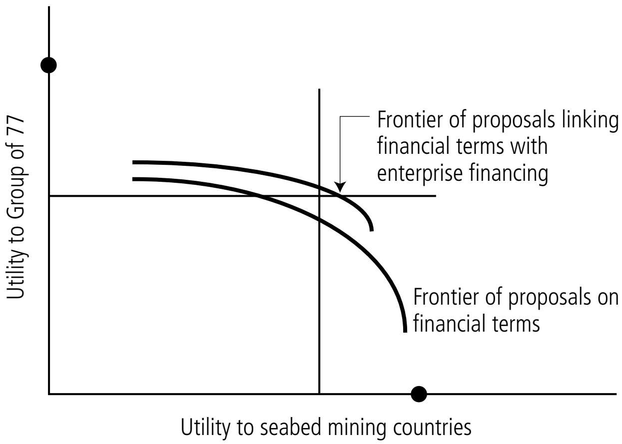

Fig. 3. Effect of issue linkage on prospects for agreement

Although the negotiations on the Law of the Sea Convention continued for nearly three more years, the financial provisions negotiated in NG 2 remained unchanged after the close of the 1979 session. Perhaps the greatest tribute to the success of the negotiations in this area was that when the United States conducted its year-long review of the Convention for the incoming Reagan administration in 1981–1982, which resulted in a lengthy list of desirable changes, the financial provisions were not brought up as a subject for change, nor, in the contentious atmosphere that resulted from the change in the US position, did the developing countries seek to reopen them.

6 Conclusions

6.1 Factors affecting the use of the MIT model in UNCLOS negotiations

The MIT model played an important role as a facilitator in the negotiations on the financial arrangements for deep-ocean mining within the larger context of the draft Law of the Sea Convention. First, the model assisted in providing informed data about the economics of deep-ocean mining when the negotiators returned to a diagnosis phase after earlier attempts at formula building failed to reach agreement. Second, through the identification of scenarios of interest to the parties, a formula was developed that encompassed the elements of a taxation regime. Finally, the model supported the development, analysis and comparison of detailed proposals, and the final compromise agreement.

Which factors put the MIT model in such an influential position? Several other economic models and analyses were available to UNCLOS, so which factors led to the adoption of the MIT model? In part, the selection was a chance opportunity. The growth of interest by the MIT faculty and students in deep-ocean mining, NOAA's decision to finance MIT research, and the recognition of the potential value of the model by members of the UNCLOS delegation were unplanned occurrences that opened the way for the model's introduction. Once discovered, however, the following five factors contributed to the success of the model:

1. *Applicability*: The model addressed a topic, the economics of deep-ocean mining that was crucial to the negotiations. The model was detailed, installed on a computer (which facilitated its use), and incorporated extensive economic data relevant to the examination of ocean-mining operations. Most important, prior to the introduction of the MIT model the conference participants had no detailed and commonly accepted source of economic information on which to base their assumptions, so the model addressed a critical information gap.

2. *Veracity*: The model was accompanied by a report that explained its assumptions, structure, and results in great detail. The model's structure utilized techniques accepted by cost engineers and financial analysts for other major engineering projects, and the data for most of the estimates were drawn from the known costs of existing equipment and systems. Both the structure and the data were subjected to extensive technical review before the model was released to the public. This review process, the reputation of the Massachusetts Institute of Technology, and the

detailed presentation and defense of the model given by the MIT team at the seminar sponsored by the Neptune Group combined to provide the Conference participants with the assurance that the model was technically correct and without ideological biases that might work against their interests. The trust in the model was maintained throughout the negotiations by the willingness and ability of the MIT team to answer the many questions raised by all parties in the negotiations.

3. *Adaptability*: The technical and economic assumptions of the model were clearly specified and could be modified to reflect different expectations about costs, efficiencies, resource endowment, and metal markets. The model could be adapted to test a variety of scenarios of interest at the Conference. Likewise, the structure of the model could be adapted to include new features such as the calculation of payments to the International Seabed Authority, the effect of the payments on the profitability of the mining operation, and the cumulative value of payments to the Authority over the life of the project. Direct access to the model operators at the Conference helped to ensure that the modifications to the model reflected the intent of the negotiators.

4. *Accessibility*: Use of the model was not restricted to any party or group of parties. The report describing the structure and assumptions of the model was distributed throughout the Conference. The MIT team members on the US delegation were available to all participants to answer questions about the model. Analyses of possible proposals based on the model were available to all participants. The MIT team, both in Cambridge, Massachusetts and at the Conference, acted as assistants to the negotiators and were available to all parties as consultants to modify the programming of the model to incorporate the legal and economic intent of the proposals' sponsors. The availability of the compact model in New York provided a speedy response to requests for an evaluation of proposals involving multiple variables. The compact model also made it possible to conduct analyses in direct consultation with the negotiators, unhindered by the lack of face-to-face contact involved in using the main computer model at MIT.

5. *Integrity*: The MIT team was trusted by all parties in the negotiations to maintain the quality of the model and to protect it from misuse. This was particularly important because few delegates actually saw the code and data of the model, while time constraints of the negotiations precluded public review of each modification to the model.

6.2 Process and substance: The model as a mediation tool

While the MIT model would not have been used if it had not provided much-needed technical and economic information, it is also true that it would not have had such a substantial role without Ambassador Richardson's decision to make the model available to Ambassador Koh and Koh's development of the negotiation process, first to educate the delegates and then to involve them in the creation of increasingly complex formulas for agreement. Koh's process for moving discussions back and forth between the full negotiation group, the smaller Group of Financial Experts and the exclusive group of technical experts that assisted Koh in developing proposals was the framework within which the model could maximize its impact.

6.3 Final considerations for modelers and mediators

The substantial contribution of the MIT model to the negotiations at UNCLOS suggests that there may be roles for analytical models in other negotiations and disputes. The experience also indicates, however, that such models will not find easy acceptance.

Two points should be considered in evaluating the potential for computer models to assist in the conduct of complex negotiations. On the positive side, a model may reduce the technical and economic uncertainties that would otherwise lead negotiators to make extreme demands in order to protect their nation's interests. The importance of the MIT model in this role was noted by a former US negotiator who said:

> What broke the log jam in this particular case was a mutual reeducation program which provided a neutral basis for negotiations. The MIT study clearly showed that some of the economic fears of the developed countries were overblown and that some of the political theories of the developing countries would not lead to economic vitality (see [1]).

However, there is concern that a model might go beyond simple evaluation of an issue in the selection of an optimum solution, based on criteria known only to the model builders. Reflecting the concern that models and model builders were increasing in importance at the expense of the negotiators, Ambassador Paul Engo of Cameroon, Chairman of Committee I on Seabed Matters, commented during the seventh session of UNCLOS in 1978:

> [We] have ourselves been dragged into adopting models and systems of calculation on fictitious data that no one, expert or magician, can make the basis of any rational determination We get more and more engrossed with each session and have been reduced to mere spectators in the inconclusive tournament among experts (see [7]).

If a model is to be useful in the resolution of a conflict, it will be necessary to keep in mind that decisions to compromise or to trade among interests are the responsibility of the representatives of the parties to a negotiation, not of the model builders. Establishment of a procedure for technical verification and political review similar to that followed by the MIT team may help models gain acceptance by those representatives. Just as important, however, is the building of trust in the builders and operators of the model to protect their creation from misuse. Ultimately, it is the confidence of the negotiators in the integrity of the model builders and operators that will determine the degree to which the model is relied upon in the resolution of dispute.

References

1. Katz R (1979) Financial arrangements for seabed mining companies. Journal of World Trade Law 13:209–222 (May/June)

2. Leitner PM (1998) A bad treaty returns. World Affairs (Winter)
3. Levering R, Levering M (1999) Citizen action for global change. Syracuse University Press, New York, USA
4. Newman B (1974) The great ocean mining race $200,000,000,000 beneath the sea. The Saturday Evening Post, March. Reprinted in: Mineral resources of the deep seabed, Hearings before the Subcommittee on Minerals, Materials and Fuels of the Committee on Interior and Insular Affairs, United States Senate, 93rd Congress, 2nd Session on Amendment No. 946 to S.1134, March 5,6,7, 1328–31
5. Nyhart JD, Sebenius JK, Richardson EL, Koh T (1983) The development and use of MIT deep ocean mining model in law of the sea negotiations: critique of development and use of model for the negotiations. Presentation and panel discussion at the New York Academy of Science
6. Nyhart JD, Antrim L, Capstaff A, Kohler A, Leshaw D (1978) A cost model of deep ocean mining and associated regulatory issues. MIT Sea Grant Report 78–4
7. Richardson EL (1980) A constitution for the oceans. Time Magazine, September 8. Available in The Time Magazine electronic archives at http://www.time.com/time/archive/preview/0,10987.924405,00.html
8. Sebenius JK (1981) The computer as mediator: law of the sea and beyond. Journal of Policy Analysis and Management, vol 1, 77–95
9. Sebenius JK, Antrim L (1979) Analysis of Norwegian proposal on financial arrangements. Informal Paper, Third United Nations Conference on the Law of the Sea, August. Copy available from author
10. US delegation to the Third United Nations Conference on the Law of the Sea (1979) Expanded analysis of "mixed system" of financial arrangements. August. Available from author

Part IV

Models for International Negotiations

Somehow these models stand between those *of* and those *in* negotiations: On the one hand, and contrary to the models *of* negotiations, they provide very special advice on solving a concrete negotiation problem; on the other hand, they do not take into account the more subtle procedural or structural aspects. A fair division solution, for example, if accepted by the parties, eliminates any kind of negotiation. Thus, in reality, models of this type may serve as guidelines at the beginning of or at some point during the negotiations without claiming that they can solve the negotiation problem immediately.

The chapters in this part illustrate and discuss the role of this type of formal model with the help of real cases. Kilgour's graph modeling and Okada's analysis of the Kyoto Protocol and its implications, as well as the other contributions, in particular, Brams, Kilgour and Sanver's, Raith's, and Schüssler's discussion of the use of the Adjusted Winner Solution, demonstrates the merits of this type of model for practitioners of international negotiations.

International Negotiations on Climate Change: A Noncooperative Game Analysis of the Kyoto Protocol*

Akira Okada

Graduate School of Economics, Hitotsubashi University, Tokyo, Japan

1 Introduction

The purpose of this chapter is to consider international negotiations on climate change in a game-theoretic framework. With the increase of environmental concerns in the late 1980s, the United Nations Framework Convention on Climate Change (UNFCCC) was signed at the Rio de Janeiro Earth Summit in June 1992. The objective of the Convention was to stabilize greenhouse gas (GHG) concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system. Developed countries promised to reduce their emissions of CO_2 and other greenhouse gases (GHGs) to their 1990 levels by the year 2000. The Convention, however, lacked any legally binding commitments, and this voluntary approach was not successful. The third Conference of the Parties (COP3) to the Convention was held at Kyoto from 1–10 December 1997. The objective of the Kyoto conference was to adopt a "legally-binding protocol or other legal instrument"[2] committing developed countries to reducing their GHG emissions.

More than 150 developed and developing countries attended the Kyoto conference. Intensive negotiations took place and the Kyoto Protocol to the Convention was finally agreed.[3] The key contents of the Protocol are summarized as follows. First, Annex I countries (Organisation for Economic Co-operation and Development [OECD] countries and countries in the former Soviet Union [FSU] and Eastern Europe) as a whole reduce emissions by 5.2 percent below 1990 levels between 2008 and 2012 (Article 3). Second, the quantified emissions limitation or reduction commitment (QELRC) is assigned to each Annex I country (Article 4 and Annex B).

* I would like to thank Jiro Akita and the participants in the workshop at the International Institute for Applied Systems Analysis (IIASA) on Formal Models of Negotiations for their helpful comments. Research support from the Sumitomo Foundation is gratefully acknowledged.

2 Ministerial Declaration at COP2, 8-19 July 1996. http://unfccc.int/sessions/cop2/l17.pdf

3 The agreement was reached late in the morning of 11 December, 36 hours after the official deadline. For a detailed survey of the Kyoto Protocol, see [7]. The full text of the Protocol is available at http://unfccc.int/resource/docs/convkp/kpeng.pdf

The reduction rates of major emitting countries are as follows: Russia and Ukraine 0 percent; Japan 6 percent; the United States (USA) 7 percent; and the European Union (EU) 8 percent; Third, the Protocol includes three "flexible" mechanisms for international emissions transfer, such as joint implementation (Article 6), the clean development mechanism (Article 12), and emissions trading (Article 17).

In this chapter, we present a game-theoretic model of international negotiations on the Kyoto Protocol. Our analysis is focused on reduction commitments of CO_2 emissions and emissions trading. The bargaining model has two consecutive stages, (i) allocating emission reductions to countries, and (ii) international emissions trading. We assume that emissions trading takes place in a competitive market and that, given a total emissions target, efficient reductions (minimizing the total reduction costs) can be attained across the countries in the emissions market. This enables us to solve the two-stage model by backward induction. Given the competitive equilibrium of emissions trading, the whole bargaining model of allocating emissions can be reduced to the n-person unanimous bargaining problem where an agreement requires the unanimous consent of countries. The countries evaluate the agreement of emissions allocation by their net costs after trading in the competitive emissions market.

We formulate a negotiation process on emission reductions by noncooperative multilateral bargaining theory (see [20], [5], [10], [14], among others).[4] Specifically, our model is based on the sequential bargaining game with random proposers introduced by Moulin [11] and Okada [14], which is a generalization of Rubinstein's [18] two-person alternating-offers model.

The negotiation process consists of a (possibly) infinite sequence of bargaining rounds. At the beginning of each round, every country is randomly selected as a proposer according to a predetermined probability distribution. If selected, a country proposes an allocation of emissions. All other countries either accept or reject the proposal sequentially. If all responders accept the proposal, then the proposal is agreed upon. If any country rejects the proposal, then negotiations can proceed to the next round with some positive probability, and the same process is repeated. Negotiations, however, may stop with the remaining probability. The *disagreement point* of negotiations describes countries' costs in the event of failure of negotiations.

We show that there is a unique stationary subgame perfect equilibrium (SSPE) of the negotiation process and, moreover, that the equilibrium proposal by every country converges to the asymmetric Nash bargaining solution as the stopping probability of negotiations goes to zero. The weights of countries in the asymmetric Nash bargaining solution are determined by their probabilities of being selected as proposers.

In the second part of the chapter we present numerical results based on actual emissions data for the European Union (EU), the FSU, Japan, and the United States. Using the marginal costs of CO_2 emissions abatement estimated by Nordhaus [13], we compute the competitive equilibrium price ($9.45 per carbon ton) in emissions trading among the four countries. Based on the numerical results we discuss how

[4] These works treat multilateral bargaining situations where coalitions of players are allowed to form. The model in this chapter can be extended to such a general situation (see [14]).

the reduction commitments by the Kyoto Protocol can be supported by the asymmetric Nash bargaining solution under three different kinds of weight: equal weight, population weight, and gross domestic product (GDP) weight.

The actual negotiations in the Kyoto conference involve many unknown factors which are not easily analyzed by any formal model. The Kyoto Protocol may be best understood as the outcome of an unseen political process compromising various conflicting objectives. Nevertheless, in our view, it is important to consider negotiations on the Kyoto Protocol by the formal methodology of game theory. A game-theoretical model clarifies the strategic structure of international negotiations on the Protocol and enables us to examine carefully the kind of rational basis on which the Protocol is established. It helps us to understand a complicated negotiation process on climate change from the strategic viewpoint of countries and gives us quantitative predictions on emissions trading.

As far as we know there are few game-theoretical works on the Kyoto Protocol. In our previous work [16], we analyzed negotiations on the Protocol by cooperative game theory. The negotiations on reduction commitments are formulated as a cooperative voting game. Several allocation rules (including the Kyoto Protocol) are evaluated from the viewpoints of three major emitting countries, the FSU, Japan, and the USA, based on the same empirical data as in this chapter. The cooperative game approach has a problem in that the stability is restrictive: the set of stable outcomes (defined by the *core* solution) of the voting game is nonempty if, and only if, there is at least one country possessing veto power. Unlike in the UN Security Council, however, all negotiating players are symmetric regarding the veto power in international negotiations on climate change. For this reason, the unanimous voting rule, where every country has veto power, is the most relevant for international negotiations on climate change. Then, the result of the cooperative game is vacant in that *all* allocations of emissions can be stable. The noncooperative game approach in this chapter overcomes this difficulty. It can select the (asymmetric) Nash [12] bargaining solution as an unique SSPE of the bargaining process. Noncooperative behavior is critical in international negotiations with countries without any centralized power. Barrett [3] analyzes "acceptable"[5] allocation schemes of emissions among China, the FSU, and the USA, and demonstrates by empirical data that the uniform percentage abatement rule is preferred by these countries, without an analysis of bargaining procedure. Tadenuma [22] presents a three-stage game model of the Kyoto Protocol in an economy with production and consumption causing emissions of greenhouse gases. Complementary to our model, his bargaining model presumes that countries first negotiate on the total amount of emissions permits, given a distribution rule for the permits. Tadenuma [22] examines whether or not the bargaining frontier (including the Nash bargaining solution) is Pareto-optimal in the economy.

The chapter is organized as follows. Part 2 presents a game-theoretic model of international negotiations on the Kyoto Protocol. The properties of a competitive

[5] In game-theoretic terminology, the acceptability here means the individual rationality that each country's net benefit is at least as great under an allocation as it would be if negotiations failed.

equilibrium of emissions trading are reviewed. Part 3 presents a noncooperative bargaining procedure on emission reduction. The main theorem is proved. Part 4 presents numerical results on the competitive equilibrium of emissions trading among the EU, the FSU, Japan, and the USA. Part 5 concludes the chapter.

2 The Model

Let $N = \{1, \cdots, n\}$ be the set of countries. For every $i \in N$, we denote by E_i country i's current level of CO_2 emissions. The total level of CO_2 emitted by n countries is given by $E = \sum_{i \in N} E_i$. Let x_i denote country i's reduction of CO_2 emissions where $0 \leq x_i \leq E_i$. The CO_2 abatement cost function of country i is denoted by $C_i(x_i)$. It is assumed that $C_i(x_i)$ is a differentiable, strictly convex, and monotonically increasing function on R_+, the set of all nonnegative real numbers. Let $MC_i(x_i)$ denote the marginal abatement cost function of country i. For a natural number $s = 1, 2, \cdots$, notation R^s_+ means the nonnegative orthant of the s-dimensional Euclidean space R^s.

Our game-theoretic model of international negotiations on climate change has the following two consecutive phases: (i) negotiations on emissions allocations and (ii) emissions trading.

In the first phase of negotiations, n countries negotiate on an allocation of emissions. Let ω_i denote an amount of emissions permits allocated to country i. Country i has to reduce $E_i - \omega_i$ amount of emissions, if emissions trading is impossible. An *emissions allocation* is formulated by a vector $\omega = (\omega_1, \cdots, \omega_n) \in R^n_+$. The total amount of CO_2 emissions permits is given by $\overline{\omega} = \sum_{i=1}^n \omega_i$. In this chapter we assume that the total emissions target $\overline{\omega}$, $0 < \overline{\omega} < E$, is exogenously determined by scientific knowledge and that it is not an issue of negotiations. The agreement of an emissions allocation $\omega = (\omega_1, \cdots, \omega_n)$ satisfying $\overline{\omega} = \sum_{i=1}^n \omega_i$ is reached as the unanimous consent among n countries. The negotiation process will be described in Part 3.

In the second phase of negotiations, n countries negotiate to sell and buy their emissions permits, given the initial emissions allocation $\omega = (\omega_1, \cdots, \omega_n)$ agreed in the first phase of negotiations. We assume that the emissions trading takes place in a competitive market. The competitive equilibrium of emissions trading is efficient in the sense that the total reduction costs across n countries attaining the emissions target $\overline{\omega}$ is minimized.

Let p be a market price of emissions and let x_i be an actual level of emission reduction by country i after the trading.

Definition 2.1. A *competitive equilibrium* of international CO_2 emissions trading with an initial emissions allocation $\omega = (\omega_1, \cdots, \omega_n) \in R^n_+$ is defined to be a vector $(p^*, x_1^*, \cdots, x_n^*) \in R^{n+1}_+$ satisfying

$$x_i^* \in \text{argmin}\, \{C_i(x_i) + p^*(E_i - x_i - \omega_i) \mid 0 \leq x_i \leq E_i\} \quad \text{for any } \; i \in N \quad (1)$$

$$\sum_{i \in N} (E_i - x_i^*) = \bar{\omega} \quad \text{where} \quad \bar{\omega} = \sum_{i \in N} \omega_i. \quad (2)$$

The minimum value of the optimization problem (1),

$$c_i^e \equiv C_i(x_i^*) + p^*(E_i - x_i^* - \omega_i) \tag{3}$$

is called the *competitive equilibrium reduction cost* for country i.

The definition can be explained as follows. If country i wants to reduce CO_2 emissions by x_i from E_i, it must possess $E_i - x_i$ amounts of emissions permits. If the initial allocation ω_i of emissions is less than this level, country i has to purchase $E_i - x_i - \omega_i$ amounts of emissions permits from other countries. Equation (1) means that every country i minimizes its CO_2 abatement net costs, given the equilibrium price p^* of emissions. Equation (2) is the balanced condition of demand and supply for emissions permits.

We review the standard properties of the competitive equilibrium. If x_i^* is an interior solution of country i's cost-minimizing problem (1), then the well-known principle of *marginal cost pricing* must hold:

$$p^* = MC_i(x_i^*) \quad \text{for all } i \in N \tag{4}$$

where MC_i is the marginal reduction cost function of country i. It can be proved that the equilibrium emission reduction $x^* = (x_1^*, \cdots, x_n^*)$ minimizes the total reduction costs for n countries to attain the reduction target $\bar{\omega}$. That is, x^* is the optimal solution of

$$\min_{x \in R_+^n} \sum_{i \in N} C_i(x_i) \qquad s.t. \qquad \sum_{i \in N} x_i = \sum_{i \in N} (E_i - \omega_i). \tag{5}$$

We denote by $c(\bar{\omega})$ the minimum value of the total emission reduction costs in (5) given the emission target $\bar{\omega}$.

Since a competitive equilibrium $(p^*, x_1^*, \cdots, x_n^*) \in R_+^{n+1}$ of the emissions trading can be characterized as a solution of the balanced equation (2) of demand and supply and the marginal cost pricing equation (4), it is important to note that the competitive equilibrium $(p^*, x_1^*, \cdots, x_n^*)$ is determined solely by the total reduction $r = E - \bar{\omega}$, given the marginal abatement cost functions of n countries. Further, if the total reduction level r is fixed throughout the analysis, we represent the competitive equilibrium reduction cost c_i^e for country i in (3) simply as a function $c_i^e(\omega_i)$ only of the initial emissions ω_i assigned to country i, independent of those assigned to other countries. The function c_i^e is called the *competitive equilibrium reduction cost function of country* i.

We summarize the properties of a competitive equilibrium of international CO_2 emissions trading in the following proposition.

Proposition 1. In the competitive equilibrium of international CO_2 emissions trading with an initial emissions allocation $\omega = (\omega_1, \cdots, \omega_n) \in R_+^n$, the emissions price p^* and the emission reduction vector $x^* = (x_1^*, \cdots, x_n^*) \in R_+^n$ are solely determined by the total emissions target $\bar{\omega} = \sum_{i \in N} \omega_i$. Given $\bar{\omega}$, the equilibrium reduction $x^* = (x_1^*, \cdots, x_n^*)$ is efficient; thus, it can minimize the total reduction costs for n countries to attain the emissions target $\bar{\omega}$. The equilibrium reduction cost $c_i^e(\omega_i)$ of every country i is a decreasing function of the initial emission ω_i allocated to country i, and it satisfies

$$\sum_{i \in N} c_i^e(\omega_i) = c(\bar{\omega}), \tag{6}$$

where $c(\bar{\omega})$ is the total equilibrium reduction costs defined by the optimal value of (5).

When countries negotiate on emissions allocations, they can anticipate the competitive equilibrium of international emissions trading. Thus, it is reasonable to assume that countries evaluate every emissions allocation by their equilibrium reduction costs after emissions trading, given the allocation. Equation (6) shows that the bargaining problem on emissions allocations can be considered as a cost-allocation problem where every country wants more emissions permits to decrease its reduction costs. A dynamic model of the negotiation process on emissions allocations will be described in the next part.

Remark. In this chapter, we investigate the outcome of emissions trading by applying the competitive equilibrium in price theory. An alternative approach is to apply some solution concept (the Shapley value, for example) in cooperative game theory. To do this, an n-person game in coalitional form can be constructed as follows. Every subset S of N is called a *coalition* of countries. All member countries in coalition S jointly minimize their total costs of emission reduction by reallocating emissions permits within the coalition. The total cost of reducing emissions for coalition S is given by:

$$\min_{x \in R_+^S} \sum_{i \in S} C_i(x_i) \quad s.t. \quad \sum_{i \in S} x_i \geq \sum_{i \in S} (E_i - \omega_i), \;\; 0 \leq x_i \leq E_i, \;\; \text{for any i} \in S,$$

assuming $\sum_{i \in S}(E_i - \omega_i) \geq 0$. The first constraint means that coalition S as a whole should not emit more CO_2 than the sum of emissions permits assigned to all member countries. Let $C^\omega(S)$ denote the minimum cost above. It can be easily seen that the cost function C^ω of coalitions is subadditive: $C^\omega(S \cup T) \leq C^\omega(S) + C^\omega(T)$ for any S and T with $S \cap T = \emptyset$. Formally, a *cooperative game of international CO_2 emissions trading* is defined by a pair (N, C^ω). For detailed discussions on cooperative games, see [21].

3 A Noncooperative Bargaining Process of Emissions Reduction

In this part we formulate a noncooperative bargaining process of emission reduction in the first negotiation phase that was briefly described in Part 2. The negotiation can be formulated as the n-*person unanimous bargaining problem* in the literature of the bargaining theory. Let $\omega = (\omega_1, \cdots, \omega_n) \in R_+^n$ be an emissions allocation for n

countries satisfying $\sum_{i \in N} \omega_i = \bar{\omega}$, and let Ω be the set of all emission allocations ω. Every country i evaluates an allocation $\omega = (\omega_1, \cdots, \omega_n)$ by its competitive equilibrium reduction cost $c_i^e(\omega_i)$ given in (3). The agreement of $\omega = (\omega_1, \cdots, \omega_n)$ is reached by the unanimous consent of all n countries.

To complete the description of the unanimous bargaining problem, we need to specify the *disagreement point* that shows what would happen in the case where negotiations fail. It is difficult for us to estimate the disagreement point of the international negotiations on the Kyoto Protocol. There are a number of uncertain factors relating to climate change from the environmental and economic viewpoints. In addition, the failure of the Kyoto conference will cause further political uncertainty in international negotiations. Because of the difficulty of forecasting future activities as a result of negotiations failure, we simply assume in this chapter that the breakdown of negotiations in the Kyoto conference delays the prevention of global warming and that each country i has to be burdened with some exogenous cost d_i. We assume

$$\sum_{i \in N} d_i > c(\bar{\omega}). \tag{7}$$

This condition means that the disagreement in the Kyoto conference causes higher reduction costs for countries as a whole than they would be under the Kyoto Protocol. The disagreement point of negotiations is given by $d = (d_1, \cdots, d_n) \in R_+^n$. By agreeing to an emissions allocation $\omega = (\omega_1, \cdots, \omega_n)$, each country i can save its reduction costs by $d_i - c_i^e(\omega_i)$.

Definition 2. The *unanimous bargaining problem of CO_2 emission reduction* is defined to be a triplet $B = (\Omega, d, (c_i^e)_{i \in N})$ where Ω is the set of all emissions allocations $\omega = (\omega_1, \cdots, \omega_n) \in R_+^n$ for n countries satisfying $\sum_{i \in N} \omega_i = \bar{\omega}$, $d = (d_1, \cdots, d_n) \in R_+^n$ the disagreement point satisfying (7), and c_i^e the competitive equilibrium reduction cost function of country i defined in (3).

Before presenting a formal model of the negotiation process, we define a cooperative solution of the unanimous bargaining problem introduced by Nash [12].

Definition 3. The *asymmetric Nash bargaining solution* of the unanimous bargaining problem $B = (\Omega, d, (c_i^e)_{i \in N})$ with a weight vector $\alpha = (\alpha_1, \cdots, \alpha_n) \in R_+^n$ is the optimal solution $\omega = (\omega_1, \cdots, \omega_n)$ of the maximization problem

$$\begin{aligned} &\max\ (d_1 - c_1^e(\omega_1))^{\alpha_1} \cdot \cdots \cdot (d_n - c_n^e(\omega_n))^{\alpha_n} \\ &\text{s.t. (i) } \omega = (\omega_1, \cdots, \omega_n) \in \Omega \\ &\qquad \text{(ii) } c_i^e(\omega_i) \leq d_i \quad \text{for all} \quad i = 1, \cdots, n, \end{aligned}$$

where c_i^e is the competitive equilibrium reduction cost function of country i.[6]

The asymmetric Nash bargaining solution maximizes the weighted product of all countries saving costs from emissions trading. The following proposition

[6] (7) guarantees that there exists an emissions allocation $\omega = (\omega_1, \cdots, \omega_n)$ satisfying constraints (i) and (ii).

characterizes the asymmetric Nash bargaining solution.

Proposition 2. If the disagreement point $d = (d_1, \cdots, d_n)$ satisfies $c_i^e(0) \geq d_i$ for all $i \in N$, then the asymmetric Nash bargaining solution $\omega = (\omega_1, \cdots, \omega_n)$ of the unanimous bargaining problem $B = (\Omega, d, (c_i^e)_{i \in N})$ for n countries with a weight vector $\alpha = (\alpha_1, \cdots, \alpha_n) \in R_+^n$ satisfies

$$\frac{d_1 - c_1^e(\omega_1)}{\alpha_1} = \cdots = \frac{d_n - c_n^e(\omega_n)}{\alpha_n}. \tag{8}$$

Proof. We would first remark that the optimal solution of the maximization problem in Definition 3 is unchanged when we replace the objective function $(d_1 - c_1^e(\omega_1))^{\alpha_1} \cdot \cdots \cdot (d_n - c_n^e(\omega_n))^{\alpha_n}$ with its logarithmic value. That is, the Nash bargaining solution $\omega = (\omega_1, \cdots, \omega_n)$ with the weight vector $\alpha = (\alpha_1, \cdots, \alpha_n)$ is the optimal solution of the maximization problem

$$\begin{aligned}
\max\ & \alpha_1 \cdot \log(d_1 - c_1^e(\omega_1)) + \cdots + \alpha_n \cdot \log(d_n - c_n^e(\omega_n)) \\
\text{s.t. } & \text{(i) } \omega_1 + \cdots + \omega_n = \bar{\omega}, \\
& \text{(ii) } \omega_i \geq 0 \quad \text{for all} \quad i = 1, \cdots, n, \\
& \text{(iii) } c_i^e(\omega_i) \leq d_i \quad \text{for all} \quad i = 1, \cdots, n.
\end{aligned}$$

The assumption that $c_i^e(0) \geq d_i$ for all $i \in N$ implies that constraint (ii) is not binding at the optimal solution. Constraint (iii) is not binding from (7), either. Therefore, it can be shown that the Nash bargaining solution $\omega = (\omega_1, \cdots, \omega_n)$ satisfies the first-order condition

$$-\frac{\alpha_i}{d_i - c_i^e(\omega_i)} \frac{dc_i^e(\omega_i)}{d\omega_i} + \lambda = 0, \quad \text{for all} \quad i = 1, \cdots, n, \tag{9}$$

where λ is the Lagrange multiplier corresponding to constraint (i). As we have $dc_i^e(\omega_i)/d\omega_i = -p^*$ for all $i = 1, \cdots, n$ from (3) where p^* is the competitive equilibrium price of emission, the theorem can be proved by (9). QED.

The condition $c_i^e(0) \geq d_i$ in the proposition presumes a natural situation that country i prefers the breakdown of negotiations to an agreement if it is not assigned permissions permits. The proposition shows that the ratio of saving costs to the weight should be equalized across n countries at the asymmetric Nash bargaining solution. Solving (3), (6), and (8), we can obtain an explicit formula of the asymmetric Nash bargaining solution $\omega^* = (\omega_1^*, \cdots, \omega_n^*)$ with the weight vector $\alpha = (\alpha_1, \cdots, \alpha_n)$,

$$\omega_i^* = \frac{1}{p^*}\{c_i^e(0) - d_i + \alpha_i(\sum_{i=1}^{n} d_i - c(\bar{\omega}))\} \tag{10}$$

where $c_i^e(0) = C_i(x_i^*) + p^*(E_i - x_i^*)$. The emission reduction cost $c_i^e(\omega_i^*)$ of country i at the Nash bargaining solution $\omega^* = (\omega_1^*, \cdots, \omega_n^*)$ is given by

$$c_i^e(\omega_i^*) = d_i + \alpha_i(c(\bar{\omega}) - \sum_{i=1}^{n} d_i). \tag{11}$$

We call (11) the *Nash bargaining reduction cost* of country i.

We now turn to our main problem of describing a bargaining process on emission reductions in the first phase of negotiations. We will prove that the asymmetric Nash bargaining solution can be attained as a noncooperative equilibrium of the model and, moreover, that the weight of each country for the Nash bargaining solution can be derived endogenously from the bargaining rule.

The bargaining process consists of a (possibly) infinite sequence of bargaining rounds. The specific rule of the bargaining procedure is as follows.

(1) At the beginning of each round t $(= 1, 2, \cdots)$, every country $i \in N$ is randomly selected as a proposer with a predetermined probability θ_i.

(2) The selected country i proposes an emission allocation $\omega^i = (\omega_1^i, \cdots, \omega_n^i) \in \Omega$.

(3) All other countries in N either accept or reject the proposal ω^i sequentially according to a predetermined order over N. The order of responders does not affect the result in any critical way.

(4) If all responders accept the proposal ω^i, then it is agreed upon. In this case, every country $j \in N$ bears its competitive equilibrium reduction cost $c_j^e(\omega_j^i)$ in the emissions trading.

(5) If any country rejects the proposal ω^i, then the following events may occur. With probability $1 - \varepsilon$ $(0 < \varepsilon < 1)$, negotiations proceed to the next round $t + 1$ and the same process as in round t is repeated. The other possibility is that, with probability ε, negotiations stop and all countries $j \in N$ are burdened with their own costs d_j at the disagreement point $d = (d_1, \cdots, d_n)$.

(6) Every country is fully aware of past moves in the process whenever it makes a decision.

(7) All countries minimize their expected costs of emission reduction after emissions trading.[7]

The flowchart of the negotiation process is illustrated in *Figure 1*. The bargaining model above is denoted by Γ^ε, where ε is the stopping probability of negotiations when a proposal is rejected. The game Γ^ε is formally represented as an infinite-length extensive game with perfect information, where all players can make their choices with perfect knowledge of all past moves.

A (pure) strategy of every country $i \in N$ in the game Γ^ε is defined as a sequence $\sigma_i = (\sigma_i^t)_{t=1}^\infty$ of t-th round strategies σ_i^t, where σ_i^t prescribes (i) a proposal in round t and (ii) a response policy assigning either "yes" or "no" to every possible proposal. Both a proposal and a response policy prescribed by the strategy σ_i may depend on the history of negotiations.

We investigate a stationary subgame perfect equilibrium of the bargaining game Γ^ε. Roughly, a subgame perfect equilibrium of Γ^ε is a strategy combination $\sigma = (\sigma_1, \cdots, \sigma_n)$ that prescribes the optimal action to every country at *every*

[7] In the event with probability zero that negotiations continue forever, we assume that the disagreement point $d = (d_1, \cdots, d_n)$ prevails.

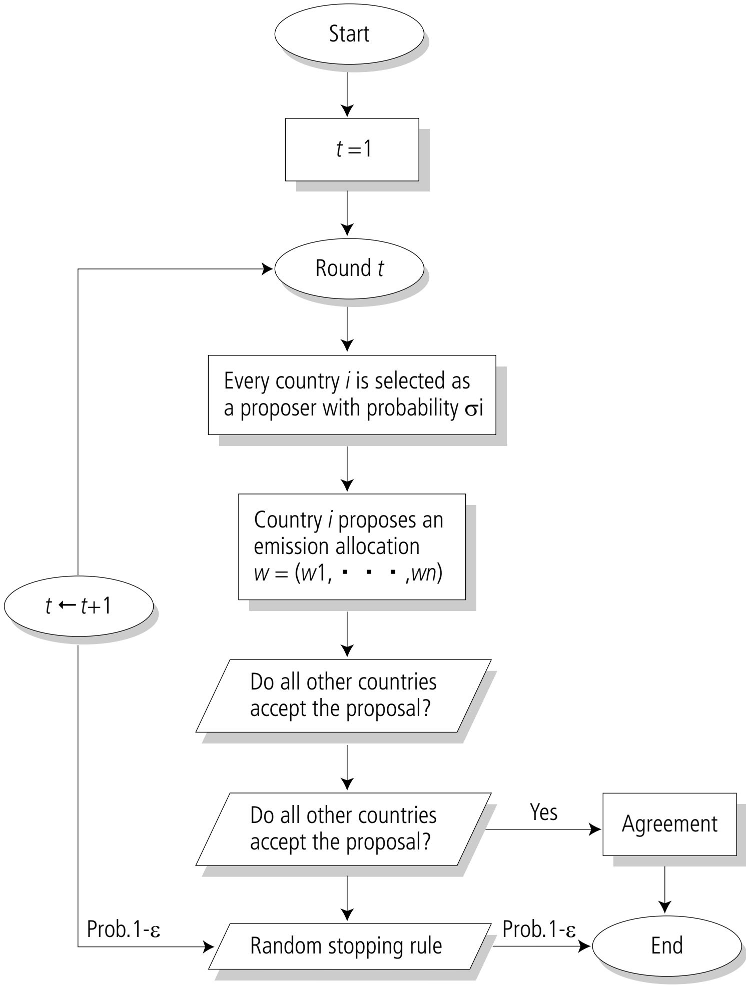

Fig. 1. Flowchart of the negotiation process

possible move of the country in Γ^{ε}, given that all other countries follow σ. For a precise definition of a subgame perfect equilibrium of an extensive game, see [8] or a standard textbook of game theory. A stationary subgame perfect equilibrium of Γ^{ε} is a subgame perfect equilibrium in which every country's equilibrium strategy is *stationary*. A stationary strategy in Γ^{ε} prescribes a proposal and a response policy, independent of negotiation history in past rounds. More precisely, a stationary subgame perfect equilibrium of Γ^{ε} is defined as follows.

Definition 4. A strategy combination $\sigma^* = (\sigma_1^*, \cdots, \sigma_n^*)$ of the bargaining game Γ^{ε} is called a *stationary subgame perfect equilibrium* (SSPE) of Γ^{ε} if it is a subgame perfect equilibrium with the property that for every $i \in N$ and every t $(= 1, 2, \cdots)$, the t-th round strategy σ_i^t of country i is independent of history before round t (the response policy may depend on history within round t).

The solution of a stationary subgame perfect equilibrium is employed in almost all the literature on noncooperative multilateral bargaining models (see [1], [17], [5], [14] and [15], among others).[8] It implies "forgiveness—let bygones be bygones" in the sense that countries do not treat others unfavorably even if they were treated unfavorably themselves in past rounds of negotiations.

In the literature of the equilibrium selection theory developed by Harsanyi and Selten [8], the SSPE of an extensive game satisfies the condition of *subgame consistency* that requires every player to behave in the same way across *isomorphic subgames*. In our bargaining game Γ^ε, all subgames starting with the beginning of all rounds can be isomorphic, as they have identical game trees. The SSPE imposes on every player the duty to make the same proposal whenever he is selected as a proposer, and to respond to every possible proposal independently of the negotiation history in past rounds. It should be noted that the response certainly depends on the proposal itself and on the in-round history, that is, who was a proposer and who has already accepted the proposal.

Two justifications for the SSPE may be possible. One is that negotiators representing countries have a low capacity for information processing and thus tend to behave according to a simple strategy such as a stationary one. This kind of justification based on the strategic complexity is considered by Baron and Kalai [2] and Chatterjee and Sabourian [6]. The other is based on the "focal point" arguments that it is easier for negotiators to coordinate their mutual expectations on stationary strategies. The focal point arguments have received much attention in game theory since the pioneering work of Schelling [19].

We are now ready to state the main theorem that characterizes an SSPE of Γ^ε.

Theorem. There is a unique SSPE in the bargaining game Γ^ε of emission reduction.[9] The expected equilibrium cost of every country in Γ^ε is equal to the asymmetric Nash bargaining reduction cost in $B = (\Omega, d, (c_i^e)_{i \in N})$ where the weight vector is given by the probability distribution $\theta = (\theta_1, \cdots, \theta_n)$ for selecting proposers. Moreover, as the stopping probability ε of negotiations becomes close to zero, the emission allocation proposed by every country converges to the asymmetric Nash bargaining solution of B.

Proof. We first prove the uniqueness of an SSPE in the bargaining game Γ^ε. Let $v^\varepsilon = (v_1^\varepsilon, \cdots, v_n^\varepsilon)$ be the expected equilibrium cost vector of n countries in an SSPE. Suppose that country i is selected as a proposer in the initial round. Each

[8] It is well known that the set of nonstationary subgame perfect equilibria is very large in sequential multilateral bargaining models like ours when the discount rate of future payoffs is close to zero, or when the stopping probability of negotiations is also close to zero.

[9] More precisely, the equilibrium path of an SSPE is unique.

responder j bears the expected cost $(1-\varepsilon)v_j^\varepsilon + \varepsilon d_j$ if negotiations break down, as the equilibrium strategies are stationary. By backward induction, this fact implies that if country i proposes an emissions allocation where all other countries $j\ (\neq i)$ bear no more than the expected costs $(1-\varepsilon)v_j^\varepsilon + \varepsilon d_j$, the proposal is accepted by them.[10] Therefore, country i bears optimally the cost $c(\bar{\omega}) - \sum_{j\neq i}\{(1-\varepsilon)v_j^\varepsilon + \varepsilon d_j\}$ if it is selected as a proposer. By the definition of the expected cost v_i^ε, the following equation must hold,

$$v_i^\varepsilon = \theta_i[c(\bar{\omega}) - \sum_{j\neq i}\{(1-\varepsilon)v_j^\varepsilon + \varepsilon d_j)\}] + (1-\theta_i)\{(1-\varepsilon)v_i^\varepsilon + \varepsilon d_i\}, \text{for all } i \in N. \quad (12)$$

Let $\bar{v} = \sum_{i=1}^n v_i^\varepsilon$ and $\bar{d} = \sum_{i=1}^n d_i$. Then, (12) can be arranged as

$$v_i^\varepsilon = \theta_i[c(\bar{\omega}) - (1-\varepsilon)\bar{v} - \varepsilon\bar{d}] + (1-\varepsilon)v_i^\varepsilon + \varepsilon d_i, \quad \text{for all } i \in N. \quad (13)$$

By summing both sides of (13) for all $i \in N$, we obtain $\bar{v} = c(\bar{\omega})$. By substituting this into (13) and solving it, we can obtain

$$v_i^\varepsilon = d_i + \theta_i[c(\bar{\omega}) - \bar{d}] \quad \text{for all } i \in N. \quad (14)$$

That is, the expected equilibrium cost of every country i is equal to the Nash bargaining reduction cost (11) in B with the weight vector $\theta = (\theta_1, \cdots, \theta_n)$. Note that v_i^ε is independent of ε. It follows from (14) that an SSPE in Γ^ε must be unique. In equilibrium, every country i proposes the emission allocation $\omega^i(\varepsilon) = (\omega_1^i(\varepsilon), \cdots, \omega_n^i(\varepsilon)) \in \Omega$, satisfying $c_i^e(\omega_i^i(\varepsilon)) = c(\bar{\omega}) - \sum_{j\neq i} c_j^e(\omega_j^i(\varepsilon))$ and $c_j^e(\omega_j^i(\varepsilon)) = (1-\varepsilon)v_j^\varepsilon + \varepsilon d_j$ for all $j \neq i$, where c_j^e is the country j's competitive equilibrium reduction cost given in (3).

Finally, the arguments above show that the reduction costs $(c_1^e(\omega_1^i(\varepsilon)), \cdots, c_n^e(\omega_n^i(\varepsilon))$ proposed by every country i converge to the Nash bargaining reduction costs $v^\varepsilon = (v_1^\varepsilon, \cdots, v_n^\varepsilon)$ as ε goes to zero. Since the correspondence in (3) between an emissions allocation and a competitive equilibrium reduction cost vector is continuous and one-to-one, every country i's proposal $\omega^i(\varepsilon)$ converges to the asymmetric Nash bargaining solution of B as the stopping probability ε of negotiations goes to zero. QED.

The theorem shows that the asymmetric Nash bargaining solution can be supported as the SSPE of the bargaining model Γ^ϵ where the stopping probability ϵ of negotiations is sufficiently small. The weights of countries for the asymmetric Nash bargaining solution are determined by their probabilities of being selected as proposers. This result implies that the bargaining power of every country comes from the opportunity to make a proposal in the bargaining process.

[10] When responders j bear exactly the expected costs $(1-\varepsilon)v_j^\varepsilon + \varepsilon d_j$, they are indifferent to accepting or rejecting the proposal. However, the proposer's cost minimization in the SSPE induces the acceptance by responders on an equilibrium path, even in this instance.

4 Numerical Results

In this part, based on actual data, we compute the competitive equilibrium of international emissions trading among four major emitting countries the EU, the FSU, Japan, and the USA, based on actual data. The Nash bargaining solution of the emissions allocation is analyzed empirically. We consider three different weights, equal weight, population weight, and GDP weight. From the numerical results we consider how the formal model of the Nash bargaining solution can explain the actual agreement of reduction rates in the Kyoto Protocol.

The most difficult task in our empirical analysis is to estimate countries' marginal abatement costs of carbon emissions. Based on a survey of nine different estimates in the USA and Western Europe, Nordhaus [13] derives the following logarithmic functional form as the marginal abatement costs of the USA:

$$MC_{USA}(R) = -185.2 \cdot \ln(1 - R), \tag{15}$$

in units of 1989 US dollars per ton of carbon where R is the rate of emission reductions (x/E).[11]

Based on the estimation by Nordhaus [13], Bohm and Larsen [4] derive marginal reduction costs of carbon emissions for other countries. Their method is to obtain the marginal cost function of emission reductions for other countries by modifying that of the USA, taking into account the differences of countries' carbon intensities (E/GDP). Let e_{USA} be the carbon intensity of the USA (reference country). Bohm and Larsen assume that if any country l has carbon intensity e_l lower than the USA, it has already taken appropriate measures to reduce its carbon intensity. Thus, its marginal cost function starts at a higher level R_l^0 along the MC-curve of the USA. *Figure 2* illustrates the USA's marginal cost function of emission reduction and the starting point R_l^0 for country l's marginal cost.[12] The curve is steeper at R_l^0 than at the origin of the USA. The starting level R_l^0 is assumed to be

$$r_l = 1 - \frac{e_l}{e_{USA}}. \tag{16}$$

Then, the marginal cost function of country l is computed as

$$\begin{aligned} MC_l(R_l) &= 185.2 \cdot \ln(1 - r_l) - 185.2 \cdot \ln(1 - r_l - R_l) \\ &= -185.2 \cdot \ln(1 - \frac{R_l}{1 - r_l}), \end{aligned} \tag{17}$$

where R_l is the rate of emission reduction by country l.

Similarly, if country h has carbon intensity e_h higher than the USA, its marginal cost of emission reduction starts at a negative level R_h^0 along the MC-curve of the USA (see *Figure 2*). The starting point R_h^0 of country h is assumed to be

[11] In the above form we assume zero intercept (Nordhaus [13] estimates an intercept of -4.13).

[12] *Figure 2* is also given in [16].

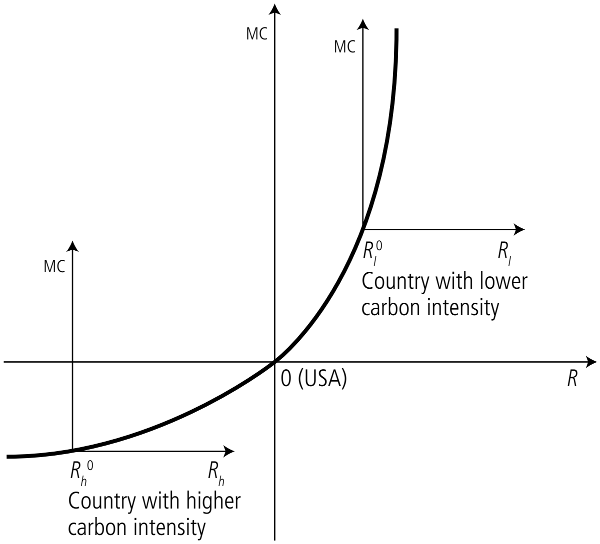

Fig. 2. Marginal cost (mc) functions of different countries

$$r_h = -1 + \frac{e_{USA}}{e_h}.$$

The marginal cost function of country h can be given by the same formula as (17), that is,

$$\begin{aligned} MC_h(R_h) &= -185.2 \cdot \ln(1 - r_h - R_h) + 185.2 \cdot \ln(1 - r_h) \\ &= -185.2 \cdot \ln(1 - \frac{R_h}{1 - r_h}). \end{aligned}$$

where R_h is the rate of emission reduction by country h. The curve is flatter at r_h than at the origin of the USA.

Under the estimated marginal reduction costs of countries, we can obtain the explicit formula of the competitive equilibrium of CO_2 emissions trading (see [16]).

Proposition 3. Under the marginal cost functions (17) of emissions abatement and an emission allocation $\omega = (\omega_1, \cdots, \omega_n)$, the equilibrium price p^* of emissions, the equilibrium reduction x_i^* of country i, and its equilibrium reduction cost $c_i^e(\omega_i)$ are given by

$$p^* = -185.2 \cdot \ln[1 - \frac{E - \bar{\omega}}{\sum_{i=1}^n E_i(1 - r_i)}] \tag{18}$$

$$x_i^* = \frac{E_i(1 - r_i)}{\sum_{j=1}^n E_j(1 - r_j)}(E - \bar{\omega}) \tag{19}$$

$$c_i^e(\omega_i) = 185.2 x_i^* + p^*(E_i r_i - \omega_i), \tag{20}$$

where $r_i = 1 - e_i/e_{USA}$ if country i's carbon intensity e_i is lower than e_{USA}, and $r_i = -1 + e_{USA}/e_i$, if otherwise.

Proof. By solving (2), (4) and (17), we can prove (18) and (19). Since the country i's equilibrium reduction cost $C_i(x_i^*)$ is computed as

$$\begin{aligned} C_i(x_i^*) &= \int_0^{x_i^*} -185.2 \cdot \ln[1 - \frac{t}{E_i(1-r_i)}]dt \\ &= 185.2[E_i(1-r_i) - x_i^*] \cdot \ln[1 - \frac{x_i^*}{E_i(1-r_i)}] + 185.2x_i^* \\ &= p^*[x_i^* - E_i(1-r_i)] + 185.2x_i^*, \end{aligned}$$

we can prove (20) by (3). QED.

With the help of Proposition 3, we can compute the competitive equilibrium of international emissions trading among the EU, the FSU, Japan, and the USA. *Table 1* shows all relevant data for 1990 on these countries.[13] Data include:

- Carbon emissions (E_i);
- Population (n_i);
- Gross domestic products (GDP_i);
- Carbon intensity ($e_i \equiv E_i/GDP_i$).

It can be seen in *Table 1* that carbon intensities are diverse among the major emitting countries. The EU and Japan have the lowest carbon intensities, and the FSU has the highest. The USA is in an intermediate position. These data imply that the EU and Japan have already made a great effort to reduce carbon emissions in their domestic production and that these countries have to bear high costs to satisfy their reduction commitment under the Kyoto Protocol if emissions trading is not allowed. *Table 2* shows the reduction cost of each country without emissions trading. The EU bears the highest reduction costs ($991 million) without emissions trading. The FSU does not bear any reduction costs.

Table 3 summarizes numerical results on the competitive equilibrium of international emissions trading under the Kyoto Protocol. The efficient reduction share among the four countries are roughly 12 percent by the EU, 47 percent by the FSU, 4 percent by Japan, and 37 percent by the USA. The small reduction share of the EU and Japan reflects the fact that the marginal reduction costs of these two countries are less than those of the FSU and the USA. The reduction shares of the EU, Japan, and the USA are lower compared with those without emissions trading. This means that these three countries will be buyers of emissions in the trading. The FSU will sell about $83 million tons of emissions. It is remarkable that the total emission reduction costs can be reduced significantly by emissions trading. The total saving costs amount to about $1,030 million. In examining the saving cost for each country,

[13] Data sources: (1) Carbon emissions: [9]. (2) Population: [24] (population of the FSU is compiled by Toyo Keizai Data Bank [23]). (3) GDP: [26] [25] (for the FSU).

Table 1. Data on countries

Country	Carbon emissions	Population	GDP	Carbon intensity
Austria	16	7.72	162	0.1
Belgium	27	9.97	197	0.14
Denmark	14	5.14	133	0.11
Finland	14	4.99	137	0.1
France	99	56.74	1,216	0.08
Germany*	243	79.37	1,770	0.14
Greece	20	10.09	84	0.24
Ireland	8	3.50	47	0.17
Italy	106	57.66	1,102	0.1
Luxembourg	3	0.38	11	0.27
Netherlands	38	14.95	295	0.13
Portugal	12	9.87	70	0.17
Spain	58	38.96	514	0.11
Sweden	13	8.56	238	0.05
United Kingdom	155	57.56	985	0.16
EU	826	365.46	6,961	0.12
FSU	1011	289.34	1,535	0.66
Japan	292	123.54	2,970	0.098
USA	1315	249.92	5,794	0.23
Total	3444	1028.26	17,260	

Note: Carbon emission indices (EI): million tons; population (1990) natural increase (NI): million; GDP: 1990 billion US dollars
*GDP of Germany is for 1991.

Table 2. CO_2 reduction under the Kyoto Protocol without emissions trading

Country	Initial permits (million tons)	Reduction rate	Reduction (million tons)	Reduction costs (million $)
EU	760	0.08	66	991
FSU	1011	0.00	0	0
Japan	274	0.06	18	251
USA	1223	0.07	92	610
Total	3268		176	1852

the EU enjoys the highest saving costs (about $470 million). The FSU can earn a profit (about $400 million) by selling emissions. Japan also benefits from emissions trading, although its saving cost is not as large as that of the EU. The saving costs of the USA is marginal ($52 million). From this result, it can be said that the EU and the FSU benefit the most from emissions trading. The equilibrium price of emissions is $9.45 per carbon ton in the emissions trading among the four countries. In our previous study [16], the equilibrium price of emissions was $6.27 among the three countries without the EU. The present result shows that the emissions price increases if the EU, a large buyer, enters the emissions market.

Table 3. CO_2 reduction under the Kyoto Protocol with emissions trading

Country	Initial permits (million tons)	Initial reduction (million tons)	Efficient reduction million tons)	Efficient reduction share	Equilibrium costs (million \$)	Saving costs (million \$)	Permit price (\$/ton)
EU	760	66	21.37	0.121	522	469	
FSU	1011	0	82.97	0.471	–398	398	
Japan	274	18	6.25	0.036	141	110	9.45
USA	1223	92	65.42	0.372	558	52	
Total	3268	176	176.00	1	823	1029	

We next consider how the asymmetric Nash bargaining solution can explain the actual agreement of reduction rates in the Kyoto Protocol. To compute the Nash bargaining solution we need data on the disagreement point $d = (d_1, \cdots, d_n)$ of negotiations. As we do not have any reliable estimate of such data, we consider a converse question: what disagreement points can support the reduction rates in the Kyoto Protocol as the asymmetric Nash bargaining solution? Mathematically, we consider a solution $d = (d_1, \cdots, d_n)$ of the following linear equations:

$$d_i = c_i^e(\omega_i) + \alpha_i(\sum_{i=1}^{n} d_i - c(\bar{\omega})), \quad \text{for all } i = \text{EU, FSU, Japan, USA}, \tag{21}$$

where ω_i is the emission assigned to country i by the Kyoto Protocol, and $c_i^e(\omega_i)$ is country i's equilibrium reduction costs given in *Table 3*. It should be noted that (21) includes three independent linear equations. This means that there exist a continuum of solutions $d = (d_1, \cdots, d_n)$ for (21). Given the total disagreement costs $\bar{d} = \sum_{i=1}^{4} d_i$, (21) has a unique solution of the disagreement point.

For the asymmetric Nash bargaining solution, we consider three kinds of weights: equal weight, population weight, and GDP weight. The population (GDP, respectively) weight determines countries' weights to be proportional to their populations (GDPs, respectively). According to *Table 1*, the EU has the largest weight under both rules of population and GDP. Japan has the smallest weight under the population rule, and FSU has the smallest weight under the GDP rule. *Table 4* gives several disagreement points which can support the reduction rates in the Kyoto Protocol as the asymmetric Nash bargaining solution with different weights.

The main findings of *Table 4* are as follows. GDP gives a very low weight to the FSU. This means that the bargaining power of the FSU is quite low in the negotiation process. Reflecting this fact, *Table 4* shows that, in order for the Nash bargaining solution with the GDP weights to support the reduction rates in the Kyoto Protocol, the FSU should earn profits when negotiations fail. As this is an unlikely event we can conclude that the Nash bargaining solution with the GDP weights fail to explain the agreement of the Kyoto Protocol. In both the equal weights and the population weights, the disagreement costs of the EU and the USA are larger than those of Japan and the FSU. This means that if the EU and the USA estimate as high their damages in the failure of the Kyoto Protocol, the reduction rates agreed in the Kyoto Protocol

Table 4. Possible disagreement points yielding reduction commitments under the Kyoto Protocol as the Nash bargaining solution

Population weight				
Total costs in disagreement (million $)	3000	4000	5000	weights
EU	1296	1652	2007	0.35
FSU	215	496	777	0.28
Japan	403	523	643	0.12
USA	1087	1331	1574	0.24
GDP weight				
Total costs in disagreement (million $)	3000	4000	5000	weights
EU	1400	1803	2207	0.4
FSU	–205	–116	–27	0.09
Japan	516	688	860	0.17
USA	1289	1625	1960	0.34

can be supported as the asymmetric Nash bargaining solutions both with the equal weights and with the population weights.

5 Conclusion

We have considered international negotiations on CO_2 emission reduction committed under the Kyoto Protocol in the framework of game theory. Specifically, we have presented a noncooperative bargaining model of negotiations on emission reduction. The main theorem shows there is a unique stationary subgame perfect equilibrium of the bargaining model and that the equilibrium proposal by every country converges to the asymmetric Nash bargaining solution of emission reduction as the stopping probability of negotiations goes to zero. In the model, one country is selected randomly as a proposer at the beginning of every round. It is proved that the weight of every country for the asymmetric Nash bargaining solution is determined by the probability of being selected as a proposer. This result implies that the bargaining power of a country emerges from the opportunity of making a proposal in the bargaining process.

In the last part of the chapter, we have considered using empirical data on the four major emitting countries/groups of countries how the formal model of the Nash bargaining solution can explain the actual agreement of reduction rates in the Kyoto Protocol. We have computed the competitive equilibrium of emissions trading among the EU, the FSU, Japan, and the USA based on the estimation of these countries' marginal costs of emission reduction in the literature. The numerical results are summarized as follows: the emissions price is $9.45 per carbon ton. The EU, Japan, and the USA buy in total about 83 million tons of emissions from the FSU. The FSU earns about $400 million from trading. The emissions trading significantly reduces the total reduction costs for the four. The total cost savings amount to $1,030 million. Regarding the cost savings of individual countries/groups of countries, the EU enjoys the highest cost savings ($470 million) and Japan the second highest ($110 million). The USA has the least cost savings ($52 million). The numerical results show that if the equal weight or the population weight is used, the asymmetric Nash bargaining solution can support the reduction rates in the Kyoto Protocol when the EU and the USA evaluate their damages, should negotiations fail, as being greater than those of Japan and the FSU. The Nash bargaining solution with the GDP weight fails to explain the actual agreement of reduction rates. The main reason is that the bargaining power of the FSU is low under the GDP rule, while the Kyoto Protocol gives the FSU a favored outcome of zero reduction for its participation in the Protocol.

In addition to the numerical results, some political events before and after the Kyoto conference seem to support the result of the formal model in this chapter. Before the Kyoto conference, the EU, Japan, and the USA made three major proposals with the intention of affecting the outcome of the agreement in the conference in their favor. The EU made an early proposal to reduce emissions to 15 percent below the 1990 levels by 2010. They also proposed a "bubble" approach that would group all the European countries together in reduction commitments. Japan proposed a 5 percent reduction, claiming to differentiate individual countries based on emissions per GDP, emissions per capita, and population. The USA proposed that developed countries should reduce their greenhouse gases to 1990 levels by 2008–2012 and that developing countries should participate in emission reductions in a meaningful way. It is conceivable that these proposals increased the bargaining power of the three countries. After the Kyoto conference, the USA, led by the its new Republican president, changed its policy to oppose the Kyoto Protocol (March 2001). The withdrawal of the USA from the Kyoto Protocol is indicated by our numerical result that its reduction in cost by emissions trading is marginal ($52 million). There is no evidence that the USA estimates damage to it caused by the failure of negotiations as high. It is argued that an early commitment to reduce emissions would result in serious harm to the United States economy.

Finally, our analysis shows the usefulness and the limits of a game-theoretic approach to international negotiations on climate change. Game theory enables us to understand strategic conflicts in international negotiations on emission reductions. On the other hand, if one wants to derive some practical results from game-theoretic analysis, reliable estimations of model parameters are necessary. In our model we need to estimate the amount of damage to countries by the failure of negotiations on the Kyoto Protocol. Although this is by no means an easy task, further empirical studies on international negotiations based on game-theoretic models should be undertaken in future research.

References

1. Baron D, Ferejohn J (1989) Bargaining in legislatures. American Political Science Review 83:1181–1206

2. Baron D, Kalai, E (1993) The simplest equilibrium of a majority rule decision game. Journal of Economic Theory 61:290–301
3. Barrett S (1992) 'Acceptable' Allocations of tradeable carbon emission entitlements in a global warming treaty. Combating Global Warming, UNCTAD, UN, New York, USA
4. Bohm P, Larsen B (1994) Fairness in a tradeable-permit treaty for carbon emissions reductions in Europe and the former Soviet Union. Environmental and Resource Economics 4:219–239
5. Chatterjee K, Dutta B, Ray D, Sengupta K (1993) A noncooperative theory of coalitional bargaining. Review of Economic Studies 60:463–477
6. Chatterjee K, Sabourian H (2000) Multiperson bargaining and strategic complexity. Econometrica 68:1491–1509
7. Grubb, M, Christiaan, Brack D (1999) The Kyoto Protocol: a guide and assessment. The Royal Institute of International Affairs, London, UK
8. Harsanyi JC, and Selten R (1988) A general theory of equilibrium selection in games. The MIT Press, Cambridge, MA, USA
9. Marland G, Boden TA, Andres RJ (2001) Global, regional and national CO_2 emission estimates from fossil fuel burning, cement production, and gas flaring: 1751–1998. NDP-030, CDIAC, Oak Ridge National Laboratory. http://cdiac.esd.ornl.gov/ndps/ndp030.html
10. Moldovanu B, Winter E (1995) Order independent equilibria. Games and Economic Behavior 9:1995, 21–34
11. Moulin H (1984) Implementing the Kalai-Smorodinsky bargaining solution. Journal of Economic Theory 33:32–45
12. Nash JF (1950) The bargaining problem. Econometrica 18:155–162
13. Nordhaus WD (1991) The cost of slowing climate change: a survey. The Energy Journal 12:37–65
14. Okada A (1996) A noncooperative coalitional bargaining game with random proposers. Games and Economic Behavior 16:97–108
15. Okada A (2000) The efficiency principle in noncooperative coalitional bargaining. Japanese Economic Review 51:34–50
16. Okada A (2002) A market game analysis of international CO_2 emissions trading: evaluating initial allocation rules. In Sawa T (ed), International frameworks and technological strategies to prevent climate change. Springer-Verlag, Tokyo, Japan, 3–21
17. Perry M, Reny PJ (1994) A noncooperative view of coalition formation and the core. Econometrica 62:795–817
18. Rubinstein A (1982) Perfect equilibrium in a bargaining model. Econometrica 50:97–109
19. Schelling TC (1960) The strategy of conflict. Harvard University Press, Cambridge, MA, USA
20. Selten R (1981) A noncooperative model of characteristic-function bargaining. In: Boehm V, Nachthanp H (eds), Essays in game theory and mathematical economics in honor of Oscar Morgenstern. Bibliographisches Institut Mannheim, Germany, 139–151
21. Shubik M (1983) Game theory in the social sciences: concepts and solutions. MIT Press, Cambridge, MA, USA
22. Tadenuma K (2004) International negotiations for reduction of greenhouse gases with emission permits trading. Project on Intergenerational Equity (PIE) Discussion paper, Series No. 191, Hitosubashi University, Tokyo, Japan
23. Toyo Keizai Data Bank (1993) Toyo Keizai Inc., Tokyo, Japan
24. United Nations Demographic Yearbook (1993) United Nations, New York, USA
25. United Nations Statistical Yearbook (1993) United Nations, New York, USA
26. United Nations Statistical Yearbook (1998) United Nations, New York, USA

The Graph Model for Conflict Resolution as a Tool for Negotiators

D. Marc Kilgour

Wilfrid Laurier University, Waterloo, Ontario, Canada

1 Supporting Negotiators

Negotiation is a common human process, but its outcome is often crucial at the personal, corporate, legal, and international levels. The importance of negotiation is evidenced by the enormous amount of advice and expertise that is available to assist negotiators. Much of this guidance, such as the books of the *Getting to Yes* series, Fisher *et al.* [6], for example, counsels negotiators not only on their conduct at the bargaining table but also on how to plan and prepare. On the analytical side of this literature (for example, [15], [16]), there are calls for systems to analyze the strategic problems of negotiators, but few reports of success.

There have been many suggestions that game theory "should be" the natural tool for analyzing strategic problems. But for various reasons, including its insistence on fixed rules of play and its strong assumptions about shared knowledge, game theory has been dismissed as "theoretical acrobatics" by some who study negotiation, specifically (Raiffa [16], p. 6). Later, Raiffa explained "(F)or a long time I have found the assumptions made in standard game theory too restrictive for it to have wide applicability (to negotiation). Such limits are hard to swallow in seeking to put this elegant theory into practice." ([16], p. 12)

This article proposes the Graph Model for Conflict Resolution (GMCR) as a tool that can assist negotiators with the strategic aspects of negotiation. This methodology draws on game theory, but is not game-theoretic. Its strength is its simplicity and flexibility, both in modeling and analyzing strategic decision problems. The graph model incorporates some restrictions on knowledge and rationality, making it appropriate for advising individuals in a multi-decision-maker process, rather than assessing normatively choices that "should" be made or outcomes that "should" be achieved.

Another advantage of the Graph Model for Conflict Resolution methodology is that it has been implemented in a decision-support system, most recently GMCR II. Computer-based support has great advantages in speed, accuracy, reliability, and versatility. The fact that the graph model can be delivered as computer-based support is what makes it most promising in terms of genuine assistance to real-world negotiators.

Many computer-based support systems for negotiations are now available. Starke and Rangaswamy [18] provide a useful review of these systems, concentrating on regulated communication systems and expert systems that provide tactical advice. In the classification of *Table 1*, the major dichotomy is support for the negotiation (i.e., for all negotiators) versus support for the negotiator (i.e., for a particular negotiator).

Table 1. A simple taxonomy of negotiation support systems.

Negotiation support	Negotiator support	
(1) Regulated communication	(1) Expert systems	
(2) Process support	(2) Decision-analysis systems	
(3) Context support	(2a) Single participant	(2b) Multiple participant

Source: [12]

Our focus is on negotiator support. Within this class, expert systems suggest bargaining tactics and principles, often tailored in the context of the negotiation or the nature of the negotiators. Decision-analysis systems are subdivided into systems based on single-participant models (for a recent survey of these, see DAS [2]) and systems based on multiple-participant models. The Graph Model for Conflict Resolution and the associated decision-support system, GMCR II, fall into this latter category.

Single-participant negotiator support systems identify a decision maker's best course of action based on some assignment of subjective probabilities (or other measure of likelihood and relevance) to actions and events outside the decision-maker's control. In contrast, multiple-participant systems concentrate on strategic analysis—asking what actions are in a particular decision-maker's interest and what responses would be in the interests of the other participants. The objective of this article is to explain how the Graph Model methodology (hereinafter simply called the Graph Model) can contribute to a negotiator's success.

2 Overview of the Graph Model for Conflict Resolution

The purpose of the Graph Model is to model and analyze strategic conflicts and to provide insights into and understanding of strategic decisions. More specifically, the Graph Model is designed to provide strategic advice to a client (who would typically be a participant in the conflict), a mediator, or a policymaker. Negotiation is strategic conflict because it involves choices by two or more parties whose goals differ. The Graph Model tool is therefore appropriate to the task of assisting negotiators. Kilgour *et al.* ([11], [12]) argue that this methodology can be extremely useful for negotiation preparation. This chapter updates and broadens that argument, taking into account not only new technical developments in the Graph Model, but also a new approach to the interpretation of Graph Model analysis. (For a more detailed overview of the methodology, see the special issue of *Group Decision and Negotiation*, with its introduction by Kilgour and Hipel [9].)

The Graph Model methodology was first published in Kilgour *et al.* [10], and the first comprehensive treatment appeared as Fang *et al.* [3]. Up-to-date references to developments from the original model, including methodological extensions, advances in technical implementation, and applications, can be found on the Conflict Analysis Group Web site [1].

The Graph Model provides a simple, flexible way to model and analyze strategic conflicts. It represents conflicts as processes of discrete movement within a set (usually finite) of "states"; the model includes decision makers (DMs), who control movements among the states. A graph model is described by a family of (directed) graphs, one for each DM. All these graphs have the same set of vertices, which correspond to the states of the model. The movements between states controlled by a particular DM are represented by (directed) arcs or edges in that DM's graph. As all DM graphs have a common vertex set, it is often convenient to show all DMs' graphs together as the *integrated graph* of the model, in which the edges of each DM are labeled with the DM's name. One such integrated graph is given at the top of *Figure 1*, which shows a complete graph model of the Canada–US Softwood Lumber Negotiation during a crucial phase in late 1986. (Some modeling details are summarized below.) The very simple model in *Figure 1* has three decision makers and 13 states.

Besides the integrated graph, *Figure 1* also exhibits the second component of a graph model, the DMs' preferences over the states (considered as final resolutions of the conflict). Note that the DMs have different preferences, as is surely to be expected in conflicts. In this model, preferences are transitive (which is usually the case); intransitive preferences can be incorporated, but they limit the analysis. Observe that one DM, the United States (US) Commerce Department, has indifferences in its preference ordering (i.e., some states are "tied" in preference, indicated in the ordering by lines joining equally preferred states). The other two DMs' preferences are strict orderings (i.e., contain no ties).

2.1 Example: The Canada–US Softwood Lumber Negotiation, 1986

We now describe the situation modeled in *Figure 1* to illustrate how the Graph Model can help us to understand a strategic conflict. (For additional details and references to original sources, see Fang *et al.* [3] and Kilgour *et al.* [12]. This specific model was originally developed as an illustration to Fang *et al.* [5].)

Softwood lumber exports from Canada to the United States (USA) were about $CDN 2 billion per year in the 1980s. By 1986 almost two-thirds of Canadian production was exported to the United States and represented about one-third of the US market. The increase in imports from Canada had coincided with the decline of the US lumber industry, and politicians and industry groups complained that the Canadian product had an unfair competitive advantage. The main issue was stumpage fees—the charges for harvesting trees on federally owned land—which were alleged to be artificially low in Canada, amounting to an unfair subsidy.

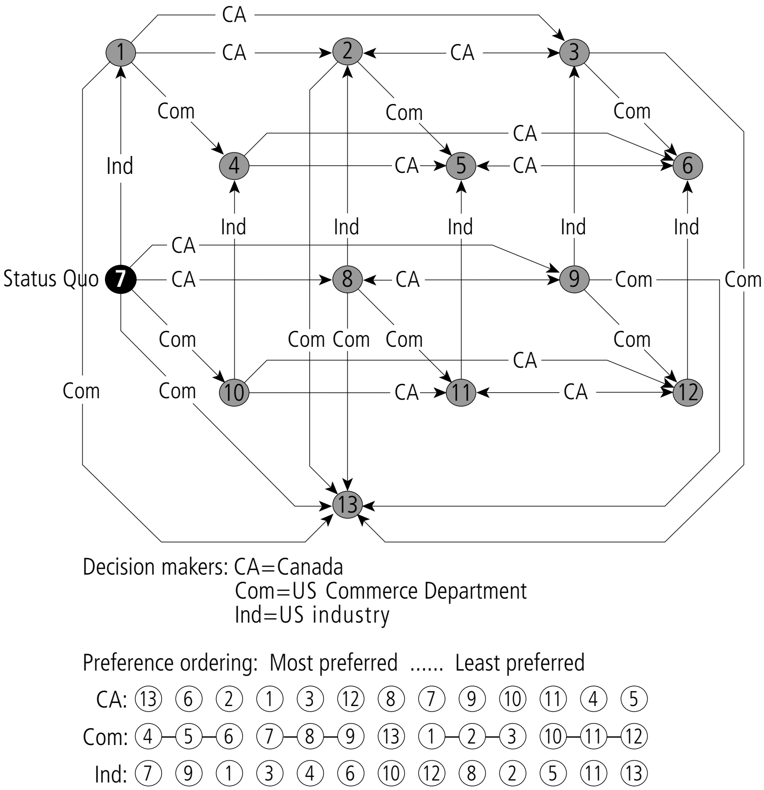

Fig. 1. Integrated Graph and Decision Makers' Preference Orderings for Softwood Lumber Negotiation Model
Sources: [3], [5]

In May 1986 the Coalition for Fair Canadian Lumber Imports, a group established by the US lumber industry, petitioned the International Trade Commission (ITC), part of the US Commerce Department, arguing that US industry had been harmed by Canadian imports. The Coalition asked for a duty of 27 percent on imports from Canada. In June the ITC found that imports from Canada were indeed harming the US industry and, in accordance with established procedure, sent the case to the International Trade Administration (ITA), another body within the Commerce Department, to determine an appropriate remedy. The ITA decision to require a 15 percent duty was announced on 16 October 1986. The decision was not final, however; it required Commerce Department ratification, which had to occur on or before 30 December 1986.

The model of *Figure 1* describes the situation after the October decision. There are three decision makers, Canada (representing the federal government and the affected provinces), the US Commerce Department (including the ITC, which was involved in the final injury ruling), and US industry, as represented by the Coalition for Fair Canadian Lumber Imports.

The states of the model are based on the following actions available to the DMs:

- Canada had to choose exactly one of the following: (Du) accept the import duty, (L) attempt other sanctions, including legal action, or (T) impose an export tax on Canadian producers.
- The US Commerce Department had to choose exactly one of: (Rt) retain the import duty, (Dr) drop the import duty in return for an export tax, or (Rj) reject the industry petition.
- US industry could either continue to press its petition (P), or withdraw it (W).

The states in the graph model of *Figure 1* consist of all possible combinations of these choices, except that Commerce's decision to Reject is considered to "trump" all other actions. In other words, if Commerce chooses Reject, the choices of the other decision makers do not matter. The states are numbered for convenience, as shown in *Table 2*. For example, the status quo state, or initial position of the model, is state 7. At the status quo, there will be a duty on Canadian Softwood Lumber exported to the US (unless the Commerce Department decides otherwise); meanwhile, US industry is continuing to press its petition (it wanted an even higher duty), and Canada accepts the situation. A contrasting outcome is state 13, which results if Commerce rejects the duty.

The possible movements between states, which are controlled by the DMs, are also specified in a graph model. In the Softwood Lumber case, these state transitions reflect the following assumptions about allowable actions: 1) Canada can move away from accepting the duty to either fighting it with legal action or imposing an export tax, and can switch back and forth between these two positions, but never move back to accepting the duty; 2) Commerce can move from retaining the duty to dropping it in return for an export tax or rejecting it altogether, but after either of these decisions it cannot move further; and 3) Industry can withdraw the petition at any time, but if it does so the petition cannot be revived. (The latter condition does not mean that another petition could not be filed in the future; it means that a new model would be needed to analyze the situation after the new petition is filed.)

Table 2. State definitions for *Figure 1* Graph Model

DM/State	1	2	3	4	5	6	7	8	9	10	11	12	13
Canada	Du	L	T	Du	L	T	Du	L	T	Du	L	T	–
Commerce	Rt	Rt	Rt	Dr	Dr	Dr	Rt	Rt	Rt	Dr	Dr	Dr	Rj
Industry	W	W	W	W	W	W	P	P	P	P	P	P	–

The second component of a graph model consists of preference orderings for all DMs. The preferences shown in *Figure 1* were constructed based on a study of the situation. For example, Canada most wanted the petition rejected but if not, it next preferred that the duty be replaced by an export tax and that this substitution be accepted by the other parties. If the export tax offer was rejected, it preferred a legal challenge, or simply acceptance of the duty, provided that the petition was withdrawn. Commerce wanted to avoid being "offside" with Industry; it had a strong preference for dropping the duty if Industry withdrew the petition, but for retaining it if Industry pressed the petition. All other aspects of the preferences shown in *Figure 1* were determined through further analysis of the situation.

3 Using the Graph Model

One important contribution of the Graph Model for Conflict Resolution methodology is to provide a simple and flexible modeling structure for strategic conflicts. Often the analyst gains useful insights merely from constructing a model. The Graph Model can also be used to predict outcomes or assess choices, but these are secondary functions. The major objective of the methodology is to provide strategic advice, which could involve giving a DM a prescription for actions to achieve the outcomes it would prefer, or suggesting to a mediator which possible resolutions of the conflict would be stable, or revealing to a policymaker how the strategic structure of the situation shapes the outcome. This strategic advice reflects the capacity of the Graph Model to identify strategic stability.

Note that the Graph Model is intended to be used prescriptively—to advise a client who may be a DM in the conflict or an outside analyst—rather than descriptively or normatively—to predict which choices will be made, or to determine which choices "should" be made. Thus the objectives of the Graph Model are closer to Decision Theory than to Game Theory. This approach is motivated in part by the practical difficulty of obtaining data that are objectively correct, as would be required for a normative or descriptive analysis. To sidestep this problem, the analysis is carried out on behalf of a "client," who is responsible for providing data. Provided certain consistency conditions are met, the data is accepted without question by the analyst, who then crafts one or more graph models to represent the situation, analyzes the models, and interprets the results of the analysis.

Analysis is central to the Graph Model. The usual way of analyzing a graph model is to evaluate the stability of every state from the point of view of every DM. To complicate the procedure, there are many ways of defining stability within the Graph Model, reflecting that decision makers in conflicts may approach their decisions in many different ways. But, although the objectives of the Graph Model are different from those of Game Theory, the stability analysis algorithms are all equivalent to the analysis of certain simple games that can be constructed from a graph model.

To understand stability, recall that a graph model provides a picture of how a strategic conflict can evolve by state-to-state transitions beginning at the status quo state. The current state is known to all DMs at all times. A DM who can cause a transition from the current state to some other state may choose to do so at any time; if several DMs have the opportunity to move, no assumption is made as to which one will seize the initiative. Likewise, if a DM has several possible moves, no assumption is made as to which one will be selected. (For example, in *Figure 1* all three DMs may move away from the status quo state, state 7; if Canada moves, it may choose whether to move to state 8 or state 9.)

A particular state is said to be *stable* for a particular DM if that DM would not strictly prefer to move away from it unilaterally, should the opportunity arise. A state is trivially stable for a DM who cannot move away from it. The main question of stability analysis is whether a DM who can move away from a state would choose *not* to do so.

A *stability definition* or *solution concept* is a procedure for answering this question. In principle, a stability definition is implemented by constructing an extensive-form game called a *departure game*, in which the first player to move is the DM in question and the initial options are to select each of the DM's available moves from the focal state or to stay at that state. Stability definitions differ as to how the departure game continues after an initial move to another state—for example, which opponent moves next? What are the opponents' payoffs? How many moves may the game may contain? Different stability definitions, reflected in the different rules for constructing these games, model different styles of strategic decision making, from conservative to manipulative, from low knowledge to high knowledge, and from low foresight to high foresight to infinite foresight. Each stability definition can be thought of as a sociological model of behavior in a strategic conflict; some of these models incorporate bounds to rationality, for example, by presuming that DMs lack knowledge of each other's preferences or by imposing a limited "horizon" of foresight.

All departure games are finite extensive form games of complete information. Game theory guarantees that such games have a well-defined solution—a subgame perfect equilibrium—which can be determined by backward induction. A state is defined to be stable for a particular DM under a particular stability definition if staying at that state is a subgame perfect equilibrium in the departure game constructed according to the rules defining the stability definition. (There are a few exceptions; for instance, stability with infinite foresight is determined by taking the limits of finite length games. Moreover, some stability definitions presuppose transitivity of preferences.)

A state that is stable for all DMs is in *equilibrium* and is usually interpreted as a possible resolution of the conflict. Should an equilibrium be attained, no DM would see any reason to move away from it. In practice, a model may have many equilibria; knowledge of this situation would be advantageous to a DM who could then move toward a preferred equilibrium. Of course, the situation is complicated by the fact that different DMs may be associated with different stability definitions. For this reason, stability information for a range of stability definitions is useful.

In fact, the usefulness of a range of stability information, recognized some time ago, was the main motivation for the development of computer systems to calculate stability. Both the Graph Model for Conflict Resolution [3] and GMCR II ([4], [5]) determine the stability of all states for all DMs under a range of stability definitions. GMCR II also incorporates an input module that enables users to enter most conflict models conveniently and an output module that assists them in interpreting the stability calculation. For example, *Figure 2* shows a screen of the Equilibria page of the GMCR II analysis of the graph model of *Figure 1*.

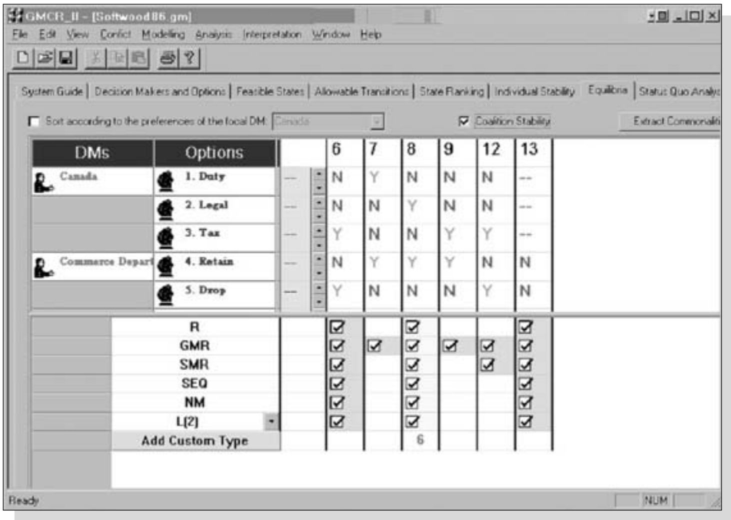

Fig. 2. Equilibria Screen of GMCR II analysis of graph model of *Figure 1*

The concept of unilateral improvement is useful in explaining some of the graph model stability definitions. A DM has a *unilateral improvement* from state s to state t if the DM controls the transition from s to t and the DM (strictly) prefers t to s. The stability definitions appearing in *Figure 2* are then as follows:

- R (Nash stability): The DM has no unilateral improvements from s.
- GMR (general meta-rationality): Any unilateral improvement for the DM from s can be *sanctioned* by other DMs (i.e., they can respond by moving to a state that the DM does not prefer to s).
- SMR (symmetric meta-rationality): A state is GMR, and furthermore the DM cannot respond to the sanction to achieve a state more preferred than s.
- SEQ (sequential stability): Same as GMR, except that the moves by the opponents that achieve any sanction must be unilateral improvements (for the mover).
- L(2) (limited-move stability, horizon 2): In a simple-length 2 departure game (one move by the DM followed by an optimal response by an opponent), the DM cannot guarantee that a state preferable than s is reached.
- NM (nonmyopic stability): In simple-length h departure games, for h large enough, the DM cannot reach a state that is more preferable than s.

For more details about these definitions and their properties see Fang *et al.* [3]. Note that different definitions presume different levels of foresight and different levels of information about other DMs. *Figure 2* shows that some states, for example, state 6, are stable for all DMs under all of the definitions shown, whereas other states, for example, state 7, are stable for all DMs under a few definitions (only GMR, in fact). States not mentioned in *Figure 2* are state 1 unstable for at least one DM under all stability definitions. Other output screens from GMCR II provide details as to the stability of every state for every DM under every stability definition.

The original concept for the Graph Model envisaged that the analyst would specify one stability definition for each DM. But recently a more sophisticated approach has evolved, facilitated by the availability of full stability information on all states, as exhibited in *Figure 2*. Somewhat informally, a state is said to be in a *strong equilibrium* if it is stable under most definitions for all DMs and in a *weak equilibrium* if it is stable under a few definitions for all DMs. The graph model of *Figure 1* is typical in that, as seen clearly in *Figure 2*, the strong equilibria—states 6, 8, and 13—are easy to distinguish from the weak equilibria. (It is not surprising that state 13 is stable for all DMs; as *Figure 1* shows, no DM can move away from it.) The first conclusion from the analysis is that, once any of the three strong equilibrium states is achieved, it will endure.

Because DMs' preferences over states are different, information about which states are likely to persist is strategically important. From *Figure 1*, it is easy to see that Canada can achieve state 8, a strong equilibrium, by a unilateral move from the status quo, state 7. (To confirm that state 8 is an equilibrium, note that no DM would prefer to move away from it: Canada prefers state 8 to state 9, Commerce prefers state 8 to state 11, and Industry prefers state 8 to state 2.) The immediate benefits of moving from state 7 to state 8 explains Canada's initial declaration of intent to fight Commerce's initial ruling. These stability properties explain why the conflict stabilized at state 8 for several months.

The graph model of *Figure 1* was selected for this presentation because it is simple and easy to understand, yet it illustrates many important points about the Graph Model for Conflict Resolution methodology. But even this small model would be tedious to analyze by hand. Fortunately, GMCR II has carefully designed data structures and efficient algorithms for calculating stability; it analyzed this model in well under a second. The largest model analyzed to date has approximately 100,000 states [8] and was analyzed within minutes on a computer.

4 Graph Model Stability Analysis: Benefits for Negotiators

Although the model of the Softwood Lumber Negotiation shown in *Figure 1* is very small, the analysis reported above does illustrate the kind of advice that the Graph Model might give to a decision maker about strategic choices during a negotiation. In October 1986 Canada moved rapidly to the equilibrium at state 8. Commerce was indifferent about this move; Industry would have preferred to avoid it, but the model suggests why there was no alternative. A move by Industry to state 1 would probably be followed by a move by Canada to state 2, which Industry finds less preferable than state 8. Thus what happened is understandable, and in this instance (unlike some others) the analysis does not suggest any alternative action that would have made some decision maker better-off. Often, knowledge of which states are likely to exhibit a high level of stability is the most important strategic advice at this stage.

The capacity of GMCR II to analyze graph models facilitates the use of the Graph Model for Conflict Resolution methodology not only before negotiations begin but also while they are ongoing. In response to developments during the negotiation process, a graph model built earlier can be updated rapidly to reflect new options and new assessments of preferences. The analysis then provides the negotiator with a new set of strategic insights into the updated model. "What-if?" questions can be explored not only prior to the negotiation but also during the process. Of course, the use of the model typically requires an expert, and as it is better to have one or more domain experts, it is unwise to attempt an analysis while seated at the negotiating table. The Graph Model is thus appropriate for ongoing negotiation support, in caucuses or during recesses in the process.

In summary, the Graph Model—especially when applied using the GMCR II system—is simple and fast. To some degree the answers it provides are definite, but there is scope for individual judgment and interpretation not only in input but also in applying analysis results to formulate a negotiation strategy. This is natural, inasmuch as the model itself reflects the information and assessments of the client, who is often a participant in the negotiation. Usually, it is easy to adjust and re-analyze a model in order to assess new information or revised judgments. In addition, if there is doubt about some components of the model, sensitivity analysis is feasible and provides useful information about exactly what input is crucial to the analysis results and therefore to the strategic conclusions.

Admittedly, the Graph Model for Conflict Resolution methodology has some disadvantages, most of which can be traced to the simplicity of the model itself. One issue is that questions involving allocation in a continuum are awkward in a graph model. They can be handled approximately by discretizing the model, using a few possibilities to approximate the possible allocations. Another disadvantage is the difficulty of explicitly incorporating partial information and beliefs into a graph model. The use of sensitivity analysis certainly helps, but the ultimate choice must depend on the client's judgment.

It might be possible to develop the graph model in ways that would permit it to address these problems or to incorporate features of real-world problems that seem to be missing. Among the latter are transaction costs (perhaps a DM might be deterred from moving by the costs of forcing a transition) and history (the preference for a state may depend on the other states that have been visited recently). These measures have not been adopted because to do so would serve to cancel out the simplicity of the Graph Model, which is probably its greatest strength. Experience has shown that the Graph Model provides valuable strategic insights with surprisingly little input. To deal directly with continuous allocations, for example, or to incorporate transaction costs, would require not ordinal preferences but cardinal utilities (for all DMs), and cardinal preference information is much more difficult to obtain than ordinal preference information. So far, the developers have chosen to preserve the simplicity of the model and work around the problems.

5 Conclusions: Recent Developments

While some directions for development of the Graph Model have been considered and rejected, others have been pursued. Most of the new developments increase or extend the applicability of the methodology without changing its information requirements. The three features that will be described here briefly are Coalition Analysis, Status Quo Analysis, and new options for preference representation.

Coalition Analysis is an algorithm that can be applied to a graph model after the standard stability analysis is complete. It assesses states that are equilibria or strong equilibria (in the Softwood Lumber model, states 6, 8, and 13—see *Figure 2*) for vulnerability to coalitions of the DMs. To elaborate, individual decision makers are not motivated to move away from equilibria, by definition; those who can move foresee that a unilateral move away from the state would not result in a more favorable state. But sometimes a group of DMs, by making several moves in succession, can achieve a state that is more preferred by every DM in the group. (For examples and discussion, see [13].)

Typically, such coalitional moves involve some but not all of the DMs. But one involving all DMs occurred in the Softwood Lumber negotiation modeled in *Figure 1*. In October 1986 the conflict quickly reached the strong equilibrium at state 8, as described above. But as the clock ticked toward midnight on 31 December of that year, the three DMs reached an agreement that effectively shifted the conflict to the strong equilibrium at state 6. Note from *Figure 1* that this shift requires a coordinated sequence of moves by all three DMs. Moreover, no individual move away from state 8 is in the interest of the mover, but the combination of moves to state 6 in fact benefits all DMs, as is clear from the preference orderings of *Figure 1*. Simply put, all DMs are better off at strong equilibrium state 6 than at strong equilibrium state 8; acting in sequence, they achieved this transition.

GMCR II includes a primitive Coalition Analysis algorithm, which searches for equilibria that could be upset in this way by any coalition or a subset of two or more DMs. Two conditions define a coalitionally unstable equilibrium: first, there is another equilibrium that is strictly preferred by all members of a coalition of two or more DMs; second, the members of the coalition, by coordinating their actions, can move from the original equilibrium to the mutually preferred equilibrium. As shown in *Figure 2*, this algorithm can be invoked by clicking on Coalition Stability on the Equilibria page of GMCR II. Coalition Analysis augments the Graph Model methodology by providing information on the long-term stability of equilibria.

Status Quo Analysis is a new feature that adds a dynamic dimension to the analysis of a graph model. The idea is to begin with a known Status Quo state and to trace and assess the possible paths from that state. For instance, in the discussion of the Softwood Lumber model of *Figure 1*, it was noted that from the Status Quo state 7, state 8 is reachable in one step (a so-called *unilateral move* by the DM Canada). Moreover, this move is in the interest of the DM, so it is a unilateral improvement. To reach state 6, three unilateral moves are required, and at least one of them is not a unilateral improvement. Furthermore, if the status quo had been state 1, the strong equilibrium at state 8 would be unattainable. A simple version of this Status Quo

Analysis, until now carried out heuristically and by hand, was developed recently but has not yet been programmed [14].

Status Quo Analysis involves algorithms to draw a directed graph that shows reachability (in the graph-theoretic sense (see [19]) from the status quo state, based on unilateral moves or unilateral improvements. Summary tables showing the minimum path length to each achievable state are constructed. The algorithms can be simplified when every DM's graph is transitive, an important consideration in the Graph Model methodology, as DMs are usually forbidden from making two moves in succession. Planned extensions of these algorithms will be based on paths which include a limited number of unilateral moves that are not unilateral improvements. Of course, methods for incorporating the Status Quo Analysis algorithms into GMCR II are now under study.

A third area of development of the Graph Model for Conflict Resolution methodology is the definition and analysis of new options for expressing preferences. There are two reasons for adding new preference options—more information may be available, or less. With incomplete preference information, the objective is to carry out a partial Graph Model analysis in instances when an outcome has not yet been fully defined or assessed by a DM or when information about the DM's preference is not available. Some ideas on how to make a preliminary, albeit coarse, analysis of a graph model when some preference information is lacking were recently put forward (see [15]).

Another idea currently under exploration is the extension of the Graph Model when the "preferred" relationship has been refined into "strongly preferred" or "mildly preferred." Then, some of the standard stability definitions can be extended to define extreme or mild stability. Methods of using information about strength of preference, if available, to infer whether stability is extreme or mild have been developed (see [7]).

Finally, we return to the question of aid to negotiators. The Graph Model methodology can provide a negotiator with simple, flexible analyses of a negotiation problem, from the negotiator's own point of view. The analysis is precise and logically consistent because it is based on mathematical models and, in particular, game theory. On the other hand, the Graph Model is different from Game Theory in that it captures many aspects of open, free-flowing processes, like negotiation, while avoiding the strong assumptions about information that have been criticized as rendering Game Theory inapplicable to negotiation. Of course there are costs—the Graph Model has nothing to say about creativity (What model does?) and it is awkward in its approach to many problems of division. Some of these issues are explored in greater depth in a recent special issue of the journal *Group Decision and Negotiation* (see [9]).

From the point of view of support for negotiators, the contribution of the Graph Model for Conflict Resolution can be expressed very simply: it can provide a negotiator with a strategy, and a negotiator without a strategy has no direction. For this reason, more than any other, negotiators will benefit from the Graph Model's combination of ease of use and insightful strategic analysis.

References

1. CAG (2006) Conflict analysis group. Web site available at http://www.systems.uwaterloo.ca/Research/CAG/
2. DAS (2002) Decision analysis society, survey of decision tree and influence diagram. Software available at http://faculty.fuqua.duke.edu/daweb/dasw6.htm
3. Fang L, Hipel KW, Kilgour DM (1993) Interactive decision making: the graph model for conflict resolution. John Wiley and Sons, New York, USA
4. Fang L, Hipel KW, Kilgour DM, Peng X (2003a) A decision support system for interactive decision making, Part 1: Model formulation. IEEE Transactions on Systems, Man and Cybernetics Part C 33(1):42–55
5. Fang L, Hipel KW, Kilgour DM, Peng X (2003b) A decision support system for interactive decision making, Part 2: Analysis and output interpretation. IEEE Transactions on Systems, Man and Cybernetics Part C 33(1):56–66
6. Fisher R, Ury W, with Patton B (1992) Getting to yes: negotiating agreement without giving in. 2nd ed., Penguin, New York, USA
7. Hamouda L, Kilgour DM, Hipel KW (2004) Strength of preference in the graph model for conflict resolution. Group Decision and Negotiation 13:449–462
8. Hipel KW, Kilgour DM, Fang L, Peng X (2001) Strategic decision support for the services industry. IEEE Transactions on Engineering Management 48(3):358–369
9. Kilgour DM, Hipel KW (2006) Introduction to the special issue on the graph model for conflict resolution. Group Decision and Negotiation (to appear)
10. Kilgour DM, Hipel KW, Fang L (1987) The graph model for conflicts, Automatic 23(1)41–55
11. Kilgour DM, Hipel KW, Fang L (1994) Negotiation support using the graph model for conflict resolution. Group Decision and Negotiation 3(1):29–46
12. Kilgour DM, Fang L, Hipel KW (1995) GMCR in negotiations, Negotiation Journal 11(2):151–156
13. Kilgour DM, Hipel KW, Peng X, Fang L (2001) Coalition analysis in group decision Support. Group Decision and Negotiation 10:159–175
14. Li KW, Kilgour DM, Hipel KW (2005) Status quo analysis in the graph model for conflict resolution. Journal of the Operational Research Society 56 6:699–707
15. Li KW, Hipel KW, Kilgour DM, Fang L (2004) Preference uncertainty in the graph model for conflict resolution. IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans 34(4):507–520
16. Raiffa H (1982, 2nd ed. 2003) The art and science of negotiation. Harvard University Press, Cambridge, MA, USA
17. Shell GR (1999) Bargaining for Advantage. Penguin, New York, USA
18. Starke K, Rangaswamy A (2000) Computer-mediated negotiations: review and research opportunities. Encyclopedia of Microcomputers, Vol. 26, Marcel Dekker, New York, USA
19. West DB (2003) Introduction to graph theory. 2nd ed. Prentice-Hall, Upper Saddle River, NJ, USA

A Minimax Procedure for Negotiating Multilateral Treaties

Steven J. Brams[1], D. Marc Kilgour[2], and M. Remzi Sanver[3]

[1] Department of Politics, New York University, New York, NY, USA
[2] Department of Mathematics, Wilfrid Laurier University, Waterloo, Ontario, Canada
[3] Department of Economics, Istanbul Bilgi University, Istanbul, Turkey

1 Introduction

We propose in this chapter a procedure for reaching agreement on multilateral treaties that produces a compromise as close as possible to the preferences of all parties. By "close" we mean that the maximum distance of the compromise from the position of any state is minimal, which we call a *minimax outcome*. We show that this procedure is relatively invulnerable to strategizing by states, reducing any incentive they might have to misrepresent their preferences to try to induce a better outcome.[4]

The procedure is different from the usual method for reaching an agreement in multilateral treaty negotiations. Normally states vote separately on each provision of a treaty, often starting from a "single negotiating text" (Sebenius [15] discusses extensive use of such a text in the Law of the Sea negotiations). A provision is included in the treaty if a simple or qualified majority—or sometimes all states—supports it; otherwise, it is excluded. If provisions are included in a treaty that are not approved by all states, then those states that do not get their way on provisions they consider important might be quite frustrated—and unwilling either to sign the treaty or to abide by its terms.

Thereby a large majority of states may agree to and achieve a treaty far removed from the ideal of one or more states. By contrast, if the votes for and against every provision are aggregated in a different way, a better compromise—one that leaves no state too aggrieved—may be found.

Our focus is in aggregating votes, which is appropriate when negotiations occur in the shadow of voting. The parties to a treaty, when deciding on what provisions to propose, must consider how those provisions will be viewed when it comes time to vote on them. Will they receive sufficient support to be accepted by most if not all

[4] This chapter is a revised version of Brams *et al.* [6]. The minimax procedure has also been analyzed in two other papers. Brams *et al.* [7] apply it to the selection of committees but alter the procedure discussed here by weighting the number of voters who cast identical ballots. Kilgour *et al.* [12] provide an efficient algorithm for making minimax calculations and propose a different weighting scheme that downgrades the influence of extreme voters.

parties in the negotiation? Even if a provision is found acceptable, will its inclusion facilitate or retard passage of the treaty in a referendum, or a parliamentary vote, in those countries that must ratify it?

Clearly, voting—or, more accurately, its anticipation—is part and parcel of negotiating a treaty, especially in influencing what gets put in and what gets left out, as well as how the treaty provisions are packaged. Although we do not model the crafting and packaging of these provisions, the fact that the minimax procedure we propose does not hurt any party as much, potentially, as the usual voting procedure suggests that it would allow the parties to be more open in their negotiations. We have no evidence to prove this, but we do present evidence that a 1954 environmental treaty might have fared better if the positions of countries had been aggregated under the minimax procedure rather than the usual procedure.

To begin the analysis, we suppose that negotiations have reached a stage whereby the provisions of a treaty in dispute (i) can be specified, (ii) are of approximately equal significance to all states, and (iii) are relatively independent of each other. While states of very different size, wealth, military capability, and the like can be weighted (perhaps by monetary contributions, as in the International Monetary Fund (IMF)), the procedure we propose does not directly reflect such differences.

An approach for addressing these differences is to group states into regional or functional blocs that are more or less equal in size, possess substantial common interests, and have a similar stake in the outcome of the treaty negotiations. Putting together such blocs, however, is not straightforward, as we discuss later.

The procedure we propose for forging consensus in multilateral treaty negotiations is based on "fallback bargaining" [2], but we modify it in an important way. Instead of assuming that states directly rank alternatives, we represent their preferences by vectors of 1s and 0s, where a "1" indicates approval and a "0" indicates disapproval of each proposed provision of a treaty.

We assume that all positions of states, given by vectors of 1s and 0s, are possible. Each state supports ("votes for") a specific vector, which we call that state's *top preference*, or ideal point. Below a state's top preference, its preference for other vectors depends on their distance from its ideal: its seond preference is any of the vectors that differ from its top preference in one component, then in two, and so on, down to the vector that differs from the top preference in every component. Thereby each state ranks vectors, based on their proximity to its ideal point.

For example, assume that only two provisions of a treaty are at issue and a state approves of the first provision and disapproves of the second. Then its top preference is 10; its ranking in our model is

$$10 > \{11, 00\} > 01.$$

That is, the state most prefers 10, next most prefers either 11 or 00—between which it is indifferent—and least prefers 01.

This chapter's plan is as follows: Part 2 describes and illustrates fallback bargaining, showing that an outcome may be highly dependent on the decision rule used. In one example we show that every possible outcome may be selected under some fallback *decision rule* r which may range from 1 to n.

We focus on *fallback bargaining with unanimity* (FB_n), comparing it with the common procedure of majority voting (*MV*) on each provision. We show that the minimax outcomes under FB_n may differ from *MV* outcomes, which we call *minisum outcomes* because they minimize the sum of the distances (or average distance) to players' top preferences. Whereas FB_n outcomes are in the Rawlsian tradition of minimizing the largest deviations from a settlement, *MV* outcomes are in the utilitarian tradition of minimizing average departures.

Part 3 compares minimax and minisum outcomes, asking which is better and under what conditions. We introduce the notion of a "weighted" minimax outcome to take into account the fact that the weight or the number of players having the same top preference may vary if negotiation is not among equals.

Part 4 analyzes the manipulability of *MV* and FB_n, showing that FB_n is vulnerable to manipulation; this is also true if the fallback decision rule is not unanimity. By contrast, *MV* is invulnerable—players always have an incentive to be truthful. But in any realistic situation with incomplete information about the preferences of two or more players, FB_n would be extremely difficult to manipulate.

FB_n outcomes seem superior in situations like the one discussed in Part 5, in which 32 states negotiated oil-pollution controls in 1954. In these negotiations, states could abstain as well as vote yes or no on treaty provisions. To break ties among the six FB_n outcomes, we suggest approval voting, which allows states to approve of one or more of the FB_n outcomes and which is quite resistant to strategic exploitation.

Part 5 concludes by suggesting that FB_n may well facilitate consensus in multilateral treaty negotiations, especially those that include most of the nearly 200 states in the world today. These complex negotiations frequently involve both individual states and overlapping blocs of states, scores of provisions, and considerable maneuvering by the players to try to achieve a strategic advantage. We believe that our proposed procedure would encourage players to be honest, render their negotiations more open, and make the compromises they achieve as acceptable as possible to all players.

2 Fallback Bargaining under Different Decision Rules

Assume there are k provisions of a treaty being negotiated by n players (countries). A possible treaty is a binary k-vector, $(p_1, p_2, \ldots, p_k)$, where p_i equals 0 or 1. Such binary vectors will be called *combinations*. To simplify notation, we write combinations such as (1, 1, 0) as 110. Note that the total number of combinations is 2^k.

A country's *top preference* is its most-preferred combination. We assume countries rank treaties according to their "Hamming distance" from their top preferences. The *Hamming distance*, $d(p, q)$, between two binary k-vectors, p and q, is the number of components on which they differ. For example, if $k = 3$ and a player's top preference is 110, the distances between it and the eight binary 3-vectors (including itself) are shown below:

Top preference	$d = 0$	$d = 1$	$d = 2$	$d = 3$
110	110	100	000	001
		010	101	
		111	011	

Note that three vectors are tied for second place at distance $d = 1$ and three more are tied for third place at $d = 2$. In general, the preference ordering is not strict because of such ties.

To find a consensus choice, we next describe fallback bargaining for some decision rule $r \leq n$: (Technically, Brams and Kilgour [2] define fallback bargaining as occurring only when preferences form a linear order, which assumes that players have strict preferences. But it is easy to extend this procedure to nonstrict preferences, as we do here.)

1. Assume countries approve only of combinations at distance $d = 0$ from their top preferences—that is, only the top-preference combinations are acceptable. If one or more combinations are approved of by at least r countries, then those with the most approvals are winning, and the process stops.
2. If no combination is approved of by at least r countries at distance $d = 0$, then consider all combinations at distance $d \geq 1$ from the players' top preferences. If one or more combinations are approved of by at least r countries, then those with the most approvals are winning, and the process stops.
3. If no combination is approved of by at least r countries at distance $d \geq 1$, continue the descent until, *for the first time*, one or more combinations are approved of by at least r countries at distance $d \geq 2$. The combinations with the most approvals are winning, and the process stops.

Exactly when the process stops depends on, among other things, the number of countries n, the number of provisions k, and the decision rule r [2]. Our first result characterizes winning outcomes under FB_n.

Proposition 1. *The FB_n winners are the minimax outcomes—they minimize the maximum distance to the top preference of any player.*

Proof. See Brams and Kilgour ([2], p 292, Theorem 3).

The idea behind the proof is the following. Suppose an FB_n outcome is not a minimax combination. Then there is some other combination for which the maximum distance to the top preference of any player is less. But the descent under FB_n, which stops at the *first* point at which all players approve of some combination, must stop at this other combination. Therefore, this other combination must be a minimax combination.

A decision rule of unanimity is frequently used to decide important questions, such as the admission of new members into a regional or international organization.

As a case in point, the Treaty of Rome in 1958 made unanimous consent of the original six-member Common Market a requirement for admission of new members; that rule is still in effect in the present 25-member European Union that has evolved from the Common Market over the last 50 years.

That different fallback decision rules can give dramatically different results is illustrated by *Example A*, in which there are $n = 10$ players and $k = 3$ provisions, which yield 8 combinations. Note that four combinations (000, 100, 010, 001) are the top preferences of one player each, and three combinations (110, 101, 011) are the top preferences of two players each; only combination 111 is nobody's top preference. Geometrically, the positions of the different players are shown as points in three-dimensional space in *Figure 1*.

Example A. Number of players approving each combination at different distances

Combination	$d = 0$	$d = 1$	$d = 2$	$d = 3$
1. 000	1	4	10***	10
2. 100	1	6**	8	10
3. 010	1	6**	8	10
4. 001	1	6**	8	10
5. 110	2*	4	9	10
6. 101	2*	4	9	10
7. 011	2*	4	9	10
8. 111	0	6*	9	10
Total	10	40	70	80

* Winner for $r = 1$ and 2.
** Winner for $r = 3, 4, 5$, and 6 (simple majority).
***Winner for $r = 7, 8, 9$, and 10 (unanimity).

We use *Example A* to prove our next proposition.

Proposition 2. *It is possible for every combination to win, or tie for winning, under some decision rule r.*

Proof. The number of players that approve of each combination at distances $d = 0, 1, 2$, and 3 are shown in *Example A*. At $d = 0$, this number is simply the number of players whose top preference is that combination. The three combinations that are approved of by 2 players (110, 101, 011) are the winners under decision rules $r = 1$ and $r = 2$: for $r = 1$, they get the most votes (2); for $r = 2$, they are the only combinations that get 2 votes.

It is easy to verify that four combinations (100, 010, 001, 111) are the winners at distance $d = 1$ under decision rules $r = 3, 4, 5$, and 6; and one combination (000) is the winner at distance $d = 2$ under decision rules $r = 7, 8, 9$, and 10.

Clearly, all players must approve of all combinations at $d = k$, which is $d = 3$ in *Example A*. But because all players find combination 000 acceptable at $d = 2$ in this example, it is the unanimity winner ($r = 10$). In sum, every combination is a

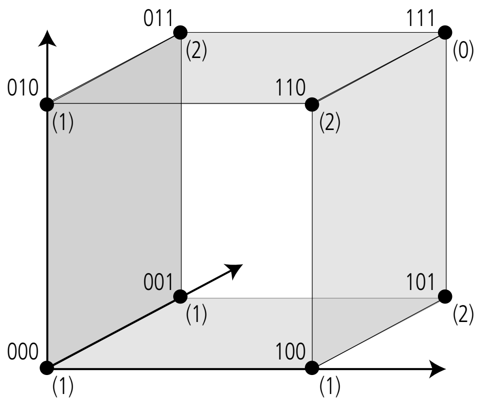

Fig. 1. Positions of $n = 10$ players on $k = 3$ provisions (*Example A*)
Note: The numbers in parenthesis give the number of players supporting the corresponding top preference.

winner, or tied for winning, under some decision rule r. QED.

Because the winning combination under fallback with unanimity (*FB*$_n$), 000, is the only combination that is a distance of 2 or less from the top preference of every player, it would seem a reasonable compromise. In *Figure 1*, this position is a distance of $d = 3$ from combination 111, but because 111 has no players supporting it, combination 000 is a maximum distance of $d = 2$ from all players.

Majority voting (*MV*) on each provision is probably the procedure most often used to select a winning combination. As we next demonstrate, the outcome it produces may not coincide with the *FB*$_n$ winner.

Proposition 3. *FB*$_n$ *may produce a unique outcome that is different from an MV outcome and is no player's top preference.*

Proof. In *Example B*, there are $n = 5$ players and $k = 4$ provisions, which yield 16 combinations, whose numbers we show in parentheses following the combination. Combination 0010 (4) is the unique winner under *FB*$_n$ at $d = 2$. But this is not the top preference of any of the 5 players, whose top preferences are combinations 1000 (2), 0001 (5), 0110 (9), 0011 (11), and 0111 (15) (shown in boldface in *Example B*).

It can be verified directly that a majority of the players prefer 0 on the first two components, and 1 on the last two components, so combination 0011 (11) is the *MV* winner. Hence, a unique *FB*$_n$ winner may not be the top preference of any player and also may differ from an *MV* winner. QED.

Because the *FB*$_n$ winner, combination 0010 (4), is selected at $d = 2$, its maximum distance from all players' top preferences is 2. By contrast, the *MV* winner, combination 0011 (11), is $d = 3$ from combination 1000 (2)—the top preference of

Example B. Number of players approving each combination at different distances (top preferences of five players in boldface)

Combination	$d = 0$	$d = 1$	$d = 2$	$d = 3$	$d = 4$
1. 0000	0	2	4	5	5
2. 1000	**1**	**1**	**2**	**4**	**5**
3. 0100	0	1	4	5	5
4. 0010	0	2	5*	5	5
5. 0001	**1**	**2**	**4**	**5**	**5**
6. 1100	0	1	2	4	5
7. 1010	0	1	3	5	5
8. 1001	0	2	3	4	5
9. 0110	**1**	**2**	**3**	**5**	**5**
10. 0101	0	2	4	5	5
11. 0011	**1**	**3**	**4**	**5**	**5**
12. 1110	0	1	3	4	5
13. 1101	0	1	4	5	5
14. 1011	0	0	3	5	5
15. 0111	**1**	**3**	**4**	**4**	**5**
16. 1111	0	1	3	5	5
Total	5	25	55	75	80

*FB_n winner; the *MV* winner is combination 0011 (11)

one of the players—so it is not as close to *all* players as the FB_n winner, which is what distinguishes FB_n winners from *MV* winners, as we will discuss in more detail shortly.

In *Example A*, FB_n (uniquely) gives combination 000, whereas *MV* gives all 8 combinations (because positions of 0 and 1 tie with 4 votes each on all 3 provisions). In this case, some tie-breaking mechanism would have to be used to select a specific winning combination. Observe that the distances of the five players from FB_n winner 0010 (4) are (2, 2, 1, 1, 2); the maximum distance is 2 and the sum of the distances is 8. In contrast, the distances from *MV* winner 0011 (11) are (3, 1, 2, 0, 1); the maximum distance is 3 and the sum of the distances is 7. As shown in Proposition 1, FB_n minimizes the maximum distance; as shown next, *MV* minimizes the sum of distances.

When FB_n and *MV* give different outcomes, does one yield a "better" compromise than the other? Consider again *Example B*, wherein we noted that FB_n gives combination 0010 (4) and *MV* gives combination 0011 (11).

Proposition 4. *The MV winners are the minisum outcomes—they minimize the sum of the distances to the top preferences of all players.*

Proof. Assume that there are k provisions and n players, and that player i supports the binary k-vector $p^i = (p_1^i, p_2^i, \ldots, p_k^i)$. For an arbitrary binary k-vector $x = (x_1, x_2, \ldots, x_k)$, define

$$d_j(x, p^i) = \begin{cases} 0 \text{ if } x_j = p_j{}^i \\ 1 \text{ if } x_j \neq p_j{}^i \end{cases}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$. Then it is clear that the Hamming distance from x to p^i is given by $d(x, p^i) = \sum_{j=1}^{k} d_j(x, p^i)$. We consider how to select x so as to minimize $D(x) = \sum_{i=1}^{n} d(x, p^i)$.

For any x and j, define $S_j(x) = \sum_{i=1}^{n} d_j(x, p^i)$. Note that $D(x) = \sum_{j=1}^{k} S_j(x)$ and that $S_j(x)$ depends only on x_j and not on the other $k - 1$ components of x. From the definition of $d_j(x, p^i)$ it follows that:

$$S_j(x) = \begin{cases} N_j(0) \text{ if } x_j = 1 \\ N_j(1) \text{ if } x_j = 0, \end{cases}$$

where $N_j(t) = |\{i : p_j{}^i = t\}|$ for $t = 0$ and 1. It also follows that x minimizes $S_j(x)$ iff (if and only if) $x_j = 1$ when $N_j(0) < N_j(1)$ and $x_j = 0$ when $N_j(0) > N_j(1)$. [Note that when $N_j(0) = N_j(1)$, there is no condition on x_j.] Consequently, x minimizes $D(x)$ iff

$$x_j = \begin{cases} 1 \text{ if } N_j(1) > N_j(0) \\ 0 \text{ if } N_j(0) > N_j(1) \end{cases}$$

for all $j = 1, 2, \ldots, k$, which is true iff x is an *MV* outcome. QED.

The idea underlying the proof of Proposition 4 is that the *MV* outcome minimizes the number of disagreements of the players on each provision, because at least as many players agree as disagree with the *MV* outcome on each provision. Because the sum of disagreements across the k provisions equals their sum across all the players, the latter—which is the total distance of all players from their top preferences (Hamming distance)—is also minimized.

3 Which Minimization Criterion Is Best?

Is it better to minimize (i) the sum of the distances from a compromise outcome (*MV* winner) or (ii) the maximum distance of players from a compromise outcome (FB_n winner), given there is a conflict? If the goal is to avoid antagonizing any player "too much," there are good grounds for choosing FB_n. In *Example B*, the FB_n winner is a maximum distance of 2 from the top preferences of the 5 players, whereas the *MV* winner is a distance of 3 from one player's top preference.

Consider how these winning outcomes would (or would not) change if the players were differently weighted. Suppose, for example, that the player supporting combination 1000 (2) were given a weight of 5, or were replaced by 5 players that all have combination 1000 (2) as their top preference. This combination would become the *MV* outcome—on each provision. It would get majority approval—but the top preference of the player supporting combination 0111 (15) is a distance d of 4 from it [because combination 0111 (15) is the *antipodal combination*—the opposite on every component—of combination 1000 (2)]. (Sufficient conditions under which minimax and minisum outcomes may be antipodal are given in Brams *et al.* [6].)

By contrast, the FB_n winner would still be combination 0010 (4), which remains at a distance d of at most 2 from each player's top preference. However, now the sum of distances of each player from combination 1000 (2) is

$$5(2) + 2 + 1 + 1 + 2 = 16,$$

whereas the sum of distances from *MV* combination 1000 (2) is

$$5(0) + 2 + 3 + 3 + 4 = 11,$$

which is considerably less, so combination 1000 (2) becomes the *MV* winner.

Patently, the players supporting combination 1000 (2) "call the shots" under *MV*. By implementing their top preference, they leave the minority—the other 4 players—at distances d of 2 to 4 from their top preferences.

Because FB_n depends only on *which* k-vectors are the top preferences of one or more players—not the numbers that support each—it is insensitive to weights (sometimes referred to as independence of "clones," or other players with the same top preference). This property suggests that FB_n would work best for *negotiation among equals*, which might require grouping countries into blocs.

This is decidedly not the case in international forums such as the General Assembly of the United Nations. But should a microstate have the same weight as the USA or China in international negotiations? Requiring states to coalesce in blocs that have more or less equal weight and take a common position is one way to ensure negotiation among equals. In fact, the 184 countries in the International Monetary Fund (IMF) are grouped into 24 so-called Constituencies—with from one member (eight largest contributors) to 25 members (all African states)—each having a weight equal to the total weight of its members. Each Constituency is represented by an Executive Director, who casts the Constituency's votes as a bloc on an Executive Board that, in principle, reflects a consensus among the Constituency's members. The decision rules used are 50 percent, 70 percent, and 85 percent, depending on the importance of the issue being decided. FB_n might also be used within a bloc to find a compromise position that best reflects the views of the bloc's members.

If negotiation is not among equals, then it is possible to amend FB_n to take into account the different numbers or different weights of players supporting each combination. As an illustration, consider *Example A*, in which three combinations (5, 6, and 7) are supported by two players each, four combinations (1, 2, 3, and 4) by one player each, and one combination (8) by no player. Let the speed of descent from a top preference be inversely proportional to the number of players who support it, as long as this number is not zero.

Thus, in *Example A*, the two players each that support combinations 5, 6, and 7 will consider acceptable combinations at $d = 1$ exactly when the one player each that supports combinations 1, 2, 3, and 4 consider acceptable combinations at $d = 2$. In effect, each of the former combinations is the top preference of twice as many players as the latter combinations, producing twice as much "inertia" in the descent. Accordingly, the winning combination may be only half as distant from the top preferences of these players as from the top preferences of the players whose top preference is supported only by themselves. It is easy to see that the first combination to become acceptable to all 10 players under these rules is 111 (combination 8), which is $d = 1$ from the 6 players that approve of two provisions each and $d = 2$ from the 4 players that approve of one provision each.

Note that combination 111 is the antipode of the minimax outcome 000. Combination 111 might be considered a *weighted minimax outcome* in the sense that it minimizes the maximum weighted distance from all players, where weight in this case reflects the numbers of players who make each combination their top preference. Because there is not a consensus in most treaty negotiations on how players should be weighted, we will not try to apply the notion of weighting here. For more on the weighting of minimax outcomes, see [7] and [12].

Many international disputes are between two countries or two sets of countries. For example, the EU now negotiates as a collectivity with the USA. In such two-player disputes, wherein the obvious compromise is to settle about half of the disagreements in favor of each of the two players, *FB*$_n$ is hardly needed. On the other hand, if the disagreements concern issues that the two sides care differently about, a point-allocation procedure called "Adjusted Winner" seems a better mechanism to use [4][5].

4 Manipulability

A bargaining procedure is *manipulable* if a player, by misrepresenting its preferences, can obtain a preferred outcome. To determine what is "preferred," we must take account of preferences between sets. We assume that if a is preferred to b_1 and a is preferred to b_2, then a is preferred to $\{a, b_1, b_2\}$.

Because both *MV* and *FB*$_n$ require that players support a specific binary k-vector representing their top preferences, the manipulability of these procedures depends on whether the players can benefit by indicating a top preference different from their sincere one.

Proposition 5. *FB*$_n$ *is manipulable, whereas MV is not.*

Proof. We start with *MV*. As the players' choices are binary on each provision, it is always in a player's interest to support whichever position (0 or 1) it prefers. Since the majority choice on each provision has no effect on the majority choices on other provisions, each player has a dominant strategy of voting sincerely on all provisions.

Now consider *FB*$_n$. In *Example C*, there are $n = 4$ players and $k = 4$ provisions, which yield 16 combinations, but not all are shown. For the "true top preferences" of the 4 players in *Example Ca*, there are three *FB*$_n$ winners (combinations 1000 [1], 0000 [5], and 1001 [6]), each of which gets unanimous approval at distance $d = 2$.

If the player supporting combination 1000 (1) in *Example Ca* misrepresents its top preference as combination 1110 (5) in *Example Cb*, then the *unique* winner is combination 1000 (1), which is in fact this player's true top preference (and which also happens to be the *MV* winner). Thus, by indicating an insincere top preference, this player can induce a preferred outcome in *Example Cb*, based on our assumption

Example C. Number of players approving each combination at different distances

Ca: True top preferences

Combination	$d = 0$	$d = 1$	$d = 2$
1. 1000	1	3**	4*
2. 0001	1	1	2
3. 1100	1	2	3
4. 1010	1	2	3
5. 0000	0	2	4*
6. 1001	0	2	4*
10 Other combinations	0	At most 2	At most 3

Cb: Player 1 above misrepresents as player 5 below

Combination	$d = 0$	$d = 1$	$d = 2$
1. 1000	0	2	4*
2. 0001	1	1	1
3. 1100	1	2	3
4. 1010	1	2	3
5. 1110	1	3**	3
11 Other combinations	0	At most 1	At most 3

* FB_n winner.

**FB_m winner.

about preferences between sets (given at the beginning of this part). QED.

In theory, therefore, FB_n is vulnerable to manipulation. In practice, however, FB_n is probably almost as resilient to manipulation as *MV* because its possible exploitation would require that a manipulative player have virtually complete information about the preferences of other players—and that they all act sincerely—which is unlikely in most real-life situations.

The resilience of FB_n as well as *MV* to manipulation would seem to negate the advantages that accrue to strategic players in negotiation with naive players. In fact, even for strategic players, there may be no gain from manipulation.

As an illustration, let us assume that the fallback decision rule in *Example C* is a simple majority (3 of the 4 players), which we indicate by FB_m rather than unanimity. As shown in *Example Ca*, the FB_m outcome is combination 1000 (1). It is not difficult to verify that no player can induce a preferred outcome under FB_m by misrepresenting its top preference.

But manipulation is certainly possible under FB_m.

Proposition 6. *If the decision rule is not unanimity, fallback bargaining is manipulable.*

Proof. If the decision rule is not unanimity, we prove this proposition by giving an example that shows that fallback bargaining is manipulable under FB_m. In *Example Da*, observe that 5 players choose 4 different combinations. Combination 0000

(1) receives the approval of 3 of the 5 players at distance $d = 1$ and is the unique winner under FB_m.

Example D. Number of players approving each combination at different distances

Da: True top preferences

Combination	$d = 0$	$d = 1$
1. 0000	1	3*
2. 1000	1	2
3. 0100	1	2
4. 1111	2	2
12 Other combinations	0	At most 2

Db: 2 Player 4's above misrepresent as 2 player 5's below

Combination	$d = 0$	$d = 1$
1. 0000	1	3
2. 1000	1	4*
3. 0100	1	4*
4. 1111	0	0
5. 1100	2	4*
11 Other combinations	0	At most 2

*FB_m winner.

Note that the most dissatisfied players will be the 2 players whose top preference is the antipodal combination, 1111 (4). If these 2 players misrepresent their top preference as 1100 (5) in *Example Db*, they induce combinations 1000 (2), 0100 (3), and 1100 (5) as the winners, each of which receives the approval of 4 players at distance $d = 1$. (Because combination 0000 (1) receives approval from 3 players, four combinations listed in *Example Db* receive approval from a majority of players. However, we assume that the FB_m winners are those combinations—namely, the three combinations, 1000 (2), 0100 (3), and 1100 (5)—that receive the most approvals.) Because these three winning combinations are closer to the true top preference of the 2 dissembling players [combination 1111 (4) in Da] than is combination 0000 (1), their misrepresentation is rational, rendering FB_m manipulable. QED.

FB_m is called the "majoritarian compromise" in Hurwicz and Sertel [11], Sertel and Sanver [16], and Sertel and Yilmaz [17], wherein it is analyzed as a voting procedure with majority rule rather than a bargaining procedure, as in Brams and Kilgour [2]. But the fallback process is essentially the same under both interpretations, whether the decision rule is simple majority, qualified majority, or unanimity.

We next turn to a real-life dispute that suggests how FB_n might be used in practice. It raises new issues (the possibility of abstention; ties among FB_n outcomes) for which we suggest some pragmatic solutions.

5 An Application to an Environmental Dispute

On 25 April 1954, a three-week conference of 32 states, representing 95 percent of the world's shipping tonnage, convened in London. It included 18 developed states from Europe, North America, and Australasia; of the others, three were from Eastern Europe, four from Asia, six from Latin America, and one from Africa.

The treaty that resulted from this conference, the International Convention for the Prevention of Oil Pollution (OILPOL '54), prohibited the discharge of oil from certain ships in specified ocean areas. It came into force in 1958 and was followed by three more conferences between 1958 and 1962 that resulted in the adoption of additional measures to strengthen those in OILPOL '54.

The principal question on the table concerned how best to prevent oil pollution of the sea from discharges from both tankers and nontankers. The account that follows, and the voting of the 28 states shown in *Example E* (no data were given for 4 states) is taken from M'Gonigle and Zacher [13], pp 85–91. A detailed discussion of compliance with this and subsequent oil-pollution treaties can be found in Mitchell [14]. (For data on the preferences of states for another treaty, see Hug and König [10].) The debate focused on

> A number of contentious issues: the seriousness of the problem; past experience with [prohibition] zones, the technical feasibility of suggested preventive measures, and cost ... [There was] a parallel between the alternative a state favored at the meeting and the information they seemed to possess on the persistence, behavior, and effects of oil in the marine environment ... In the main ... it was differing perceptions of cost/benefit that dictated the choice between a general prohibition and a system of zonal controls (M'Gonigle and Zacher [13], pp 86–87).

The 12 different positions of both individual states and subsets of states on four issues are shown in *Example E.* The first issue indicates a state's position on the desirability of a general prohibition on discharges, based on its statements for or against such a prohibition. The other three issues are votes on specific resolutions: (i) forbidding discharges by tankers; (ii) extending the prohibition zone off the British Isles; and (iii) making the North Sea a prohibition zone.

In addition to the positions of 1 (yes) and 0 (no) on these issues, we have included "A" to indicate either no statement or a vote of abstention by a state—in effect, no announced position or a position of neutrality. (Five states did not vote on resolution [iii]; we presume their positions to be A even though they did not formally abstain.) In measuring distance, we assume A to be a distance of zero from both 0 and 1. Thus, A indicates agreement, or at least acquiescence, with both the pro and con positions.

As can be seen from *Example E*, the positions of the states range from highly pro-environment (1111) to highly anti-environment (0000). As a significant number of states chose A on one or more issues, neither 1 nor 0 wins by a majority on any of the four issues—nor, for that matter, does A.

On the first two issues, the plurality winner is 0, with this position getting the support of 14 states on each issue. On the second two issues, the plurality winner

Example E. Positions of 28 states on four different issues at the 1954 Convention on the Prevention of Oil Pollution of the Sea

Positions	Number of states	States
1. 1111	5	Australia, Germany, Ireland, New Zealand, UK
2. 1A11	1	The Netherlands
3. 111A	1	India
4. 1AA1	2	Canada, Israel
5. 11AA	2	Poland, USSR
6. 10A1	1	Sweden
7. 01AA	2	Brazil, Portugal
8. AAAA	1	Venezuela
9. A0AA	1	Mexico
10. 00AA	4	Chile, Greece, Italy, Yugoslavia
11. 000A	3	France, Spain, USA
12. 0000	5	Belgium, Denmark, Finland, Japan, Norway

MV outcome (A not permitted): 0001.
FB_n outcome (A not permitted): 1100, 1001, 1010, 0110, 0101, 0011.

is A (13 players support this position on the third issue, and 14 states on the fourth issue). Consequently, 00AA (10) is the *MV* outcome if it can include As.

Because A signifies either the lack of a position or neutrality, it is questionable whether it should be counted as a component of an outcome. Ruling out A as a component, we ask whether 0 or 1 would obtain the most votes on the third and fourth issues. It turns out that 0 would win on the third issue (with the support of 8 states) and 1 would win on the fourth issue (with the support of 9 states). Thus, if A is not permitted as a component of an outcome, the *MV* outcome would be 0001.

The latter outcome is different from the six binary outcomes under FB_n: 1100; 1001; 1001; 0110; 0101; 0011. Each of these outcomes is a maximum distance of $d = 2$ from the 12 positions of the 28 states shown in *Table E*, including the two extreme positions of 1111 (1) and 0000 (12). It is worth noting that position AAAA (8) is distance $d = 0$ from *all* positions, including combinations not shown in *Example E*, and therefore would be the unique FB_n outcome if A were permitted. We think it proper to exclude this "compromise" as well as all other combinations that have any As as components. Because they are failures to act, either by approving or not approving a provision that addresses some issues, they are not compromises at all. In our view, a multilateral treaty on the environment—or anything else—should be a statement of what will and will not be done on salient issues, not a deferral of action, or a copout.

Observe that these compromises all involve two 1s and two 0s, putting them half-way between the extreme positions. By Proposition 1, they minimize the maximum distance separating them from the positions of all states, which is $d = 2$. By

Proposition 4, the *MV* outcome, 0001, minimizes the sum of the distances (or the average distance) to all states. Notice that the *MV* outcome is a greater distance ($d = 3$) from position 1111 (1) than is any of the six FB_n outcomes.

The FB_n outcomes seem more defensible as compromises than the *MV* outcome 0001, in part because of the greater distance of the latter outcome from pro-environment position 1111 (1) that is supported by 5 states. (Whether the FB_n outcomes would have resulted in greater compliance with OILPOL '54, which was spotty at best [14], is difficult to say. Subsequent treaties, especially those that raised equipment standards, did lead to greater compliance, but it is not evident that a more pro-environmental treaty in 1954 would have had this effect. While the size and importance of the 28 states shown in *Example E* states is very different when we compare the 14 states that lie above neutral position AAAA (8) with the 13 states that lie below this position, both subsets comprise similar mixes of large and small states, suggesting that the six FB_n outcomes offer a tolerable balance that reflects the different sizes, as well as the different interests, of all states.

If the six FB_n outcomes are considered more compelling as compromise choices than the single *MV* outcome, the question of which of the former should be chosen remains open. To make a choice we recommend *approval voting* [1], whereby states can approve of as many of the FB_n outcomes as they like. The outcome that receives the most approval is the winner, with ties broken randomly.

In *Example E*, let us suppose that the states are demanding and approve only FB_n outcomes that are a distance of $d = 0$ from their positions. Then combination 0011 would win, garnering approval from the 6 states that take positions AAAA (8), A0AA (9), and 00AA (10). If the states are more forgiving and approve of FB_n outcomes that are a distance of $d \geq 1$ from their positions, then combination 0011 would win again, but so would combinations 1001 and 0101, all tying with approval from 15 states.

If the states are still more forgiving and approve FB_n outcomes that are a distance of $d \geq 2$, the six FB_n outcomes would all tie, receiving the approval of all 28 states. As the latter strategy admits all outcomes as acceptable, rational players would presumably be more discriminating in order to influence the choice of an outcome.

While approval voting seems a good way for players to narrow down the set of compromise outcomes if several tie under FB_n, like all voting procedures it is vulnerable to manipulation. (Approval voting is invulnerable if the voters' preferences are dichotomous, in which case voters have a dominant strategy of voting sincerely [1]. More generally, Condorcet winners are strong Nash-equilibrium outcomes under approval voting, but other outcomes may be (nonstrong) equilibrium outcomes as well [3].) But how likely are ties? There were none in *Example B*, *Example Cb*, or *Example Da*, in each of which the number of players was a significant proportion of the number of combinations (at least 25 percent).

It is not unreasonable to suppose that treaty negotiations may occur over, say, 20 provisions, in which case there would be approximately 1.05 million combinations. If 200 states all had different top preferences, there would be less than 1 state for every 5,000 combinations, making the probability that an FB_n winner is unique exceedingly high for most assumptions about the distribution of top preferences. (In the case of voting by 1.8 million Los Angeles County voters on 28 propositions (268.4 million combinations) in 1990, the ratio of voters in combinations was much less (1 voter for every 149 combinations). Still, the most votes received by any single yes–no combinations was only 0.20 percent (1 out of every 500 voters), and that was not for the combination recommended by the *Los Angeles Times* [8][9]. Without such coordination, the probability of ties seems negligible if FB_n is applied to a significant number of treaty provisions. Supporting evidence in a different context can be found in the 2003 elections by the Game Theory Society (GTS) of 12 new council members from a list of 24 candidates. Of the 161 members who voted, which is about the maximum number of countries that participate in treaty negotiations, only two voters cast an identical ballot. It is worth noting that even though 16.8 million combinations (ballots) had to be checked, it was not difficult to determine minimax outcome (with computer assistance) (Brams *et al.* [7]), indicating that the application of FB_n to every 25 or so treaty provisions is feasible.

A more difficult question in applying FB_n is how to write provisions of a treaty so that they are (i) of approximately equal significance to all states and (ii) relatively independent of each other. If these criteria are not satisfied, then the assumption that a state's ranking of combinations is inversely related to the distance of these combinations from its top preference is not tenable. Drafting treaty provisions that satisfy these criteria is a delicate art that will surely require considerable intellectual effort and substantial goodwill on the part of the states, or blocs of states, negotiating a treaty.

6 Conclusions

We proposed a procedure, based on fallback bargaining, that we believe could help states reach consensus in negotiating and agreeing on multilateral treaties. While this procedure directly addresses voting rather than the negotiation of treaty provisions, we suggested that because the crafting and packaging of treaty provisions is done in anticipation of the voting procedure that will be used—and outcomes likely to occur under it—negotiation is difficult to separate from voting.

Because the procedure we proposed selects an outcome that is not too distant from the preferences of any party, it is likely to encourage a more frank exchange of views in the negotiation phase. While the procedure is manipulable in theory, in practice it is likely to be difficult to manipulate it because of the complexity of calculations required to make manipulation successful.

This procedure and the usual procedure for voting on treaty provisions begin with a set of proposed provisions. Under fallback bargaining with unanimity (FB_n), they should be as equal in importance and as independent of each other, as possible.

Each state indicates a top preference—either by approving or not approving each of the provisions—but abstentions are possible, as we demonstrated. From the combination that is a player's top preference, a preference order is induced, based on proximity as measured by the Hamming distance: combinations that differ by one component are ranked next highest, combinations that differ by two components next, and so on down to the combination that differs on every component from a player's top preference.

FB_n selects combinations that minimize the maximum distance of players from the compromise outcome(s) it selects (minimax outcome). If there are several such outcomes, as there were in the empirical example we analyzed, approval voting seems a good procedure for breaking ties among them. As we illustrated, the compromise selected may be quite different from the outcome that majority voting (*MV*) on each provision, which minimizes the sum of distances of players from a compromise outcome (minisum outcome), would select.

Insofar as the players, or blocs of players, are more or less equal, an FB_n compromise seems to us appealing. If the players or blocs are unequal in size or importance, this will not be reflected in the standard FB_n outcome. For this reason we suggested ways of modifying the procedure so as to reflect these different weights via different rates of descent. In contrast, *MV* takes weight into account, but it may yield outcomes that are compromises in name only—they may be far from the top preferences of some players.

While unanimity is sensible in many bargaining situations, a decision rule of simple or qualified majority may be more practicable in certain situations. Instead of requiring that the distance of all players from a compromise outcome be minimized, it would allow some players to be more distant as long as a (qualified) majority is as close as possible to the compromise outcome. (In *Example E*, FB_m gives combinations 0100 and 0000 [12] with 16 approvals at $d = 1$, which are distances of 3 and 4, respectively, from combination 1111 [1], the top preference of 5 states.) As with FB_n, FB_r ($r = m$ or a qualified majority) is vulnerable to manipulation, but fallback-bargaining procedures are likely to be difficult to manipulate in most practical situations.

While no bargaining procedure satisfies all desirable properties, FB_n seems to us a promising way of promoting consensus in complex negotiations involving several states or blocs of states. In the environmental treaty negotiations that we analyzed, a compromise resulted that was superior to both the *MV* and FB_m outcomes. Besides its technical properties, FB_n is practicable and transparent, which should help to instill confidence that its outcome respects, insofar as possible, everyone's interests.

References

1. Brams SJ, Fishburn PC (1983) Approval voting. Birkhäuser Boston, Cambridge, MA, USA
2. Brams SJ, Kilgour DM (2001) Fallback bargaining. Group Decision and Negotiation 10(4):287–316 (July)
3. Brams SJ, Sanver MR (2006) Critical strategies under approval voting: who gets ruled in and ruled out. Electoral Studies 25:287–305
4. Brams SJ, Taylor AD (1996) Fair division: from cake-cutting to dispute resolution. Cambridge University Press, Cambridge, UK, USA
5. Brams, SJ, Taylor AD (1999) The win-win solution: guaranteeing fair shares to everybody. W.W. Norton, New York, USA

6. Brams SJ, Kilgour DM, Sanver MR (2004) A minimax procedure for negotiating multilateral treaties. In: Wiberg M (ed) Reasoned choices: essays in honor of Hannu Nurmi. Finnish Political Science Association, Turku, Finland
7. Brams SJ, Kilgour DM, Sanver MR (2006) A minimax procedure for electing committees. Preprint, Department of Politics, New York University, NY, USA
8. Brams SJ, Kilgour DM, WS Zwicker (1997) Voting on referenda: the separability problem and possible solutions. Electoral Studies 16(3):359–377 (September)
9. Brams SJ, Kilgour DM, Zwicker WS (1998) The paradox of multiple elections. Social Choice and Welfare 15(2):211–236 (February)
10. Hug S, König T (2002) In view of ratification: governmental preferences and domestic constraints at the Amsterdam Intergovernmental Conference. International Organization 56(2):447–476 (Spring)
11. Hurwicz L, Sertel MR (1999) Designing mechanisms, in particular for electoral systems: the majoritarian compromise. In: Sertel MR (ed) Economic design and behaviour. Macmillan, London, UK
12. Kilgour DM, Brams SJ, Sanver MR (2006) How to elect a representative committee using approval balloting. In: Simeone B and Pukelsheim F (eds) Mathematics and democracy: recent advances in voting systems and collective choice. Springer, Heidelberg, Germany, 83–96
13. M'Gonigle RM, Zacher MW (1979) Pollution, politics, and international law: tankers at sea. University of California Press, Berkeley, CA, USA
14. Mitchell RB (1994) Intentional oil pollution at sea: environmental policy and treaty compliance. MIT Press, Cambridge, MA, USA
15. Sebenius JK (1984) Negotiating the law of the sea. Harvard University Press, Cambridge, MA, USA
16. Sertel M, Sanver MR (1999) Designing public choice mechanisms. In: Limam I (ed) Institutional reform and development in the MENA region. Cairo, Arab Planning Institute, Cairo, Egypt, 129–148
17. Sertel M, Yilmaz B (1999) The majoritarian compromise is majoritarian-optimal and subgame-perfect implementable. Social Choice and Welfare 16(4):615–627 (August)

“Adjusted Winner” (AW) Analyses of the 1978 Camp David Accords—Valuable Tools for Negotiators?

Rudolf Schüssler

University of Bayreuth, Bayreuth, Germany

More than half a century after the birth of game theory as an academic (sub)discipline, many experienced negotiators and mediators continue to work without any formal support from game theory or related subjects. In the last decades, hopes have risen that mathematical procedures for the fair division of goods might be able to bridge this gulf. “Adjusted Winner” (AW), a procedure suggested by Brams and Taylor [2], promises to satisfy several standards of fairness simultaneously and, for this reason will be scrutinized here.

AW is an algorithm for the fair division of an arbitrary number of goods between two rival parties. It has the stunning ability to produce simultaneously (a) Pareto optimality, (b) envy freeness, and (c) equal satisfaction of interests. This means that (a) nobody can achieve a better result without making anyone else worse-off, (b) nobody will want to swap his bundle of goods for someone else’s, and (c) for each player an equal percentage of his total interest in a given set of goods is satisfied. Moreover, Brams and Taylor emphasize that AW is no mere mathematical gadget but can be applied to real-life problems such as divorce or peace between Israel and its Arab neighbors. Not surprisingly, therefore, one can find the following enthusiastic evaluation from *The New Yorker* on the cover of *The Win-Win Solution*, a book in which Brams and Taylor advertise AW: “One can hire a lawyer and spend several years and thousands of dollars fighting, or one can make use of a neat new formula devised by Brams and Taylor.”

In view of this praise, it seems strange that AW is not applied more widely by practitioners of mediation and negotiation. In fact, my own informal inquiries into the use of AW by German mediators have revealed that AW has hardly been used by any of my contact persons—for a number of varied reasons. Many knew nothing of AW, let alone of more complicated bargaining theory. Those who did know or were willing to follow an introduction to AW responded with doubts. Some of these doubts should not be discarded easily. Brams and Taylor themselves acknowledge that AW faces problems. A solution to these problems is probably required before AW can fulfill the expectations of its authors, but we should be satisfied if the problems can be sufficiently mitigated in interesting classes of applications. I also believe that solutions for application problems of AW have to be tailored to the subject matter in question. Divorce cases will require different adjustments of AW than problems of international relations. I intend to concentrate on one subject area here—international relations as exemplified by the 1978–1979 United States (US)-brokered peace negotiations between Egypt and Israel (mainly Camp David One).

Camp David One has been used as an example of the application of AW by Brams and Togman [4] whose analysis was incorporated into Brams and Taylor [3]. Massoud [5] has applied AW to the closely related conflict between Israel and the Palestinians. Thus, we have enough points of departure for discussion. Part 1 of this chapter will provide a short sketch of the AW procedure as applied by Brams and Togman to Camp David One. Part 2 discusses general problems of formal approaches such as AW. Part 3 deals with more specific problems that follow from AW's mathematical properties. Part 4 contains suggestions as to how we may proceed toward a mitigation of the aforementioned problems and make AW fit for real-life applications at the same time. Part 5 provides a final assessment of the question of how the proposed use of AW can contribute to real-life conflict resolution.

1 AW and its Application to Camp David One

The AW procedure is an algorithm for the allocation of a finite number of goods to two parties (AW cannot guarantee good results if more parties are involved). The goods cannot be jointly consumed, thus the consumption option of one party leads to a loss on the other side. Nevertheless, agreements on allocation proposals that profit both parties are possible. Similar assumptions characterize standard bargaining theory to which AW is related. Raith ([10], [11]) has shown that the AW solution to a bargaining problem reproduces the well-known Kalai–Smorodinsky solution. However, AW seems much easier to understand for laypeople and mathematical pedestrians. It makes sense, therefore, to discuss AW without recourse to its background in formal bargaining theory.

AW assumes that two parties are to a varying degree interested in the same goods. The AW allocation reflects these interests in a way that satisfies the criteria of Pareto optimality, envy-freeness and equal percentual satisfaction of interests (equitability). This is achieved in a sequence of steps. First, AW confronts the parties with a complete list of the goods in question. The interests of each party in the whole bundle of goods is calibrated as 100 percent. The parties then have to decide how many percent of this total interest they assign to each of the listed goods. These percentages—or points, as they are also called—form the basis for further calculations. Note that independence and additivity is assumed for interest percentages, which means that they can be simply summed up.

Let us exemplify the first steps of AW now with our political subject matter in view. Brams and Taylor ([3] p. 95) have applied AW to the conflict between Israel and Egypt, mainly in order to assess the fairness of the Camp David One negotiation outcomes. In this context AW operates with a list of issues instead of goods. As each issue has to be allocated to one of the two sides, apparently something like control over an issue is to be distributed—but I will say more about the question of issue meanings later. For the moment, let us simply follow Brams and Taylor's estimation of the interests of Egypt and Israel concerning the most important issues at Camp David:

Table 1. Interests at Camp David

Issue	Israel	Egypt
Sinai	35	<u>55</u>
Diplomatic recognition	<u>10</u>	5
West Bank/Gaza Strip	<u>20</u>	10
Linkage	<u>10</u>	5
Palestinian rights	5	<u>20</u>
Jerusalem	<u>20</u>	5

(Each party distributes 100 points among the issues.)
Note: The issue titles stand for:
Sinai: Control of the Sinai. Diplomatic recognition: of Israel. West Bank/Gaza: Israeli control of West Bank/Gaza. Linkage: Linkage of accords with Palestinian autonomy. Palestinian rights: Israeli recognition of Palestinian rights. Jerusalem: Control of Jerusalem.

AW prescribes that all goods are to be handed over to the side that values them more highly in the first round of the procedure (we need not discuss the handling of equal valuations because none exists in the present case). In terms of AW the stated issues have to be regarded as "goods." This means that the side to which the column in *Table 1* belongs gains control over the underlined issues. In consequence, Israel has 60 percent of its interests satisfied and Egypt 75 percent. Therefore, Egyptian goods have to be redistributed to Israel until both receive the same amount of interest points. AW uses a ratio a_{AW} as yardstick of redistribution. This ratio is calculated for each of the Egyptian goods as the quotient of the percentage of interest of Egypt to that of Israel in this good (e.g., for the Sinai: $a_{AW} = 55/35 \approx 1.57$). The good with the smallest a_{AW} ratio is to be redistributed first. If redistributing this good does not reverse the order of point sums, the whole good is transferred and the process is repeated for the good with the next smallest a_{AW}. The good with the lowest a_{AW} which cannot fully be transferred without reversing the order of point sums is transferred partly. After a partial transfer the redistribution process always stops. The amount x of a partial transfer of a good is calculated according to the formula (the point sums are measured after the last full transfer):

$$x = \frac{\text{higher point sum} - \text{lower point sum}}{\text{point sum of both valuations for the good in question}} \,.$$

After the transfer of the xth part of a good, both players' interests will be satisfied to the same degree (as measured by AW). Furthermore, the distribution achieved will be Pareto-optimal and envy-free (see [2] p. 68; [3]). In the present case, control of one-sixth of the Sinai would have to be transferred to Israel in order to achieve the desired fairness features. Brams and Togman ([4] p. 317) assume in rather an ad hoc manner that the restrictions of Egyptian sovereignty in Sinai, which were agreed at Camp David, amount to one-sixth of Sinai's value to Egypt and Israel, respectively. Brams and Togman also tell us that Egypt fully got its way on the issue of Palestinian rights because Israel acknowledged the "legitimate rights" of the Palestinians. But neither the interpretation nor the enforcement of this issue came into Egyptian hands. This should make us wonder what "controlling" an issue means and what the interest points of AW really measure. We will discuss these questions in the following parts of the chapter, along with several other application problems of AW. Only by facing these problems can we assess the value of AW as a compass for negotiations or for the evaluation of their results.[1] Despite these problems, however, it seems important not to abandon AW without giving it a fair chance. AW allows us to minimize compromise in the sense that compromise has to be sought on maximally one issue. Hence, AW may direct us to the field where compromise appears most propitious in real-life cases, which makes it difficult to compromise on several issues. Therefore, AW deserves a close practice-oriented inspection before we can put it aside.

2 General Problems of Applying AW

Some of the criticisms that can be directed against AW are quite general. The most general inveigh against any use of formal procedures in conflict resolution or even in the social sciences. This kind of general critique goes back to Aristotle's assertion that human affairs cannot be measured with mathematical precision[2] [1]. According to the old and new Aristotelian view, political reality seems too complex and changing to be treated mathematically. Some of this is reflected by a dilemma that formal models in the social sciences have to face. Simple ideal models are often accused of oversimplification; more complicated models represent more aspects of reality but often lack predictive power. Neither of these problems is alien to formal bargaining theory, which still largely follows the path of Nash's Nobel laureate work. The Aristotelian critique, however, can be turned against itself. The conditions of formal modeling in the social sciences are so complex, changing, and diverse that a uniform answer to questions about the value of formal modeling seems inappropriate. The success of modeling should be judged case by case and model by model. A modest defense of AW should, therefore, be content with showing that AW can have valuable real-life applications.

Furthermore, AW should not be defended against the Aristotelian critique by emphasizing that it is a normative and not a descriptive tool. Normative instruments, too, can fail on the complexity and diversity dimensions. They can be too simplistic to be acceptable to prudent decision makers in real problem contexts, and AW is an obvious candidate for this objection. AW is, for example, completely unconnected to the personal side of politics, conflict, and negotiations. Empirical research

[1] But note that some of the discussed problems mainly pertain to international disputes. Other areas of application might pose other problems.

[2] cf., Aristotle (1984), I, 1098a26; II, 1104a6. [1]

in these areas shows that personal traits or the personal style of actors can have a large influence on the success or failure of conflict resolution ([6], [7]). The personal style or predilections of the decision makers often seem to count more than objective interests—not least because the specification of objective interests is often controversial. Therefore, negative emotions have to be pacified and personal relations have to be established before AW can be employed. Brams and Taylor [3] indicate that they see this point, but they also seem to believe that it suffices for a pacification of emotions and an understanding of persons to have occurred before AW is applied. However, a simple sequential approach—leading from the pacification of emotions through the buildup of personal relations to AW—may be too schematic. Mediators tend to let their clients work on finding a solution to their conflict throughout the mediation process. They create a human bond between the parties, which often proves essential for the stability of a mediation outcome. Interweaving the work on personal relations and on a fair division solution may therefore be necessary to stabilize the solution. If AW is framed as a quick-and-ready formal solution procedure for conflicts about the division of goods, it may well create a stability risk because it separates the work on the personal side and on the solution of their fair division problem.

Another general point of critique concerns the issue of rights. AW promises to find a fair and equitable satisfaction of interests. But what if one or both parties believes it has a right to some of the goods in question? This assumption appears quite natural in political contexts. Of course, one could try to settle legal questions before interests are dealt with by AW. But differing perceptions of rights are often so closely tied to a conflict that they cannot be handled independently from attempts to solve the whole conflict. This is especially true in the international arena where no judge may exist whom all parties accept.

The rights problem shows that normative claims can become too strong to be handled by a straightforward application of AW. On the other hand, AW also has a problem with the so-called realism in international affairs. AW presumes that both sides in a conflict are interested in a fair deal, whereas political "realists" assume that fairness talk is just a camouflage of self-interest. Even mediated negotiations would, therefore, be a means of getting the most out of an opponent, regardless of fairness considerations and related niceties. Under these premises, self-interested parties may even strive for relative gains, contenting themselves with getting more than an opponent gets to the detriment of the total of benefits. In such cases, of course, AW appears displaced, but there are also cases in which rivals got fed up with unsuccessful superiority struggles. AW depends on the realism of the insight that "live and let live" among people and nations may make the own life better. Still, political realists would insist that the veracity of conflicting parties should not generally be assumed. This creates a permanent problem for AW that will be addressed throughout this chapter.

3 Specific Problems of Applying AW

Besides the general problems that pertain to all formal fair-division procedures, there are also problems specific to AW. Brams and Taylor [2] have identified several of these problems (a–d) and I will add some more (e–f):

a Identification of issues [5];
b Mutual independence of issues ([4], [5]);
c Interest measurement ([4], [5]);
d Manipulability ([4], [2]);
e Existence of nonnegotiable claims;
f Nonlinear utility in single issues.

Let us deal with these problems point by point.

(a) It does not seem difficult to identify the most important or most salient issues of a dispute among political groups or countries. Usually, all sides in a political conflict are eager to let the world know of their claims and demands. Some of these pronouncements will be strategic and designed to deceive. But critical surveys of established bargaining positions, as well as behind-the scene communication and expert opinions may help to identify the true objectives of political rivals—if not in all cases, then at least in many. Massoud's [5] empirical assessment of the Israeli–Palestinian conflict underlines this point. No matter how skeptical one may be about the application of AW in this conflict, Massoud's or Brams and Taylor's list of most important issues seems acceptable.

However, it remains open as to how detailed issue specifications have to be. The stenographed issue descriptions of Part 1 are derived from somewhat fuller descriptions by Brams and Taylor [2]. Unfortunately, even these fuller descriptions state the issues in very general terms. It can be doubted that such general issue labels could meet the demands of AW's users. In order to handle issues like goods, AW has to assign rights to decide issues. But Egypt and Israel (or Israel and the Palestinians) would probably never accept that the other side freely uses a vaguely defined right. Furthermore, this question is complicated by the fact that interests in negotiation achievements have to be distinguished from interests in the public perception of one's results. Observers have ascribed to Egypt's former president, Sadat, a considerable interest in being perceived as an advocate of the Palestinians, and have quipped that this interest was not matched by a similarly large interest in the Palestinian case itself (see [8], [9]). Therefore, the Israelis could satisfy Sadat with a suitably formulated agreement, without giving away much of their substantial interests in the West Bank, Jerusalem, and in the blockade of Palestinian rights. Thus, to the extent that most negotiations are also negotiations about words, we will have to distinguish between interests in the wording of an issue and interests in its substance. It is not a priori clear that both kinds of interests can be bundled and dealt with by AW simultaneously.

(b) AW presupposes that the distributed goods or issues are mutually independent. This cannot be generally presumed. Massoud [5] uses cross-tabulation and statistical methods in order to show that his issues are sufficiently independent to apply AW. Brams and Taylor's [2] argument for the plausibility of an independence assumption. In fact, the possibility of using statistical techniques or of polling opinions about independence shows that the independence problem can be solved by a controlled and suitable carving of issues.

(c) None of the authors who have applied AW proceed straightforwardly by assigning point scores to issues. Instead, they start with a four-to-six-step evaluation of issues varying from least important to most important. The results of this evaluation are mapped on to a point scale which ranges from 0 to 100. This procedure generates methodological problems. But these problems seem to be of the same kind as elsewhere in the social sciences. The social sciences have accumulated considerable knowledge about scaling methods. Massoud takes this into account by testing alternatives to his scaling assumptions. This shows that AW's problem of interest measurement can be assumed to be of the standard variety in the social sciences. If you do not believe at all in the measurement of interests AW cannot get it right for you.

(d) Brams and Taylor [3] admit that AW is not strategy-proof. Players who know the preferences of their opponents can gain by suitably misrepresenting their own preferences. Of course, the fairness properties of AW vanish if such manipulation occurs. Nevertheless, Brams and Taylor. remain confident about AW because both players will very likely suffer if *both* try to exploit their knowledge of the other side's preferences. Therefore, Brams and Taylor ([3] p. 83) advise AW users:

> Without "knowing the likely announced allocations, each party may end up being too clever by half"—that is, hurting itself by being overly clever. Unquestionably, it is safer to be naive or sincere, or almost so.

Massoud [5] acknowledges Brams and Taylor's [2] assessment of the hazards of manipulation in a footnote, apparently assuming that it suffices to repel the objection to manipulation. However, the hazards of manipulation should not be underestimated. It has to be admitted that a good knowledge of one's opponent's preferences is a prerequisite of successful manipulation. But marital partners who file for divorce usually know the preferences of their spouses. And if Brams and Taylor or Massoud are able to measure the interests of political rivals, the rivals may do so as well. The results of attempts to use this knowledge are hard to predict. A significant expectation of manipulation by both sides can easily lead to complex interdependent expectations about the behavior of the other side. Nevertheless, the strategic structure of this interdependence seems fairly simple under some idealizing assumptions about manipulative moves. Imagine that the above-stated honest point allocations of the Egyptians and Israelis (see Part 1) were known to the other side. Let us proceed on the assumption that one side (side A) engages in manipulation by using the following strategy. First, side A tries to obtain the same goods that it would get in a truthful application of AW. This is done by assigning one point more than the opponent to each of the goods side A would receive in the case of a truthful representation. The remaining points are distributed in a way that minimizes the difference between the remaining point assignments of both sides, but (i) without losing one of the obtained goods completely or partly to side B and (ii) by trying to get a maximal percentage of the good that side A values most among the goods still held by side B. Although I have no mathematical proof, this way of unilateral manipulation was the most profitable in my tests. In the Camp David case it produces roughly an advantage of 79 to 51 points if Egypt engages in manipulation, and 75 to 51 if Israel is the manipulator. A truthful representation of preferences would lead to an equitable outcome of 66 percent of their own interest points for both sides. But—corroborating Brams and Taylor—simultaneous manipulation by both sides would lead to an outcome of only 34 points for both parties. Therefore, both sides lose more than twice their potential gains when both manipulate (32 versus 13 or 9). Furthermore, Brams and Taylor [3] emphasize that sincerity guarantees envy-freeness. This means that those who reveal their preferences truthfully believe they will never get a smaller bundle of goods than their opponents. Envy-freeness, however, cannot be guaranteed to manipulators. A score of 34 out of 100 points means that both sides would rather have the goods of the other side. We may derive a neat moralistic lesson from this, namely, that envy is the predicament of manipulators.

Brams and Taylor think that this lesson suffices to defend AW, but in fact we should not be satisfied. Our presumed Egyptian and Israeli negotiators are caught in the following game-theoretical structure:

Table 2. Agricultural ”Chicken” game

Egypt/Israel	Tell truth	Manipulate
Tell truth	66, 66	51, 75
Manipulate	79, 51	34, 34

Table 2 represents a (asymmetrical) “Chicken” game. The truth–truth cell contains no Nash equilibrium, and the players have no dominant pure strategies. “Chicken” games have two Nash equilibria (here the lower-left and upper-right cells) in pure strategies and one Nash equilibrium in probabilistically mixed strategies. If we accept that Nash equilibria represent stable strategy combinations that rational actors could play, mutual truth-telling is no stable result. In both pure strategy equilibria one of the actors manipulates. Moreover, if each actor attempts to reach the equilibrium that is better for him, both will manipulate and achieve the worst possible result. If, instead, a mixed strategy equilibrium is played, the probability for truth-telling would be (roughly) 0.65 for Egypt and 0.57 for Israel. Each player would then (roughly) receive an expected 60 points. But the probability of truth-telling is not very high, after all, and truth-telling by both would only occur in (roughly) 37 percent of all cases. In my opinion this is not really satisfactory for making our example a case for truth-telling in AW. Moreover, the “Chicken” structure is not peculiar to the present case. Under the premises of AW, nobody will receive less than 50 percent of his interest points after a truthful revelation of interests, whereas mutual cheating can easily lead to worse results. If the latter should happen, a “Chicken” game will emerge, and in all such cases mutual truth-telling is not a Nash equilibrium.

(e) Negotiators sometimes use “all-or-nothing” claims, indicating that if these claims are not fully met, all negotiation efforts will fail. AW’s attitude to nonnegotiable claims is complex. Theoretically, AW cannot accept nonnegotiable claims, because it must remain possible to allocate each good or issue to the side that attaches a higher value to it. Even if a party were to put all its interest points on one nonnegotiable issue, it would—in principle—have to be prepared to give away part of this good during the "adjustment" stage of AW. On the other hand, AW has to divide maximally one good. Therefore, under fortuitous conditions, each side could receive all the goods it categorically wants. It follows that AW can, ceteris paribus, cope better with nonnegotiable claims than other procedures—such as Proportional Allocation[3] —which would divide more than one good. Obviously, the smaller the number of divided goods, the higher, ceteris paribus, will be the probability that non-negotiable claims do not become involved in divisions. Thus, the problem of hard claims can be turned to AW's advantage.

(f) Nonlinearity has made an appearance under (c) as a problem of the distribution of points between goods. But there is a second kind of nonlinearity that may arise if a single good has to be divided according to the AW formula. AW's adjustment procedure embodies a linearity assumption for the redistribution of goods because it redistributes in proportion to the complete interest of the parties in a good. It is not obvious that this assumption reflects the parties' interests. For example, it is far from clear that, to Egypt, the complete possession of, say, 80 percent of the Sinai is, in fact, worth 80 percent of the value of the complete possession of the Sinai. We may expect nonlinearities here, with considerable added value for the complete control of a good. In divorce cases nonlinearities can be handled by selling goods and redistributing money—assuming that the utility of money is almost linear in the envisioned region of both sides' utility functions. However, neither the Sinai nor Jerusalem are for sale, nor do countries normally sell parts of their territory. Nonlinearities or thresholds in the valuation of limited property rights can therefore become very serious obstacles to the application of AW in international politics.

4 A Suggested Mode of Application for AW

At first sight the problems seem to inveigh strongly against the application of AW in international affairs. However, I do not want to draw such a pessimistic conclusion. Instead, I want to suggest an "embedded" mode of application of AW which, I think, will help to bypass or overcome many of the problems mentioned. In this mode AW is used as a heuristic tool for mediators in international negotiations. It is not used to evaluate negotiation outcomes with hindsight or to compute solutions with only the interests of the involved parties given. In my view the use of AW should become embedded in a process of mediated negotiation; it should alternate repeatedly with the use of standard mediation and negotiation techniques; and it may become dependent on them (see [12]). Under these premises, AW can inform its users that a fair and equitable solution could be achieved if certain compromises were made. A mediator is needed because the input into AW and the interpretation of a solution will have

[3] For a general description of this mechanism, cf. Brams and Taylor [2].

to be negotiated (see [3]). The circumstances of these negotiations raise questions of evenhandedness, which can best be dealt with by mediators. Sometimes it may also be necessary to have a mediator who uses expert opinions to mitigate the impact of possible misrepresentations of interests by the conflicting parties. Finally, a mediator seems best suited to guide the embedded application of AW through its three steps.

The first step is concerned with the clarification of issues. In Brams and Taylor or Massoud the word “issue” denotes a rather general label of main areas of interest. A catalog of such issues is a natural starting point for AW because it tells us how to parcel negotiation topics and how to create a list of subjects that are to be dealt with by AW. Issues that serve such purposes, however, are usually not specific enough to allow for a meaningful application of AW. Massoud [5] deals with this point in a footnote. He acknowledges that the experts he has asked could have different ideas of what the issue of, for example, Jerusalem signifies. In the end he marginalizes this problem, whereas I regard it as crucial. Negotiators will want to know what “getting the Sinai” means before they might let the other side get the Sinai. Things have to be spelled out, therefore, before AW can be applied to real-life cases. It is assumed here that what should be spelled out are “fair maximal concessions” (FMCs). We are interested in maximal concessions because AW seeks compromise only in the latter stages of its application. We should want to know how much each side would maximally let the other side get away with in the beginning. Later, corrections follow, and the idea is that each side can expect to be fairly and equitably compensated for on other issues for maximally conceding one issue to the other side. One can imagine, therefore, that both sides answer issue for issue the question of how much they would concede to the other side if they were fairly and equitably compensated in other areas of interest. The answers to this question, every one of which generates an FMC, might result in quite complicated texts. Nevertheless, these texts define what control over a good means in the embedded use of AW.

We can already speak of an embedded use, because the FMCs will very likely have to be negotiated. Asking the parties will raise the problem of veracity, so the process of investigating the FMC question should be more complex. It may have to rely on expert knowledge and on information about the former negotiating positions of rival parties. A “mediator with clout” may also push the parties to FMCs that seem to be realistic and not merely strategic traps. Finally, a consensus of the parties concerning all FMCs would “heal” the problem of veracity to a large degree. If one side accepts a proposal of the other side as FMC, it becomes a secondary matter as to whether a better FMC would have existed under conditions of full interest revelation. Acceptance implies that a workable starting point for the further application of AW has been reached. But acceptance will, even if things go well, probably only emerge after a certain amount of negotiation about FMCs.

How does reliance on FMCs mitigate other problems of AW besides the problems of veracity and of a detailed description of issues? First, the interest of a party concerning its public image can be satisfied by suitable formulations of an FMC. FMCs embody the result of a weighted interest consideration by both parties, and as texts they can be so formulated so as to serve public relations purposes. It is not necessary to account for different kinds of interests in the fair division procedure if the present approach is used. The elaboration of FMCs will also reveal whether incompatible nonnegotiable claims of both sides exist. If so, a set of mutually acceptable FMCs for all issues will not exist. In such cases, negotiations end in a deadlock, unless progress toward compatibility is made. Finally, FMCs enable both parties to introduce their perceptions of rights into the AW process. Concessions from both sides will probably depend on the rights they believe they have. An agreement on FMCs, therefore, already reflects interpretations of rights and intentions to use them.

After an agreed list of FMCs has been established, the second step of the embedded use of AW can be taken. This step deals with the measurement of interests and point allocations. The problems of interest measurement belong, as indicated, to the standard variety in the social sciences. Unfortunately, there is the problem of veracity concerning point allocations. If the advice not to underestimate this problem is heeded, we should distrust the sincerity of declarations of interest made by negotiating parties. However, in most cases in which the involved parties can be expected to know the interests of their opponents, we may also assume that third parties or mediators have a similarly precise knowledge of the interests of both sides. After all, Massoud [5] asked neither the Israeli nor the Palestinian government for a quantified statement of interests but did reconstruct the interests in question on the basis of collateral information. Mediators who are willing to make suggestions to their clients (and such mediators are likely to be found in international politics) may employ AW to point out where compromises could most profitably and parsimoniously occur. This can prove helpful, although it might not work in all cases. We should also take into account that additional filtering procedures may help to identify and remove strategically motivated misrepresentations of interests.

In a further third step the AW solution is calculated and related to the negotiation situation. AW's mathematical properties ensure that in all except one issue one side will gain a full FMC. However, the one pair of FMCs for which compromise has to be sought will force the parties back to the negotiating table. The partial satisfaction of political proposals cannot be as easily measured as the division of a homogeneous, infinitely divisible good. Intermediate compromise proposals will have to be developed for the compromise issue. More precisely, we will have to establish a set of intermediate positions between the opposed FMCs in the issue that is marked for compromise. Note the parsimony of AW. We do not have to find intermediate positions for more than one issue. Nevertheless, the suggested positions have to be acceptable as possible compromises for each party. A realistic set of intermediate positions can therefore only be developed after consultation and—probably—negotiation with both parties.

For the application of AW the intermediate positions should be thought of as "pearls on a string," which means that they should be ordered according to their greater or smaller acceptability to the negotiating parties. Of course, problems can arise during the ordering of intermediate proposals. Some proposals may be better than others for both parties. Such win–win options are to be preferred and inferior options should be dropped, so that the final string consists only of Pareto-undominated proposals. It could also happen that new Pareto-superior maximal positions become known after AW has already been applied. In this case, unfortunately, AW's application process would have to be repeated. If a suitable "string of pearls" has been found, however, it has to be assessed as to whether one of the pearls comes close enough to the AW solution or not so that the process can be terminated. This assessment cannot be made without consulting the parties again because the value of a suggested compromise to the parties is not obvious. The logic of AW requires that the whole of the "string of pearls" should be evaluated separately by both parties, uncovering what fraction of the total interest in one issue each "pearl" represents. This can be done by standard methods of utility measurement. Unfortunately, there is no guarantee that the combination of valuations that constitutes the AW solution will be embodied by a compromise proposal. AW assumes that the solution compromise has a certain value v_A to side A and v_B to side B. Both values are calculated under the assumption of linear utility. With nonlinear utility, v_A and v_B may become assigned to different compromise proposals. It would be helpful if AW could solve the problem of nonlinearity mathematically, but standard AW is not designed to accomplish this task. We have to hope that future refinements of AW will fill this gap, as the task seems mathematically feasible.

In the meantime we may pragmatically assume that the AW solution guides us to a fair and equitable result if the difference between the parties' real valuations and the ideal linear AW solution remains small. A yardstick for judging this question is, again, the consensus of the parties. If they regard a deviation from the AW solution as acceptable, they accept the result as sufficiently fair and equitable for them. If necessary, both parties may also demand the creation of new intermediate proposals between the established proposals that come closest to the AW solution and see whether some of them are more acceptable than the former proposals. Such haggling may not satisfy mathematicians, who may fear for the cherished ideal fairness properties of AW, but minor deviations from ideal worlds may go through in reality.

There may also exist cases in which the valuations of both players do not converge sufficiently to support the described solution. In these situations, the application of AW will fail as long as the problem of nonlinearity has not been solved mathematically. We will now have a closer look at the question whether this shows that AW should be abandoned.

5 AW as a Tool for Some Occasions

In the preceding part the idea of an embedded application of AW was introduced. This approach assumes that rather specific "fair maximum concessions" are negotiated for each issue. Finally, a "string of proposals" has to be negotiated in order to implement AW's redistribution suggestion. AW will succeed only if both sides accept that one of these proposals (roughly) represents the AW result. The application of AW becomes deeply embedded into a supporting structure of negotiations, which accompanies the use of AW in most of its stages. But if a formal procedure such as AW cannot succeed without informal support, why do we need it? Moreover, if informal negotiations continue if AW fails, will a prior use of AW not endanger the success of the final rescue attempt? The rationality of trying out AW in the first place depends, indeed, very much on the burden that a failing AW application would create for an informal continuation of negotiations. A burden could result from the time-consuming need to seek new information or to trace alternative advance routes for negotiations. Similar problems arise for all negotiations that end in an impasse. Nevertheless, something more specific can be said. As mentioned, AW minimizes the number of issues in which compromise has to be sought. Thus, AW leads us down a wrong path if we realize at some time that compromise in all or many issues will be the best way to resolve a conflict. On the other hand, AW's application steps coincide naturally with informal negotiation strategies that also seek to minimize the number of compromise issues. Such strategies make sense in environments where, for example, a mediator, the rival negotiators, or their principals have to satisfy the interests of several clients, each of whom is uncompromisingly interested in only a few of the negotiated issues. In this case, limiting the number of compromise issues may help to soften the mood of resistance among clients and—by the logic of *divide et impera*—to isolate and overcome the stiff resistance of a minority of clients. One could also modify this strategy so that a major compromise has only to be sought in one issue, whereas in all others only minor compromises are made. At first sight, it sounds odd to combine such strategies with AW, because Machiavellian *divide et impera* recommends seeking compromises where the expected resistance is lowest and not where fairness and equitability is maximized. However, we have to distinguish between long-term and short-term considerations. The fairness and equitability of a solution may not only be valuable in itself but will probably also increase the stability of the solution and enhance future cooperation chances with former opponents. Looking for the point of weakest resistance may, on the other hand, pay off only for a short time. An embedded use of AW may help to develop more prudent and less brutal negotiation strategies of a type that seeks to minimize the number of compromise issues. If an AW application succeeds, the results of such strategies become crowned with AW's fairness and equitability properties. If the application of AW comes to naught, the given FMCs will at least satisfy a majority of clients, while the task of finding an appropriate compromise issue can be considered under new auspices.

What we end up with is a plea to take AW seriously as an instrument that should belong in the toolkit of (political) mediators. It has become clear that AW will not be an instrument for all occasions. But the occasions for which its embedded application seems appropriate constitute a nonnegligible class of real-life cases. In these cases, AW can guide a mediator on his way and indicate chances, if there should be any, of reaching a result with remarkable fairness and equitability features. The opportunity costs of looking for such chances are small if a negotiation strategy of minimizing the number of compromise issues is pursued.

References

1. Aristotle (1984) Nicomachean ethics. In: The complete works of Aristotle, Vol. II, Princeton University Press, Princeton, USA
2. Brams S, Taylor A (1996) Fair division. Cambridge University Press, Cambridge, UK

3. Brams S, Taylor A (1999) The Win–win–solution: guaranteeing fair shares to everybody. W.W. Norton, New York, USA
4. Brams S, Togman J (1998) Camp David – Was the agreement fair? In: Harvey F, Mor B (eds) Conflict in world politics. Macmillan, Houndsmills, 306–323
5. Massoud TG (2000) Fair division, adjusted winner procedure (AW), and the Israeli–Palestinian conflict. Journal of Conflict Resolution 44:333–358
6. Mastenbroek W (1991) Development of negotiating skills. In: Kremenyuk V (ed) International negotiation. San Francisco, CA, USA, 379–399
7. Meerts P (1997) Negotiating in the European Union, Group Decision and Negotiation 6:463–482
8. Quandt W (1986) Camp David. Brookings Institute Press, Washington DC, USA
9. Quandt W (2001) Peace process. Brookings Institute Press, Washington DC, USA
10. Raith M (1999) The Structure of fair-division and the design of fair-negotiation procedures. In: Patrone I, Garcia-Jurado I, Tijs T (eds) Game practice contributions from applied game theory. Dordrecht, The Netherlands, Chapter 14
11. Raith M (2000) Fair-negotiation procedures. Mathematical Social Sciences 39:303–322
12. Touval S, Zartman W (1985) Introduction – mediation in theory. In: Touval S, Zartman IW (eds) International mediation in theory and practice. Westview Press, Boulder, CO, USA, 7–17

Procedural Design for Conflict Resolution

Matthias G. Raith

Department of Economics and Management, Otto-von-Guericke University, Magdeburg, Germany

1 Introduction

Negotiation over conflicting interests and demands involves two separate but interacting dimensions. One dimension is given by the *problem*, which includes the content and the structure of the negotiation. The other dimension comes with the *players*, and it characterizes how the game is approached. This perspective offers a broader understanding of the negotiators' behavior because it allows one to distinguish between the actual actors in negotiation and the role they are playing in the conflict. Consequently, there are two components of bargaining power, which Underdal [27] classifies as *structural* and *behavioral*, that have an influence on the negotiated agreement. Since the roles are part of the structure (i.e., the problem), structural bargaining power is what remains when the players of the game are exchanged. To what extent the structure affects the outcome, of course, depends on the players' individual bargaining skills. However, if negotiators are similar in their skills—a situation one might expect in high-level negotiations—then the outcome of negotiation will be determined by the structure. Moreover, only knowledge of the game allows one to assess how well negotiators play and to appreciate the quality of their agreement—Is it fair? Is it satisfactory? Who won? and so forth. In order to give negotiators advice on how to achieve their desired outcome, one needs an analytical approach that is focused on the game (the problem), while at the same time acknowledging that it is being played by real players (the negotiators), who are interacting with one another in some way.

In this chapter we wish to emphasize the importance of understanding not only the structure of the conflict but also the process of interaction. We highlight the limits of both cooperative and strategic bargaining theory, as bargaining theory, or more generally game theory, is a normative approach characterizing merely role interaction—the player dimension is typically neglected.[1] As an alternative, we advocate here the development of fair-division procedures as a practicable joint

[1] This apparent deficit presumably accounts for the increasing popularity of experimental research with real subjects. Despite the many new insights that experimental research has

decision-making approach to enrich the framework of negotiation analysis. We show how fair-division procedures, designed to operate within a specific conflict structure can provide guidance for the actual negotiators involved in the conflict. However, our main interest is not algorithmic elegance but rather practicability. We provide several examples of how different fair-division procedures can and should be (re)designed for practicability. Finally, we argue how and where these procedures can be integrated into the conflict-resolution process, in order to complement the actual negotiation, to support the negotiators or their facilitator, or even to substitute for the complete negotiation process.

2 Negotiation Analysis and Bargaining Theory

In his characterization of the negotiation-analytic approach, Sebenius [23] identifies four significant features. First, negotiation analysis seeks to generate prescriptive advice to one party given a descriptive view of the others. The apparent deviation from the normative view of game theory highlights the importance of the descriptive elements of decision analysis and acknowledges the influence of cognitive and social psychology in characterizing parties' behavior. Second, negotiation analysis postulates a subjective perspective, which encompasses not only negotiators' view of the bargaining problem itself but also their concern about procedures, relationships, and notions of fairness. Thus, there is an explicit recognition of the behavioral dimension of negotiation. Third, the negotiation-analytic approach acknowledges that negotiators often reach an agreement, where there is "value left on the table." This is not to say that practitioners are not interested in efficiency but that they are limited in their ability to deal with complexity. Fourth, negotiation analysis typically addresses practical questions of negotiation that mathematical bargaining or game theory has either deliberately excluded or has not been able to cope with.

Axiomatic solution concepts of *cooperative* game theory focus on the desirable properties of bargaining solutions but they do not specify the process of interaction that leads to them.[2] With a list of potentially crucial axioms that a solution should satisfy, Sebenius critically observes that there is "an impressive number of candidate solution concepts" to choose from. Moreover, even when a solution has been found, the uniqueness of an agreement with the desired properties is not guaranteed. It seems that multiplicity of equilibria is generally seen as a disadvantage: even though there may be more possibilities for improvement, thus enhancing the negotiators' flexibility, the choice of alternatives usually gives rise to coordination problems. On the other hand, one could argue that the nonuniqueness of equilibrium provides more room for acknowledging the behavioral component of bargaining power. What is apparently needed is a formal theory characterizing the behavior of players, that is, the behavioral dimension of the game, which has largely been neglected by game theory.

to offer, it is interesting to note that the experimental approach itself is often criticized for operating mainly in an artificial laboratory environment.

[2] To a certain extent, *implementation theory* tries to deal with this problem.

The desire to explain how cooperative solutions are realized has initiated the noncooperative approach to bargaining, also known as the "Nash program." This field of research focuses in particular on sequential strategic interactions allowing one to include procedural aspects and institutional frameworks in the description of a bargaining game. As Binmore *et al.* [2] argue, with a clear-cut and unambiguous procedure the bargaining problem can be analyzed as an noncooperative game. By modeling bargaining procedures that lead to or approximate axiomatic solutions, one can derive a process-oriented characterization and thus an alternative basis for the comparison of solution concepts.

In strategic bargaining theory the procedure organizes the plan of play: What is a feasible proposal? When can a proposal be made? Who can make a proposal? What is a legitimate response to a proposal? Given the procedure, players then determine their optimal strategies (plans of action) in interaction with the other players. According to Raiffa *et al.* ([18], p. 84), "an interactive decision-making perspective obliges a negotiator to consider carefully the alternatives, interests, aspirations, and behaviors of the other side." Because of the postulated optimizing behavior of the players, these models offer powerful theoretical tools, such as the use of equilibrium concepts, for studying the relationship between the procedure and the negotiated outcome (e.g., by applying comparative statistic analysis).

Following this line of research, theoretical contributions dealing with multi-stage negotiations provide answers to important procedural questions such as: When should a negotiation be decomposed into stages and negotiated according to an agenda? For example, should peace negotiations be conducted in one big round or in several smaller rounds? In what order should issues be dealt with? Should the complicated issues be negotiated at the beginning or at the end? How many stages are optimal? How large (important) should individual stages be? Winter [28] and John and Raith [10] address these questions within a normative strategic setting, thus focusing exclusively on the structural dimension of negotiation. Remarkably, their results are qualitatively similar to those of descriptive analyses emphasizing mainly the psychological aspects of negotiation (e.g., Gillespie and Bazerman [8]) (i.e., the behavioral dimension).

Although the idea of the Nash program is to classify noncooperative settings, the main criticism of the strategic approach is that it is difficult to derive general implications from a specific model of the bargaining process. According to Binmore, *et al.* [2], with *incomplete* information the models become even more particular, and there is only little room for firmly grounded guiding principles. Moreover, as many experiments have shown, strategic bargaining models are less suited to predicting the outcomes of actual negotiations, which is largely because of the underlying assumption of strict rationality. A further important aspect, which generally seems to be taken for granted, is the undeniable fact that the traditional strategic approach *is* noncooperative. This is a strong behavioral postulate, because even when parties have conflicting interests, it is not guaranteed that the actual actors negotiating automatically behave noncooperatively. The interests are part of the structure, that is, they belong to the problem dimension, but the behavior comes with the players. Indeed, many descriptions of professional political or business negotiations seem to indicate a more cooperative, problem-solving attitude. For this reason, Raiffa [17] devotes most of his analysis to situations where negotiators are engaged in what he refers to as "full" or "partially open truthful exchange." Why one finds negotiators playing truthfully and cooperatively, even when there is no enforcement mechanism aside from self-commitment, is not entirely clear. But, just because theorists are not able to fully explain cooperative behavior, this does not mean that the game is being played noncooperatively.

3 The Fair-Division Approach

According to Brams and Taylor ([4], p. 66), "bargaining theory has proved singularly inapplicable to the settlement of real-life disputes" because of its "divorce" from theories of fair division. Indeed, fair-division models seem particularly attractive for the negotiation-analytic approach outlined above because of two characteristic features.

First, the focus on *fairness* emphasizes the importance of cooperative norms in finding an agreement. This implies a behavioral assumption that clearly deviates from strategic bargaining. According to Young [29] norms are plausible standards that help parties coordinate their expectations, thereby narrowing the range of acceptable agreements. In addition, norms enhance the durability of agreements because they are relatively stable over time. Of course, if the number of standards is as large as the number of possible agreements, they become useless as focal points, but this does not seem to be a problem for most real-life negotiations. On the contrary, Schelling ([21], p. 67) is impressed by

> the remarkable frequency with which long negotiations over complicated quantitative formulas or ad hoc shares in some costs or benefits converge ultimately on something as crudely simple as equal shares, shares proportionate to some common magnitude (gross national product, population, foreign-exchange deficit, etc.), or the shares agreed on in some previous but logically irrelevant negotiation.

If norms in fair division are standards of distribution, then they play the same role as the solution concepts of cooperative game or bargaining theory. A solution, derived within a normative framework and characterized by a bundle of desirable properties (axioms), is a *norm*. Thus, the Nash program may be criticized for having narrowed our view of cooperative solutions as *predictors* for noncooperative games. As such they can be (and have been) tested and sometimes falsified. As norms, cooperative solutions cannot be falsified. However, they can be relevant or irrelevant in practice, and in order to assess their relevance, one must check to see if those negotiators who know and understand them also use them.

Should a theory be criticized for producing too many norms? From a decision-analytic perspective, as outlined by Keeney and Raiffa [12], a solution concept not only provides negotiators with an agreement but also delivers the argumentative basis that separates the chosen agreement from other alternatives. The argumentation is important not only for the psychological comfort of the negotiator but also for justifying the proposal to others. Indeed, the difference between an experienced and inexperienced negotiator seems mainly to be the knowledge of norms and how to use them. Again, it is Schelling ([21], p. 69) who emphasizes the behavioral dimension of negotiation by observing that bargaining skill is "the ability to set the stage in such a way as to give prominence to some particular outcome that would be favorable." Thus, the case against cooperative-bargaining-theoretic solutions cannot be that there are too many alternative concepts to choose from but only that they are too abstract to argue in practice.

The second characteristic feature of the fair-division approach is that the agreement is the outcome of a procedure, which is why Young [30] speaks of "fair bargains." By focusing on procedures to achieve fairness, one obtains a shift in perspective from *interactive* to *joint* decision making, which Raiffa *et al.* [18] see as a crucial feature of the negotiation-analytic approach. A procedure decomposes the implementation of a solution into successive steps that negotiators can understand and follow together. If the steps are intuitive, they will support the communication process by providing the arguments that, according to Keeney and Raiffa [12], are necessary to implement a solution in practice. This view also corresponds to the philosophical argumentation-theoretic perspective that, according to Tindale [26], requires a synthesis of logic, dialectic, and rhetoric, that is, a formal theoretical foundation characterizing the argumentative dialog between players.

The psychological advantage of a procedure is that it also enables parties to actively participate in the implementation of a solution. As Ackoff [1] observes, "one can enjoy a game played by others, but one can only have fun by playing it oneself." Negotiation support systems often fail to acknowledge this aspect. This could explain why Spector [24] finds empirical evidence that quantitative analytical support to negotiations is not popular in actual bargaining situations. What is needed, in his view, is a procedural assistance that guides negotiators to efficient outcomes.

4 The Design of Practicable Fair-Division Procedures

By listing the interesting features of a procedure, as we have done above, one does not yet have a formal definition of a procedure. Moreover, there still remain open questions that need to be addressed: How does a negotiation procedure relate to the negotiated outcome? Or, what are the behavioral components that distinguish a negotiation *procedure* from an *algorithm*? A procedural approach to conflict resolution operates on the structural level, but in order to explicitly involve the actual players it must be conducted on the behavioral level. In the context of fair division, Haake *et al.* [9] characterize a procedure as featuring the following characteristics: it must be *intuitive*, meaning that each step fits the parties' way of reasoning; it must be *plausible*, meaning that the sequence of steps naturally follows their line of argument; and it must be *manageable*, meaning that all calculations must be basic enough for them to compute without technical support.

The development of practicable procedures for cooperative negotiations has become quite popular over the past decade. In particular, the work of Brams and Taylor

[4] has continued the line of research on fair-division procedures initiated by Knaster [14] and Steinhaus [25]. In order to compare different approaches in this line of research, let us first consider a simple two-party fair-division problem.

Assume that two parties A and B have the opportunity of dividing an estate consisting of five different items.[3] In an international setting the items could represent various issues, for example, the disputed regions in a territorial conflict, specified emission rights in an environmental negotiation, or the crucial issues in a business merger.

If the individual items cannot be divided, there are $2^5 = 32$ alternative ways of allocating the five items to the two parties. Suppose that the estate is decomposed in such a way that parties have additively separable preferences over the five items. The total value that each party receives in the division of the estate is then given by the sum of the values of the received items. There are different ways of valuing the items, depending on the specific context. For example, parties may be able to assess the value of the items on a monetary scale, say, in (millions of) euros, thus being able to place bids as in an auction. In an international negotiation over the distribution of emission rights this may well be the case. An example of individual subjective monetary valuations by both parties is given in panel a) of *Table 1.*

Table 1. Players' subjective valuation of items

a) Assessment of items in euros

Item	A	B
1	4	18
2	4	9
3	32	9
4	20	6
5	20	18
Σ	80	60

b) Assessment of items in %

Item	A	B
1	5	30
2	5	15
3	40	15
4	25	10
5	25	30
Σ	100	100

The structure of the fair-division problem, as characterized by panel a), is visualized in *Figure 1* by the 32 possible allocations of the five items. Each point, characterizing a specific allocation, shows the received monetary values, as perceived by party A or B, respectively. Thus, the structure of the problem depends on the perception of the parties, irrespective of how the actual negotiators (representing the parties) intend to interact within this structure.

As can be verified, the 32 allocations are distributed pair-wise point-symmetrically around the point (40,30), characterizing the (50%, 50%) allocation.[4]

[3] We adopt the terminology of Raiffa *et al.* [18] by referring to decision entities as "parties" rather than "players" in order to emphasize the deviation from the strategic approach to bargaining.

[4] Consider, for example, the allocation marked R, where party A receives a value of €72 million (with items 3, 4, and 5) and party B receives €27 million (with items 1 and 2). If both parties simply exchange their packages, thus realizing the allocation R', then party A

The 50:50 point separates the 32 possible allocations into four regions. In the southeast region party A receives more than 50 percent of the estate, while player B receives less. In the northwest it is the other way around (because of the point symmetry). Thus, in both of these regions one of the two parties would be better-off with the package of the other party and would therefore prefer to trade the packages of items. In the southwest region both parties receive less than 50 percent of the estate. Consequently, both would want to trade packages, leading to the northeast region, where each party receives more than 50 percent of the estate. The allocations in this region can be considered as "envy-free," in the sense that neither party would prefer the package of the other. If one is interested in a fair division of the estate, then envy-freeness is an attractive feature. As one can verify in *Figure 1* for the specific case given in *Table 1*, only four allocations provide each party with at least 50 percent of the estate.

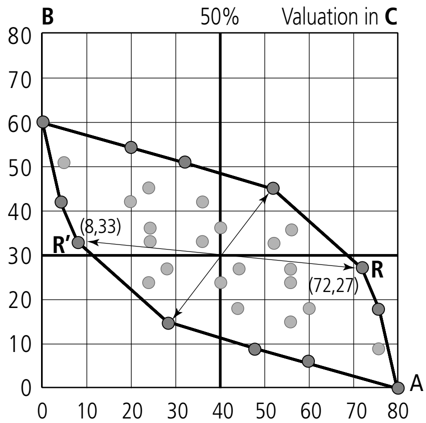

Fig. 1. The structure of a bilateral division problem

The range of allocations in *Figure 1* and, in particular, the region of envy-free allocations is bounded to the northeast. This boundary line connects what Raiffa [17] refers to as the "extremely" efficient allocations.[5] To illustrate the point symmetry, and to visualize the structure of the fair-division problem, we have also added the lower boundary in the southwest direction. Efficiency is a further interesting property for the division of an estate because it makes sure that there are no gains left on the table. As *Figure 1* illustrates, only two allocations feature both envy-freeness and efficiency, with only one, (viz., [52], [45]), being extremely efficient.

will end up with items 1 and 2 and a value of 80 – 72 = €8 million, while party B receives 60 – 27 = €33 million.

[5] "Extremely" efficient allocations are efficient allocations that cannot be dominated by convex combinations of other efficient allocations.

Because of their differing valuations, parties generally place differing bids on the items and, as a consequence, also value the total estate differently. In our example, party A's valuation of the estate is €20 million higher than party B's. An alternative characterization is to view the estate, that is, the five items together as the complete "cake" worth 100 percent and then to assess the relative importance of each individual item in contributing to the value of the estate. This is shown in panel b) of *Table 1*. Indeed, the method of relative valuations is the more general approach because it also works in many situations in which a monetary valuation is inconceivable or may simply appear immoral. According to Raiffa [16], this method of weighting items was first introduced as an analytical assessment tool for the Panama Canal negotiations in 1974, where there were 10 items on the agenda. Brams and Taylor [5] use the method of relative valuation in their analysis of the Camp David accords of 1978, where geographical issues, viz., the status of the Sinai Peninsula or the West Bank/Gaza Strip were simultaneously negotiated with political issues such as the diplomatic recognition of Israel, the recognition of Palestinian rights, or the control of Jerusalem, all of which can hardly be assessed in monetary terms. Of course, whenever monetary assessments are possible, the values in panel b) of *Table 1* should be consistent with those of panel a).

There is a long history of established procedures for dealing with problems of this type. Brams and Taylor [5] offer an insightful overview of different approaches and perspectives, beginning over 5,000 years ago with the procedure known as "divide and choose," which guarantees envy-freeness. In fact, this simple procedure was implemented in 1982 at the United Nations Conference on the Law of the Sea (UNCLOS) to allocate undersea mining tracts fairly between industrialized and developing countries (see Sebenius [22]).[6]

Although it has not yet been put into practice in the context of seabed mining, one may doubt whether "divide and choose" is the best choice for this particular problem, as it does not ensure efficiency. Moreover, it is becoming increasingly difficult to simply divide the world into the two blocs of developing and developed countries, as many even small countries want their individual situations to be acknowledged. Since UNCLOS over 20 years ago, the research on fair division has revealed alternative procedures that not only provide efficiency but would also presumably be easier to implement in practice.

4.1 The Knaster–Steinhaus procedure

Research on the mathematical design of fair-division procedures began most prominently in the last century with the work of Knaster [14] and Steinhaus [25], which we also choose as the starting point of our procedural analysis.

The setting for the Knaster–Steinhaus procedure is as given in panel a) of *Table 1*: Parties are able to value the individual items in a common monetary unit of account, for example, as one would expect at an auction. Formally, we are considering a bargaining game with transferable utility, which has important practical implications

[6] For a detailed analysis of this case see Sebenius.

for the design of a procedure. The procedure begins by having each party (confidentially) submit monetary bids for each of the items. After the bids are revealed, each item goes to the highest bidder (with equal bids, a toss of a coin may decide). The received monetary value for each party is shown in *Table 2*.

Table 2. The outcome of the Knaster–Steinhaus procedure

Item	A	B
1	4	18
2	4	9
3	32	9
4	20	6
5	20	18
Total value	80	60
Received value	72	27
50% of total value	40	30
Individual surplus	32	−3
50% of total surplus	14.5	14.5
Transfer payment	−17.5	+17.5
Final value	54.5	44.5
% of total value	68%	74%

Note that this method of allocating items already ensures efficiency. Mathematically, giving each item to the party that values it most is equivalent to maximizing the (unweighted) sum of parties' interests over the possible allocations.[7] The resulting "utilitarian" allocation is characterized graphically in *Figure 2* by the tangency point R between a negatively sloped 45° line, representing the sum of parties' valuations and the efficiency curve.

The objective of the Knaster–Steinhaus procedure is not only *efficiency* but also to give each party a "fair" share of the estate. More specifically, with two parties, each party should receive at least 50 percent of the estate. In our example, 50 percent of the total value for party A or B implies a monetary value of €40 million and €30 million, respectively. By comparing these 50 percent values with the parties' received values from the allocation of items, we see that party A experiences a surplus of 72 – 40 = €32 million, while party B, with a surplus of 27 – 30 = €–3 million, remains below the 50 percent mark. Under the Knaster–Steinhaus procedure the individual surpluses are paid to or compensated by a (virtual) "third party," so that each party now owns exactly 50 percent of the estate, according to its own subjective valuation. The collected aggregate surplus of 32 – 3 = €29 million is then divided equally, paying out to each party a value of €14.5 million. Thus, under the Knaster–Steinhaus procedure "fairness" means giving each party an "equal surplus over an equal share" (see [20]).[8] As *Table 2* shows, the surplus payment and refund policy guarantee an

[7] This is shown by Brams and Taylor [4] and Raiffa [17].
[8] This is the interpretation given by Raith.

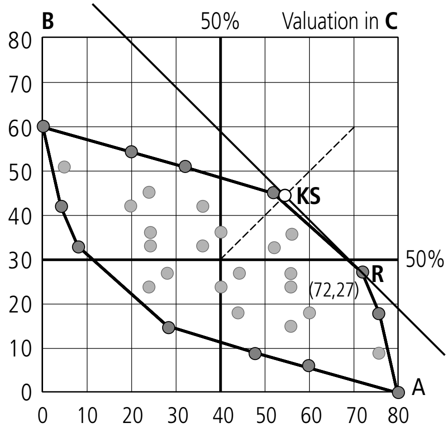

Fig. 2. The outcome of the Knaster–Steinhaus procedure

implicit transfer payment from the party that has received too much value to the party who has received too little. The outcome marked KS is shown in *Figure 1*.[9]

As the final outcome shows, both parties are always guaranteed at least 50 percent of the cake. Because of the point symmetry of the allocations, neither party would want to trade its received portion with the other party. In the two-party case, the Knaster–Steinhaus procedure implements an outcome that is not only efficient but also *envy-free*, meaning that neither party would prefer the other's outcome over its own.

For two-party fair-division problems, where monetary assessments are possible, the Knaster–Steinhaus procedure undoubtedly has attractive features, viz., *efficiency* and *envy-freeness* of the outcome. Moreover, the straightforward transfer scheme is understandable even for negotiators without a technical background, as only differences and shares need to be calculated. Nevertheless, the feasibility of this method must be assessed in practice, and here we simply do not find reports on the success of its use in practice. Although the procedure is easy to understand and the properties of the outcome are tempting for parties searching for a mutually acceptable agreement, negotiators (i.e., the people representing the parties) in real life nevertheless seem reluctant to actually use the Knaster–Steinhaus approach. The steps that negotiators need to follow somehow do not fit the way real people think. The reference points used by the procedure and, in particular, the notion of the surplus as well as the argumentation behind the transfer payment, do not resemble aspects that are actually discussed during negotiations. Apparently, the procedure is too technical to be natu-

[9] Note that point KS is on the 45° line and above the efficiency curve. Although parties value the allocations individually, they are using a common monetary unit of account, which is transferable. The distribution of the surplus thus induces a transfer of value from one party to the other.

rally integrated at some point of the process—it is not clear, where and how it would fit into the interaction between the parties involved in the division problem.

4.2 The market procedure

We can contrast the Knaster–Steinhaus approach with an alternative fair-division procedure, which features specific market-oriented procedural elements that we frequently observe in fair-division negotiations, for example, with practitioners in workshops on negotiation analysis. In their negotiation of fair-division problems, as shown in *Table 1*, parties often try to assess the "real" or "objective" value of the items in order to determine the most appropriate allocation. Moreover, efficiency as well as the desire not to receive less than the other party play an important role in looking for a mutually satisfactory agreement.

The "objective" standard that negotiators typically refer to is the *market*, and the suggested fair division would entail selling items on the market and splitting the proceeds equally. Of course, if this is done with all items, parties cannot expect to receive more than 50 percent of the estate, given their subjective ex ante expectations. As we have seen, 50 percent for each party is generally an inefficient target if parties have differing preferences. Nevertheless, the market price is a strong standard, and negotiators seem to be more worried about the fact that they do not have accurate market information at hand in the negotiation. Therefore, it is quite common for parties to use as an estimate the average of their individual valuations, where a "fair" estimate would typically lie halfway between both parties' bids.[10]

The challenging idea of a *practical* fair-division procedure is to pick up "plausible" lines of argument and to integrate them into a tractable algorithm, for which one can mathematically design desirable properties. We illustrate this with the description of what we label the *market procedure*, which is summarized in *Table 3*.

The market procedure begins just like the Knaster–Steinhaus procedure by having parties confidentially place bids on the individual items. However, after the bids are simultaneously revealed, rather than using the bids to *estimate* the actual market price, parties now consider an internal market, where the market price for each item is determined by parties' individual valuations. Here we simply take the unweighted average of the two bids as the internal market value. The middle column in *Table 3* shows the "market" values of all items, which add up to the internal market value of the total estate.

With the market values defined for all items, the items can now be "sold" to the party willing to pay the market price. Because of our construction of market values, only the highest bidder will be willing to pay the market price. Note that, if all items are sold in this fashion, the resulting allocation will again be efficient because each item is allocated to the party with the highest bid. Of course, each party must pay

[10] The fact that the average subjective valuation of these two parties is rarely a good estimate of the actual market price, is realized only *ex post*, when parties take the items to the market. However, at that point they must accept whatever the market has to offer in terms of payment.

Table 3. The market procedure

Item	A	Market value	B
1	4	11	18
2	4	6.5	9
3	32	20.5	9
4	20	13	6
5	20	19	18
Total value	80	70	60
Received value	72		27
Individual payment	52.5		17.5
50% of market value	35		35
Implicit transfer	−17.5	⟶	+17.5
Final value	54.5		44.5
% of total value	68%		74%

for items that it bought, where the individual payments add up to the market value of the total estate, which is €70 million in our example. As the hypothetical market consists only of the parties involved in the division, the market revenue can be split up equally between both parties, giving each a refund of €35 million. The difference between the payment of and the refund to each party implicitly defines a monetary transfer between parties.

The most remarkable aspect in the present context of procedural design is revealed in the last rows of *Table 3*, which show exactly the same final outcome as in *Table 2* for the Knaster–Steinhaus procedure—this is not a coincidence. Indeed, it is straightforward to show that the Knaster–Steinhaus procedure and the market procedure are technically equivalent in bilateral fair-division problems. The market procedure thus features the same interesting outcome properties as the Knaster–Steinhaus procedure. In addition, its procedural structure incorporates many of the relevant arguments used in actual fair-division negotiations.[11]

The equivalence of the market procedure and the Knaster–Steinhaus procedure results from our specific definition of the market value. More generally, the market procedure will induce an efficient and envy-free outcome for any given market value which lies between the two players' bids. At the low end of the spectrum, the market value could be determined directly by the second highest bid, where only the highest bidder has the advantage of receiving the item. It is interesting to note that the market procedure in this form is identical to the fair-division procedure referred to by Raiffa [16] as the *auction procedure* because it shows some resemblance to a second-price sealed-bid auction. At the high end of the price spectrum, the market value could be determined by the highest bid, so that only the highest bidder will not suffer from buying it. Raiffa [16] refers to this variant as the *naive procedure*, presumably because of its simplicity. The market procedure, as described above, with the market value given by the average bid, can be seen as a compromise between these two extremes.

[11] Several of our workshop participants reported having used a procedure similar to the market procedure in actual fair-division negotiations.

4.3 Multilateral extensions of the Knaster–Steinhaus and market procedures

An important feature of a practicable fair-division procedure is the extension to more than two parties because the necessity of employing a procedure as the mode for joint decision making is most apparent when a large number of parties are involved in the negotiation. For example, when the Convention on the Law of the Sea was finally signed in 1994, 159 countries participated [5]. Moreover, conferences currently dealing with worldwide externalities, such as carbon dioxide (CO_2) emissions, are typically conducted on a large scale, for example, the Kyoto Protocol, mainly because pollution does not respect national boundaries. And environmental catastrophes or severe terrorist attacks immediately take on an international dimension when many insurance companies are jointly responsible for underwriting the economic damage.

Both the Knaster–Steinhaus procedure and the market procedure are readily extendable to fair-division problems involving more than two parties. However, it is interesting to see how the technical equivalence of the two procedures in a bilateral setting vanishes as soon as three or more parties are involved. This emphasizes the differing procedural philosophies behind the two approaches, and it also indicates in which direction one should look for further possible extensions to more complex division problems.

Without loss of generality, consider the allocation of only one indivisible item, say, a deep-sea oil tract, valued differently by three parties A, B, and C. Assume that party A values the item highest at €27 billion, party B at €21 billion, while party C sees the value only at €6 billion. The fair-division according to the Knaster–Steinhaus procedure is demonstrated in *Table 4*.

Table 4. The Knaster–Steinhaus procedure for three players

	A	B	C	
Total value	27	21	6	
Received value	27	0	0	1/3
Equal share of total value	9	7	2	
Individual surplus	18	−7	−2	
Equal share of total surplus	3	3	3	
Transfer	−15	+10	+5	
Final value	12	10	5	
% of total value	44.44%	47.62%	83.33%	

Efficiency is established by giving the item to the highest bidder, party A. As each player is entitled to an equal share of one-third of the subjective value, party A realizes an individual surplus of 18 (the received value of €27 billion minus the entitlement of €9 billion). Since parties B and C receive nothing, their individual surpluses are €–7 billion and €–2 billion, respectively. The aggregate surplus of 18 – 7 – 2 = €9 billion is shared equally, so that party A ends up paying a transfer of €10 billion to player B and €5 billion to player C. As in the two-party case, the final values shown in *Table 4* leave each player with an equal surplus over an equal share, thus confirming the straightforward generalization of this approach to more than two parties. Irrespective of the number of parties n involved, the Knaster–Steinhaus procedure ensures proportionality, meaning that each party receives a final outcome worth at least the equal entitlement of $1/n$.

Nevertheless, as the last two rows of *Table 4* reveal, the final outcomes may be highly asymmetric and thus perceived as unfair. First, note that party C receives a final outcome, which (for this party) is worth over 80 percent of the estate, while parties A and B each receive less than 50 percent in their view, leaving both significantly further from bliss than party C. However, there is a further aspect to consider. Since neither party B nor C receives the oil tract, both are compensated with a transfer payment from party A. By the logic of the Knaster–Steinhaus procedure, the transfer payments depend on the subjective individual valuations. Consequently, the higher a party values an item that it does not receive, the larger the compensation for the perceived loss will be. From party C's perspective, it is only because party B values the oil tract at an "unrealistically" high value of €21 billion, that B receives compensation of €10 billion, which is twice as high as the compensation that party C receives. According to our previous definition, this induces envy, implying that, with three or more parties, the Knaster–Steinhaus procedure establishes efficiency but that its procedural mechanism can no longer ensure envy-freeness.

In searching for a better outcome, the relevant question that now arises is: Which type of envy should be addressed—party C's envy of what party B *has* (viz., more money) or party B's envy of what party C *is* (viz., closer to bliss)? This is not a question that the theorist can answer, but the majority of negotiators (that we have observed and consulted) tend to perceive the first (materialistic) type of envy, focusing on what the other parties actually receive, to be more basic than the second (idealistic) type of envy of how each of the other parties feel about this outcome. For many, idealistic envy becomes relevant only after materialistic envy has been eliminated.

As we have seen, the market procedure, which for two players induces the same outcome as the Knaster–Steinhaus procedure, nevertheless follows a different procedural logic. It is therefore worth investigating how this procedural alternative performs in a multilateral setting. This is demonstrated in *Table 5*.

Table 5. The market procedure for three players

	A	Market value	B	C
Total value	27	24	21	6
Received value	27		0	0
Individual payments	24		0	0
Equal share of market value	8		8	8
Implicit transfer	−16		+8	+8
Final value	11		8	8

The market value is again assumed to be given by the midpoint between the first- and second-highest bids. This ensures that only one party, viz., party A, will be interested in purchasing the oil tract.[12] After party A has paid for the tract, the market value of €24 billion is distributed equally among all three parties, giving each party a payment of €8 billion. Implicitly, this means a transfer payment of €16 billion from party A to B and C, resulting in the final outcomes shown in the last row of *Table 5*. With respect to our definition of "materialistic" envy, one can see that the market procedure leads to an envy-free outcome. Obviously, party A is not envious of parties B or C because it believes that it has received a higher value (27 – 16 = €11 billion). Conversely, neither party B nor C is envious of party A because B perceives that A has received a value of only 21 – 16 = €5 billion, while party C even thinks that A has made a loss (6 – 16 = €–10 billion). Moreover, parties B and C have no reason to be envious of each other because both have received the same compensation.

In contrast to the Knaster–Steinhaus procedure, the market procedure guarantees envy-freeness, and this result generalizes to any number of players as well as any number of items to be distributed. The market procedure not only follows a more intuitive procedural logic but also maintains its characteristics when the numbers of parties involved in the fair-division conflict increases. It is no surprise that the Knaster–Steinhaus procedure has failed to become popular in practice.

From the perspective of procedural design the performance of the market procedure makes it interesting to see if and how the logic of this approach can be extended to more challenging problems of fair division. For example, in situations where n parties must share at least n items fairly, it may appear reasonable to impose the restriction that each party should receive some minimal number of items, say, at least one.[13] Or, if one takes into consideration the content of the items, it may become important to package specific items together, for example, as in a territorial dispute, rather than simply to give each item to the party that values it most. In addition, the scope of fair-division problems can be widened considerably if the items under consideration not only include *goods* but also *bads* (e.g., unpopular or costly tasks that need to be carried out by the whole group). For example, emission permits may be distributed along with specific policy measures for environmental protection or economic development. Or consider insurance cases, where not only insurance premiums need to be allocated but also the coverage of the damage. Haake *et al.* [9] introduce a practical mediation procedure, which naturally generalizes the market

[12] Alternatively, one could take the unweighted average over all bids, yielding a market value of €18 billion. However, selling the oil tract to the highest bidder, party A, will then make party B envious, who would also be willing to pay the market price. For this slightly more complicated initial setting, Haake *et al.* [9] provide an extension of the market procedure.

[13] A popular example of this type of problem is the so-called "rent problem," where several parties wish to share an apartment with as many rooms as there are parties. The task is to find a distribution, where each party receives one room and no party wishes to trade rooms with any other. Several authors have addressed this problems over the past decade, with the general objective of developing a constructive (and thus implementable) proof of the existence of an *efficient, envy-free* outcome. As an alternative, Brams and Kilgour [3] have introduced a fair-division procedure that acknowledges *competitiveness* between players, thus departing from envy-freeness as the relevant notion of fairness.

procedure described above in order to establish efficient and envy-free outcomes for fair-division problems with restrictions of the mentioned type.

4.4 Adjusted Winner

The fair-division procedures we have discussed so far are all based on the condition that parties are able to value the items under consideration in a common monetary unit of account and that they have enough money to transfer value between parties in order to establish fairness. Needless to say, a sufficient budget for participating in fair division is a strong prerequisite. Moreover, limited control over monetary resources may well affect a party's valuation of the items, thus distorting the whole approach. However, the most critical point is that placing a monetary bid on an item may not be appropriate or simply not possible at all. This not only applies to political negotiations but often also holds for pure business deals.

Consider, for example, the international acquisition of the British auto maker Rover by Germany's BMW in 1994, which developed into a major disaster for BMW only a few years later. Crucial items on the agenda for Rover's future were not only the products to be produced but also the choice of production facilities, the composition of Rover's management, the market presentation of Rover/BMW, and the design of new Rover cars. Even if it does not seem immoral, it is often nevertheless inappropriate or just too difficult to assess the importance of individual issues in monetary terms.

The fair-division procedure *Adjusted Winner*, introduced by Brams and Taylor [4], overcomes these difficulties, as it does not require monetary assessments and establishes fairness without monetary compensations. The objective of Adjusted Winner is to establish an *efficient* allocation that is *envy-free* and, in addition, *equitable*, meaning that the final allocation should give both parties the same percentage of the estate.

Table 6. Adjusted Winner

Item	A	B		
1	5	30	30/5=6	3
2	5	15	15/5=3	2
3	40	15		
4	25	10		
5	25	30	30/25=1.2	1
Total points	100	100		
Received points	65	75		
Transfer of item 5	+25	−30		
Adjusted points	90	45		
Equitability	65+α25 = 45+(1-α)30			
A's share of item 5	α = 2/11			
Final points	69.55	69.55		

We discuss this approach using the fair division problem of *Table 1*, but we focus here on panel b), where the relative importance of the individual items is given as a percentage. The assessments are reproduced in *Table 6*. The package of all five items is viewed by both parties as the complete cake, worth 100 percent which is to be divided.

Rather than having parties place bids on the individual items, Adjusted Winner begins by having each party (in seclusion) distribute 100 points among all items, thereby expressing the relative importance of the individual items. Again, there are 32 possible allocations of undivided items. These are shown in *Figure 3*, where the valuations are now given in percentages.

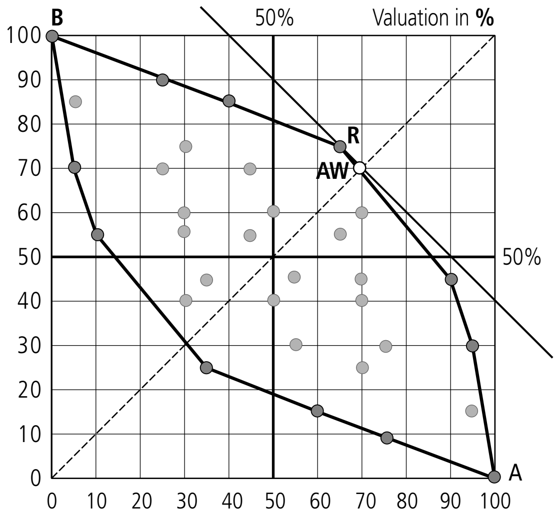

Fig. 3. Adjusted Winner—implementing the Raiffa–Kalai–Smorodinsky solution

When the point distributions are revealed, each item is assigned to the party with the greater relative interest; with equal point assignments, a coin may decide. Analogous to the Knaster–Steinhaus procedure, distributing items in this form ensures an efficient allocation. In *Table 6* one can verify for the example that items 3 and 4 go to party A, which receives a value of 65 points, whereas items 1, 2, and 5 are assigned to party B, so that it achieves 75 points. In *Figure 3* this initial (utilitarian) allocation, marked R, is given by the tangency point between the efficiency frontier and the negatively sloped 45° line, which characterizes the (unweighted) aggregate of both parties' interests.

The next step is to adjust the distribution in order to equalize points. Without money as a means of compensation, the only way to transfer points between parties is to redistribute items. With the initial allocation already efficient, the challenging task is to redistribute items so as to maintain efficiency.

As party B, with 75 points, is the temporary winner, item 1, 2, or 5 has to be transferred to party A. We can calculate for each the loss of points for party B in relation to the gain for party A. When shifting item 1, party B's loss is six times greater than party A's gain; with item 2 it is only three times as high, and with item 5 party B loses only 1.2 times as much as party A wins. Hence, an efficient redistribution would imply transferring first item 5, then item 2, and finally item 1.

In *Figure 3* a movement along the efficiency frontier from one extremely efficient allocation to the next always requires moving only one item at a time. It is never necessary to have two items transferred, and a trade of items between parties is never required. This has important procedural implications because, even if one allows the division of all items, an efficient allocation will never require the division of more than one single item, with all other items fully allocated to the party that values it relatively higher. Indeed, procedures suggesting the "fair" division of every item can typically be expected to induce inefficient outcomes.

With the complete transfer of item 5, party A receives 25 points, yielding a total score of 90, and party B loses 30, leaving it with only 45 points. The complete transfer of item 5 is, therefore, too much for an equitable allocation, implying that this is the one item to divide. The equitability equation in *Table 6* reveals that only 2/11 of item 5 need to be transferred to player A in order to give both parties an equal share of 69.55 points. In *Figure 3* the final outcome is marked AW. As *Figure 3* also illustrates, Adjusted Winner is a procedure for actually implementing the *Raiffa–Kalai–Smorodinsky* bargaining solution, which was introduced by Raiffa [15] and later formalized axiomatically by Kalai and Smorodinsky [11].

Adjusted Winner has received a remarkable amount of media coverage, presumably because it works in conflict situations, where parties are not able to place monetary bids on the items under consideration. This has raised considerable interest, in particular among lawyers and mediators. Nevertheless, aside from instructive examples of hypothetical applications mainly to the divorces of celebrities, well documented in the press, there is little evidence that Adjusted Winner has accomplished a breakthrough in practice. Perhaps this is because of a general misconception of its procedural merits. Indeed, the media seem to focus more on the solution induced by Adjusted Winner than on the procedure itself. For example, one United States (US) magazine commented "One can hire a lawyer and spend years and thousands of dollars fighting, or one can make use of a neat new formula devised by Steven Brams and Alan Taylor," which indicates that the magic of Adjusted Winner is seen to lie in the equitability equation. A German newspaper went even further, referring to the equitability condition as the "formula for peace" (author's translation). If the most important feature of Adjusted Winner is, in fact, the equitability condition, then its application in a real-life conflict requires solving an equation. However, this is a step that most people are reluctant to take in practice. Just imagine a judge trying to justify a verdict with the help of an equation.

One must understand, though, that Adjusted Winner offers more than just the implementation of a specific bargaining solution. The procedure takes parties to the efficiency frontier and shows them how to negotiate efficiently until they reach their desired outcome. In order to illustrate this point, consider the Panama Canal negotiations in 1974, as documented by Raiffa [16]. In this international negotiation there were 10 issues (items) on the agenda, most of which could not be valued in monetary terms. *Table 7* shows the different issues and the relative valuations of both negotiating parties, as perceived by the United States (USA).

Table 7. Panama Canal Negotiations (*Source*: [16], p. 177)

	Items	Units	Range	US	PAN
1	US Defense rights	% to be given up	10–25	22*	9
2	Use rights	Number of rights	20–30	22*	15
3	Land and water	% to be given up	20–70	15	15
4	Expansion rights	Years	5–30	14*	3
5	Duration	Years	20–50	11	15*
6	Expansion routes	Nominal	3 choices	6*	5
7	Compensation	Million $	30–75	4	11*
8	Jurisdiction	Years	0–20	2	7*
9	US military rights	% to be given up	10–25	2	7*
10	Defense role of Panama	% to be given up	10–25	2	13*
				100	100

If the issues in this negotiation are viewed as indivisible items, then this negotiation problem entails a total of $2^{10} = 1,024$ possible outcomes. With Adjusted Winner, each item is first fully allocated to the party that values it most. Hence, the United States' maximum demands are satisfied with issues 1, 2, 4, and 6, whereas Panama achieves full points with issues 5, 7, 8, 9, and 10, leaving both parties with 64 and 53 points, respectively. Issue 3, valued the same by both parties, can be assigned to either one. Assume that it is given to Panama, which is now the temporary winner with 68 points. One should note that this deal is already efficient.

In order to achieve equitability, parties must now move along the efficiency frontier, transferring points from Panama to the USA in the least costly way. Since the first item to be transferred is item 3, it becomes immediately clear that this is the one item to be divided. If 2/15 of item 3 are given to the USA and 13/15 to Panama, the equitable outcome (i.e., the *Raiffa–Kalai–Smorodinsky* solution) yields 66 points for each party, that is, approximately two-thirds of the entire cake for each. An important aspect is also that the proposed compromise on issue 3 is readily implemented in practice. Indeed, if parties have linear preferences over the negotiable range of percentages, then the compromise agreement on "land and water" (issue 3) would require 63.3 percent of the Canal zone to be turned over to Panama when a new treaty was ratified.

5 Multiple-Issue Multiple-Option Negotiations

The fair-division examples that we have considered so far belong to a specific class of negotiation problems, where the individual items constitute the *issues* to be negotiated. Moreover, in a bilateral setting with indivisible items, there are essentially only two *options* to each issue: either party A receives the whole item and party B gets nothing, or the other way around.

More generally, however, the issues in a negotiation can involve anything of concern to the parties, and for each issue, several options may be conceivable as realization possibilities. In addition, negotiating parties typically take as a reference for success not the 50 percent allocation, but determine the gains from negotiation by contrasting possible outcomes with their **b**est **a**lternative **t**o a **n**egotiated **a**greement (BATNA) (cf. [7]). In this section we show how these two important aspects can be integrated into the procedural design.

Consider again a multiple-issue bilateral negotiation, where parties view the individual issues as preferentially separable. As Keeney and Raiffa [13] argue, a negotiation problem with this structure is not as restrictive as it may seem—to a large extent, it is a question of appropriately defining the issues. If each party is able to express its relative concern for each issue by distributing 100 points, the issues resemble the items of a fair-division problem.

However, as Raith [20] demonstrates, by viewing negotiations—for example, the Panama Canal negotiations in 1974, the Camp David accords of 1978, or BMW's takeover of Rover in 1994—as pure fair-division problems with only two options, parties will generally leave value on the table. Inefficiency emerges even if negotiators employ a fair-division procedure. Not that the procedure itself is flawed; it is simply not designed for the underlying structure of the problem. Therefore we allow more than two alternative realization possibilities for each issue. Some may be suggested by one party, some by the other, and some may arise in the course of negotiation. The main aspect is that parties are not restricted only to the two winner-takes-all options, where one party enjoys the whole value of an item while the other receives nothing, as in the fair-division problems considered above.

After weighting each issue, an appropriate way to value the different options to an issue is to give a player's most preferred option the same number of points as the corresponding issue and the worst option zero points. Intermediate options are valued accordingly in between (see [12]).[14] An example of a three-issue negotiation is shown in *Table 8*.

An agreement in a multiple-issue, multiple-option negotiation consists of a package of options—one for each issue. In the example given in *Table 8*, there are $4 \times 3 \times 3 = 36$ distinct possible agreements, each characterized by a triple of options. The value of each agreement to each player can be calculated by adding the points of the individual options contained in the agreed package. In *Figure 4* the 36 agreements

[14] The technique of weighting issues and valuing options is analogous to the method of weighting criteria and valuing alternatives, which Keeney and Raiffa introduced for multi-attribute decision problems.

Table 8. Interest Matrix

		A	B	max sum	A→B	B→A
Issue 1		30	60			
Option 1a	A's best option	30	5			55/10 2
Option 1b	B's best option	20	60	✱		
Option 1c	B's worst option	15	0			
Option 1d	A's worst option	0	30			
Issue 2		10	20			
Option 2a	B's best option	5	20	✱		
Option 2b	A's / B's worst option	0	0			
Option 2c	A's best option	10	10			10/5 1
Issue 3		60	20			
Option 3a	B's best option	30	20		30/20 1	
Option 3b	A's best / B's worst option	60	0	✱		
Option 3c	A's worst option	0	15		60/15	
BATNA		50	40			

are visualized by plotting the values of the agreements for player A against those for player B. The assessment technique described above guarantees that a player's most preferred agreement (for player A it is acb, while for player B it is baa) is worth 100 points, the least preferred is worth zero points.

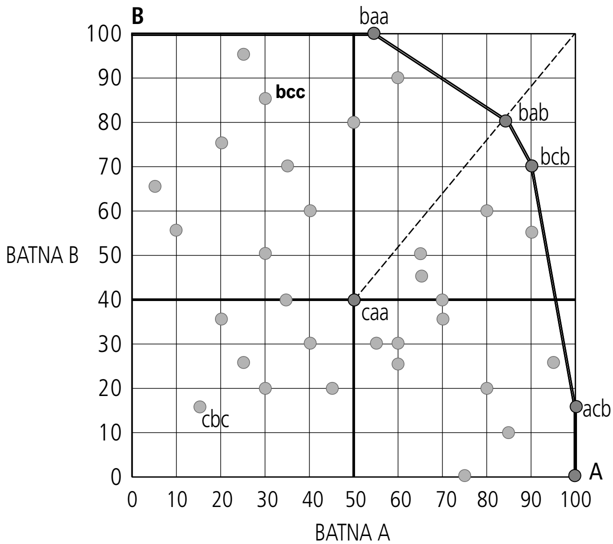

Fig. 4. The $4 \times 3 \times 3 = 36$ distinct agreements of the negotiation problem

The value of players' outside options (i.e., their BATNA) is assessed by comparing them directly to possible agreements in the negotiation. In our example, the outside options of player A and B are both comparable to the negotiation outcome caa, yielding individual BATNA values for parties A and B of 50 and 40, respectively. In *Figure 4*, each party's BATNA separates gains and losses in negotiation. In particular, the area of joint gains is the region of possible agreements lying to the northeast of the intersection of both parties' BATNA values.

The qualitative similarity of *Figures 3* and *4* indicates the possibility of extending our knowledge of fair-division procedures to more general problems of negotiation. Here we take the research perspective of Raiffa ([16], p. 360), who observes that "the need is not for the creation of new analytical techniques specially designed for the negotiation process but rather for the creative use of analytical thinking that exploits simple existing techniques." Since *Table 8* contains nonmonetary assessments of the relative importance of the issues, we continue our analysis with the Adjusted Winner procedure (see Raith [19], [20]).[15]

In a fair-division problem, the first step of Adjusted Winner ensures an efficient initial allocation by maximizing the sum of both parties' interests. By applying the same reasoning to the options in *Table 8*, one finds that the agreement bab yields the highest joint score. As one can verify in *Figure 4*, this agreement is, indeed, efficient. In the second step, Adjusted Winner then redistributes points efficiently from party A to B, or vice versa, depending on the desired solution. The only possibility of shifting points from A to B is a change in the option associated with issue 3, as party B already realizes its most preferred option in the other two issues. As one can easily verify, option 3a offers a lower loss/gain ratio than option 3c. Thus, efficiency requires moving from agreement bab to baa. Party A, on the other hand, benefits from changing options in issues 1 and 2, and here we find that the move from option 2a to 2c involves a lower loss/gain ratio than switching from 1b to 1a. Hence, an efficient point transfer to party A requires moving from bab to bcb and further on to acb. Irrespective of the direction, a movement from one allocation to the next along the efficiency frontier always requires only a single option of a single issue to be changed. From a procedural point of view, once an efficient package of options is found, an efficient modification of this agreement involves only one issue at a time.[16]

Equitability in the present context implies giving each party the same percentage gain over its BATNA value. If we consider the initial efficient allocation bab, the percentage gain for party A is (85 minus 50)/(100 minus 50) = 0.70, while the relative gain for party B is (80 minus 40)/(100 minus 40) = 0.67. Equitability, therefore, requires a transfer from A to B. From the discussion above, we know that this implies

[15] Raith also shows how other procedures for bilateral division problems can be translated into a bargaining-theoretic context, thus making them applicable to more general multi-issue negotiation problems.

[16] Note that the efficiency of single-issue modification does not contradict the common practice of trading options across several issues (in an intertemporal context also known as logrolling), as the latter method is employed to achieve efficiency while the former aims at preserving efficiency.

a linear combination of options 3a and 3b. The precise combination is given by the equitability condition

$$\frac{[20+5+\alpha 30+(1-\alpha)60]-50}{100-50}=\frac{[50+20+\alpha 20+(1-\alpha)0]-40}{100-40}.$$

From this equation one obtains a value of α = 0.036. Thus, the *Raiffa–Kalai–Smorodinsky* solution induces an agreement consisting of options 1b, 2a, and a new option for issue 3, which (in content) consists of 96.4 percent of option 3b and 3.6 percent of option 3a. Given that the assessments in *Table 8* are subject to some uncertainty, parties in practice will presumably find the agreement bab sufficiently close to equitability and not bother solving an equation. Moreover, as the number of issues and options increases, the agreements along the efficiency frontier typically lie closer together, implying that the discrete efficient package of options closest to equitability becomes a satisfactory implementation of the *Raiffa–Kalai–Smorodinsky* solution.

6 Fitting the Procedure into the Process of Conflict Resolution

The challenging question for the negotiation analyst is *where* and *how* the theoretical procedures described in the preceding sections could fit into the actual process of conflict resolution. From the viewpoint of cooperative as well as strategic bargaining theory, conflict resolution implies the specification of a well-defined solution within a fully structured conflict environment. In contrast, conflict resolution in practice consists of several distinct phases of party interaction, of which solution finding is only one and which only becomes relevant after much of the most difficult work of party interaction has been accomplished.

A well-designed fair-division procedure mimics and, thereby, is a substitute for much of the conflict resolution *process*, in particular, the phase in which parties are involved in an actual exchange of arguments and proposals. The procedure chosen by the *actors* is based on a cooperative *strategy* and focused on a *solution*. When applied to the conflict *structure*, the procedure implements an *outcome*. The procedure thus links what Faure [6] refers to as the key components of negotiation.

As we have argued in the case of the market procedure, the acceptance of procedural support depends on how well it fits actors' lines of reasoning, their way of arguing, and their analytical capabilities. The market procedure, in contrast to the Knaster–Steinhaus procedure, automates what parties seeking a fair division would intuitively do on their own. The procedure helps parties avoid mistakes, and it saves them time when many items are at stake—similar to a pocket calculator, which one ideally uses only after having learned the basic mathematics.

Adjusted Winner, in contrast, is an implementation algorithm, which is difficult to conceive of as a substitute for actual party interaction. It operates in a setting in which monetary compensations are also not possible, often because monetary assessments of the issues at stake are not possible. It is unlikely that parties would begin their interaction by placing point bids on the issues, as at an auction. Moreover, after points are assigned, Adjusted Winner simply takes over the process with its own technical logic. Neither the first step, which establishes efficiency, nor the second step of efficient redistribution resembles elements of actual negotiation. In particular, the movement along the efficiency frontier does not mimic a solution-finding process, as the solution (e.g., the *Raiffa–Kalai–Smorodinsky* solution) must be chosen before the procedure even starts.

However, Adjusted Winner can serve as a powerful analytical and mediation tool. In mediated conflicts parties are usually unable to manage their encounter without the support of a third party, and the facilitated encounter typically has an artificial structure, where direct interaction between parties is often replaced by bilateral interaction between each party and the mediator. Before the encounter, the mediator has the possibility of exploring the point assessments with each party separately. Raiffa *et al.* [18] carefully describe the tasks of the analyst in this phase of negotiation. With the well-established scoring technique described above, the mediator has the necessary knowledge of the conflict structure and, with Adjusted Winner, also has a technique of developing an efficient "single negotiation text," which can be adjusted to an outcome with other properties that parties find desirable.

Adjusted Winner not only implements the *Raiffa–Kalai–Smorodinsky* solution but also any other bilateral solution concept. Thus, parties have the opportunity to actually contract a solution together with the mediation process, that is, before they begin to work with the mediator. This is an important aspect to consider, because, as Ackoff [1] argues, parties often seem to have much less difficulty agreeing on general principles, that is, the desirable properties of a solution, than on the means to satisfy them.

Although Adjusted Winner provides analytical support in conflict resolution, the employment of the procedure does not necessarily change the nature of the mediation process, as the individual steps can be performed without extensive technical support. Indeed, even if the complexity of the negotiation problem increases because of additional issues or options, the individual steps of Adjusted Winner require at most the use of a pocket calculator to compute loss/gain ratios.

Crucial for the use of any of the procedures that we have discussed is the acceptance of the quantitative assessment of the conflict structure, which allows the mediator to "separate the people from the problem" (see [7]). The point assessment technique, which serves as the basis for the procedural design, has a long tradition in multiattribute decision theory as well as negotiation analysis, with many case studies documenting its successful use in practice. Nevertheless, the application in conflict resolution requires an analytical training of facilitators and the acknowledgment of its usefulness by the parties involved in the conflict. Although none of the procedures we discussed require graphical support, visualizations of the conflict structure, as we have shown in *Figures 1–4*, often help parties involved in a conflict to develop an awareness that win/win agreements are not at all Utopian goals.

References

1. Ackoff RL (1978) The art of problem solving – accompanied by Ackoff's fables. Wiley, New York, USA
2. Binmore K, Osborne MJ, Rubinstein A (1992) Noncooperative models of bargaining. In: Aumann RJ, Hart S (eds) Handbook of Game Theory. Vol. I, Elsevier Science Publishers, New York, USA
3. Brams SJ, Kilgour DM (2001) Competitive fair division. Journal of Political Economy 109:418–443
4. Brams SJ, Taylor AD (1996) Fair division, Cambridge University Press, Cambridge, MA, USA
5. Brams SJ, Taylor AD (1999) The win/win solution: guaranteeing fair shares to everybody. Norton, New York, USA
6. Faure GO, (2002) International negotiation: the cultural dimension. In Kremenyuk VA (ed) International negotiation: analysis, approaches, issues. 2nd edition, Jossey-Bass Publishers, San Francisco, CA, USA
7. Fisher R, Ury W (1981) Getting to yes. 2nd edition 1991, Business Books, London, UK
8. Gillespie JJ, Bazerman MH (1998) Pre-settlement settlement (PreSS): a simple technique for initiating complex negotiations. Negotiation Journal 14:149–159
9. Haake CJ, Raith MG, Su FE (2002) Bidding for envy-freeness – a procedural approach to n-player fair-division problems. Social Choice and Welfare 19:723–749
10. John R, Raith MG (2001) Optimizing multi-stage negotiations. Journal of Economic Behavior and Organization 45:155-173
11. Kalai E, Smorodinsky M (1975) Other solutions to Nash's bargaining problem. Econometrica 43:513–518
12. Keeney RL, Raiffa H (1976) Decisions with multiple objectives. Wiley, New York, USA
13. Keeney RL, Raiffa H (1991) Structuring and analyzing values for multiple-issue negotiations. In: Young HP (ed) Negotiation analysis. University of Michigan Press, Ann Arbor, MI, USA
14. Knaster B (1946) Sur le problème du partage pragmatique de H. Steinhaus. Annales de la Société Polonaise de Mathématique 19:228–230
15. Raiffa H (1951) Arbitration schemes for generalized two-person games. University of Michigan, Engineering Research Institute, Report M720–1, R30
16. Raiffa H (1982) The art and science of negotiation. Harvard University Press, Cambridge, MA, USA
17. Raiffa H (1996) Lectures on negotiation analysis. The Program on Negotiation, Cambridge, MA, USA
18. Raiffa H, Richardson J, Metcalfe D (2002) Negotiation analysis. Harvard University Press, Cambridge, MA, USA
19. Raith MG (1999) The structure of fair-division problems and the design of fair-negotiation procedures. In: Garcia-Jurado I, Patrone F, Tijs S (eds) Game practice. Kluwer Academic Publishers, The Netherlands
20. Raith MG (2000) Fair-negotiation procedures. Mathematical Social Sciences 39:303–322
21. Schelling TC (1960) The strategy of conflict. 2nd edition 1980, Harvard University Press, Cambridge, MA, USA
22. Sebenius JK (1984) Negotiating the law of the sea. Harvard University Press, Cambridge, MA, USA
23. Sebenius JK (1992) Negotiation analysis: a characterization and review. Management Science 38:18–38

24. Spector BI (1997) Analytical support to negotiations: an empirical assessment. Group Decision and Negotiation 6:421–436
25. Steinhaus H (1948) The problem of fair division. Econometrica 16:101–104
26. Tindale CW (1999) Acts of arguing—A rhetorical model of argument. SUNY Press, New York, USA
27. Underdal A (1991) The outcomes of negotiation. In: Kremenyuk VA (ed) International negotiation: analysis, approaches, issues. 2nd edition 2002, Jossey-Bass Publishers, San Francisco, CA, USA
28. Winter E (1997) Negotiations in multi-issue committees. Journal of Public Economics 65:323–342
29. Young HP (1991a) Fair division. In: Young HP (ed) Negotiation analysis. University of Michigan Press, Ann Arbor, MI, USA, 25–45
30. Young HP (1991b) Negotiation analysis. In: Young HP (ed) Negotiation analysis. University of Michigan Press, Ann Arbor, MI, USA, 1–23

Conclusion: Lessons for Theory and Practice

I. William Zartman[1] and Rudolf Avenhaus[2]

[1] Nitze School of Advanced International Studies, Johns Hopkins University, Washington DC, USA
[2] University of the Federal Armed Forces Munich, Neubiberg, Germany

The focus of this chapter is on the usefulness of formal models to the practice of international negotiations—quite in the spirit of the mission of the International Institute for Applied Systems Analysis (IIASA)—and also to the theory of international negotiations itself, and less on lessons for the formal theory, even though many ideas about what should be done in the theory itself can be found in previous chapters.[3] These conclusions deal with lessons for practitioners of international negotiations such as businesspeople, diplomats, and politicians, as well as providing lessons for modelers who want or have to develop formal models for the purpose of analyzing past negotiations in a descriptive way or, more challenging, for providing formal aid to actual negotiation.

1 Lessons for Negotiation Theory

Lessons relevant to negotiation theory can be sought from formal models *for*, and even more from models *of*, international negotiations; models *in* international negotiations, by their nature, do not claim to contribute to negotiation theory, although they do provide some helpful illustrations of relevant concepts. Individual research findings are highlighted in O'Neill's chapter, which revaluates the dispute over the usefulness of rational choice models in *International Security* in 1999, mentioned in the introduction to this publication. O'Neill provides many examples of lessons to be learned in various areas from formal modeling. These include such topics as the interaction of domestic politics in international politics, deterrence and escalation, bargaining over the initiation or termination of war, and the role of honor, prestige, face, and symbolism. A number of special effects in negotiations have been brought to light in formal studies, such as the role of audience costs [9], honor [19], trust ([11], [12]), and internal factional dynamics ([5], and Gabbay's chapter).

[3] Discussion of theoretical improvements is left to the many publications that describe trends and necessities of theoretical negotiation research (e.g., [16]).

In a summary of compound efforts within formal modeling of peacemaking to confront key objectives of the broader research program, Carment and Rowland evaluate the accumulation (building on previous findings and modifying or discarding empirically unsupported arguments), integration (drawing on alternative methodologies that provide similar findings in a different context), and synthesis (using a multilevel analysis approach) of research, finding some progress in accumulation but little or none in integration and synthesis. Nevertheless, they conclude optimistically that "... it might be suggested that formal modeling could serve as a panacea to the problems identified by the editors of this publication, because it offers a more explicit, concise, and dynamic account of the negotiation and intervention process." Indeed, formal studies have brought out some broader insights into the negotiation process on the basis of their own concepts, such as the presence of multiple equilibria or salient solutions, as a cause of escalation, as discussed in Wierzbicki's chapter (see also Zartman and Faure [28]), or of unrealized or unclaimed gains, as a spur to further bargaining ([3], [14]).

2 Lessons from Models *of* Negotiation: Strategies

A major conceptual lesson for theory from formal modeling *of* negotiation is that the appropriate game-theoretical tool for modeling negotiations is their representation in extensive form, usually performed graphically as a tree describing the strategic possibilities of the players from the beginning until the end of the "game." It has been shown (see also Avenhaus [1]) that the elements of this representation exactly match the analytical framework of negotiation theory developed by the Processes of International Negotiation (PIN) Program. This tool nicely describes simple negotiations, but also permits the analysis in words of negotiations that are too complicated to be analyzed informally.

Another lesson is the general principle that, if one wants to keep a quantitative analysis tractable, technical intricacies quickly limit the number of assumptions and the complexity of the set of rules, and therefore formal methods and models are by nature incomplete tools for representing real life issues, as discussed at the end of Rudnianski and Bestougeff's chapter. Such models do not encompass the nuances and, even more importantly, they do not take into account the fact that real players are most often irrational. Similarly, Wierzbicki argues that "in more complex situations, characterized by nonlinear or multicriteria models, equilibria are essentially nonunique" and he continues that game theory "concentrated rather on arguments for choosing a unique equilibrium as a rational outcome of a game." One can turn the argument around in a way that may be helpful for negotiation theory: the more criteria one takes into account, the more equilibria exist, which renders the agreement on one equilibrium more difficult. Purely formal arguments cannot solve this problem. As an example, consider Dostoyevsky's description of Raskolnikov's negotiation with himself, as has been analyzed in game-theoretical terms [7]. The existence of technical intricacies and multiple equilibria can be considered as an indication of a difficult negotiation problem that cannot be solved with formal arguments, and there is no simple answer to these challenges. But the way in which Wierzbicki develops his argument and Rudnianski and Bestougeff develop their models, proceeding from the simple to the sophisticated, indeed represents a contribution to this basic debate.

As current game-theoretic and rational choice literature emphasizes, better information is an important aspect of reaching agreement, a point shared by analysts of models both *for* and *in* negotiations ([9], [21]). But “in particular in strongly competitive situations, decision makers are not inclined to reveal their preferences,” as Wierzbicki notes. Formal theory has developed methods to deal with this problem, but, at least according to our understanding, its contributions have not yet been acknowledged properly by the nonformal political science community. A situation where one of the two antagonists does not know which “type” (hard or soft) the other is, can be modeled. As Güner shows, the consequences of the situation from the model can be derived. The graphical representation of the resulting equilibria indicates that this analysis cannot be performed with nonformal tools. In the workshop discussion of this issue, a striking example was given from the field of arms races and escalation ([15], [2]): in a confrontation between an aggressor and a defender, there will be no escalation if the aggressor knows whether the defender is hard or soft, but there may be escalation if the aggressor does not know the type of his adversary. However, there is other literature that discusses why weaker states attack stronger states even when they know them to respond emphatically. The relationship to negotiation, however, remains tenuous.

3 Lessons from Models *in* Negotiation: Process

Models *in* negotiation may initially limit themselves to simply providing useful data about conditions and proposals and their implications, as the studies of Regional Air Pollution INformation and Simulation (RAINS) and the United Nations Conference on the Law of the Sea (UNCLOS) models by Amann and Antrim show. Such models are even more useful when they incorporate negotiations, concepts, and practices and thus participate in the process itself, such as focusing on a formula and its use to harmonize the parties’ interests and bring the parties’ positions together in an agreement. The methodological approach presented in the chapter by Druckman emphasizes process versus outcome, one of the major themes of this book. Four modeling approaches are discussed: Negotiation Support Systems and Stochastic Modeling being process-oriented, and Decision Analysis and Game Theory being outcome-oriented. It is argued that their combination can help to address a key question in research on negotiation: What is the relationship between process and outcome?

4 Lessons from Models *for* Negotiations: Outcome

Models *for* negotiation are, in fact, alternatives to negotiation, but they carry some powerful lessons of their own. As they purport to present optimal outcomes and offer procedures to do so, they can provide outcomes that serve as targets for the negotiating process against which negotiated outcomes can be judged. Indeed, the basic value of the first such effort, the Nash [17] point, was to show that under specified conditions one Pareto-optimal outcome was possible, not that negotiations would or even should achieve that outcome. Yet the various attempts at fair division and the various efforts to define equality provide a goal toward which negotiators can aim and at the same time a spirit of sharing and mutual satisfaction—an ethos of "envy-freeness"—that would be a useful ingredient to introduce into any negotiation.

In addition to the goal and ethos, models *for* negotiation incorporate, and even depend on, some concepts already established in the negotiation process, confirming their utility from a more abstract point of view. These include packaging in Brams *et al.*'s [4] fallback bargaining (FB_u), successive intermediate positions—"a string of pearls"—between two equilibria in Schüssler's Adjusted Winner (AW) model, and Raith's use of AW as a single negotiation text (SNT) "which can be adjusted to an outcome with other properties that parties find desirable." The point is that these model solutions require tinkering as they approach the real world of negotiation and that they suggest and support various ways of doing so.

Negotiation is exorcized by models *for* it. It was explicitly excluded in the original Nash [17] article by an "equality of bargaining skill," later clarified [18] to mean an absence of bargaining skill: "there should not be any question of bargaining ability." Since there is a major question of bargaining ability in negotiation, the challenge remains as to how to achieve optimal outcomes *through*, not *for*, negotiation. Negotiation needs to be exercised rather than be avoided. As Antrim's chapter emphasizes, the decisions to compromise using the modeled outcomes as a starting point (or SNT) is the job of the skilled negotiators.

Models *for* appear to assume that negotiation is a bad, unnecessary, or undesirable thing, and in a certain sense it is. Negotiation is most inefficient; like democracy, it is wasteful in time and effort as a way of making decisions. However, like democracy, as Churchill implied, negotiation contains a number of values not attainable in any other decision-making procedure, and that are of such importance that they ultimately justify or outweigh the inefficiency of the process. One such value is found within the modeling process; others are found outside. The internal value lies in the need to overcome the multiple equilibria or the various proposals for fair allocation, each of which has strong procedural and substantive reasons behind it. The effort to find the best or the fairest is important, but for the time being, such model outcomes are best seen as starting points for serious negotiation in a search for the most satisfactory in the string of pearls or the best adjustment in the SNT.

The external value lies in a number of elements that cannot fit into a model but are significant side benefits of negotiation. These include the three "ships" of negotiation—relationship, ownership, and craftsmanship—as well as the value of education. They need to be explained, if only briefly. The value of the relationship between the parties is a hidden element in negotiation and cannot be subsumed under an ethos of cooperation. In some cases it outweighs the apparent stakes at issue, while in others it is negligible. More constant is the emotional value of ownership, a crucial element is making the final agreement stick. The parties must themselves feel and be able to make their home teams feel that they fought hard for what they got and gave, that they did their best in defending the home turf, and the fairness of the outcome in itself is not enough to outweigh the absence of a good effort to achieve it. Craftsmanship refers to the hugely important element of creativity required in negotiation to provide inducements, negative and positive, as added value to the innate value of the stakes, and to consider alternatives to the proposed outcomes on the table, both as security points or best alternative to negotiated agreement(s) (BATNAs) and as improvements to standing offers. The first two "ships" are not in the models; the third is either absent or provided beforehand by an inhumanly infinite list of possibilities. Finally, negotiation is persuasion and persuasion is education, about the stakes and about the parties. The negotiation process itself can generate a better understanding of the issues and of one or other of the parties and may thus modify preferences and thereby the actual structure of the game. Models *for* negotiation may provide starting points for the insertion of these values but they cannot contain them.

5 Lessons for Negotiation Practice

The first concern in discussing lessons of formal modeling for practitioners is the general issue of usefulness and accessibility. A short story may be relevant here. Two university departments have begun a search for someone to fill a new position, and have interviewed a number of candidates. The two search committee members sit down to evaluate them. At Formal State University (FSU), a list of criteria weighted by their importance is drawn up and the candidates evaluated numerically; when the committee calculates the winner, they ignore the system and pick their favorite candidate, bypassing a number of the established criteria. At Informal State University (ISU), committee members maintain their priorities but argue the merits of each candidate on the established criteria and reach a deadlock on different candidates, exemplifying different priorities. Of course, these stories have no correspondence in academic politics and are purely fictitious. It is not the outcome that is the point of the story but rather the fact that formalizing has the virtue of making explicit what we carry implicitly in our heads, with the drawback of requiring constraining quantification. Its greatest benefit may be to clarify the output consequences of given inputs—both preferences and characteristics—as a spur to reconsideration.

This section of the conclusions evaluates the formal modeling *of*, *for*, and *in* negotiation in the eyes of practitioners. Those two eyes focus on two sides of the model, asking whether its findings are insightful and helpful for the practice and whether they are clear and transparent to the practitioner. The first is useless if the second does not obtain, and the reverse is also true. The judgments will differ according to the models' relationship to negotiation. Models *in* negotiation can be seen as making important contributions to the diagnosis and formulation of negotiators' positions; models *of* negotiations are user-unfriendly and of only indirect relevance; and models *for* negotiations, as the name implies, describe optimal outcomes but say little about the process of achieving them.

Models *in* negotiation are dynamic, heuristic mechanisms presenting external data that can inform the process of negotiation. The two examples given above on pollution control and the law of the sea illustrate the type of work done and its relationship to the negotiation process. The RAINS model developed at IIASA combines data on emissions, dispersion, and impact, and then evaluates the cost–benefit effectiveness of various control measures in relation to targets set by the negotiations. Initially, the data were useful in the diagnosis phase of the negotiations as informational inputs into the development of the parties' understanding of the issues and positions on them. The model then becomes part of the formulation phase, as modelers and negotiators work together to identify cost-effective paths to meet the negotiators' targets. The same role of the model in the negotiating process was evident in the use of the Massachusetts Institute of Technology (MIT) model of deep-sea mining in the UNCLOS negotiations. The data on design, distribution, pricing, and taxation were run to show the profitability of the various proposals for a new mining regime for underwater minerals. After premature attempts at formulation without information had failed, the model was used for diagnosis and education and then became the basis for the development of several formulas for a new tax regime and eventually for a negotiated agreement. As purveyors of needed data in a dynamic form, the models provided insights useful to an improved negotiation process.

The models for negotiation do not simply provide information but also carry with it a requirement. Initially seen as a "black box," both the RAINS and the MIT models demanded additional interaction between modelers and negotiators to bring out their full utility. Not only did the modelers have to instruct the practitioners about their new tool in order to overcome suspicions about mysterious mechanisms inside but they also needed to hear practitioners' reactions and corrections in order to perfect their own work. Thus, the first criterion, usefulness, was dependent on the second, transparency.

Models *of* negotiation, which seek to replicate the work of practitioners in formal language, are—perhaps counterintuitively—necessarily inhospitable to practitioners. It is the models' language that causes the disjuncture: formalization depends on a sparse, controlled encapsulation of the diplomatic process in abstract or symbolic—and sometimes mathematical or quantitative—terms. It is a translation from one language to another, from the language of political analysis to that of symbolic analysis, and has no more prima facie comprehensibility than the translation of an English author into Chinese would have for the author. Of course, if the author were also fluent in Chinese, he or she could check the translation and learn some more Chinese in the process, but that would require a high degree of bilingual fluency and would not make the original text any better. This is a different matter from the transparency developed in the previous set of models, those *in* negotiations, which involve an external source of data, not a replication of the process itself in new and unfamiliar terms. It is the language and terms of analysis themselves that allow the modeler to perform his science, and this is different from the practitioners' art. The necessary nontransparency reduces the model's usefulness for the practitioner.

The same question asked *in* negotiations is then posed again: Would a greater familiarity with the language bring the practitioner new and necessary insights into the process from models *of* negotiation? It is only half accurate to say, as Rudnianski and Bestougeff do, that game theory simply repeats what everyone knows in a language no one understands—and one might add, "in a misleadingly certain way," as will be discussed below. The language is (as all languages are) both a vehicle and a problem, but a constant dialog between the linguist and the practitioner would be only a partial key to the removal of the problem. Such a dialog is, of course, one of the goals of this book, but it is not self-compelling. It is necessary first to stimulate the practitioner's curiosity by pointing out some counterintuitive implications. So piqued, a few intellectually curious practitioners may then be impelled to find out where the insights come from—to be intrigued enough by the poetry in translation to want to learn the original language, to continue the metaphor. Such practitioners are rare, and perhaps such counterintuitive insights may too be hidden underneath the things everybody knows.

The models *of* negotiation are torn between two fates: in their own Prisoner's Dilemma, so to speak. They can either explain what has occurred, using their analytical language or they can provide new possibilities and insights, showing better options than those chosen. The strength of their counterintuitive findings varies. The cognitive model presented here develops some implications for ripeness theory and broadens its application. The design of evolutionary systems underlines the importance of cultural awareness and opens the possibility of finding positive-sum solutions. Some game-theoretic analysis makes the value of maintaining a coalition of the great powers obvious; that option may be intuitively attractive but not a solution, since it itself is the problem.

However, some of these models compound their communications problems by not seeking counterintuitive implications at all but rather by limiting themselves to using their chosen analysis merely to confirm what actually happened. Rudnianski and Bestougeff's quote is again only half-right; it is not that models can explain everything, just everything that happened, although they can explain how or why what did not happen did not, which is after all what all good theories seek to achieve. Thus, game-theoretic analysis shows that ground war was not an option in Kosovo but that an airstrike was indicated. Such implications can be seen as important to the development of the language, even if they do not provide a new view on negotiation processes and possibilities. Their value is for modelers, not for practitioners.

To produce such insights, models *of* negotiation tend to focus on a particular parameter in their simplified presentations of the negotiation process. Among the game-theoretic presentations, for Rudnianski and Bestougeff the focus is preference ordering; for Avenhaus and Krieger, it is Pareto optimality; for Güner it is reaction types; and for Gabbay it is coupling values.

Models *for* negotiation create for themselves a different challenge if they seek to attract practitioners. Their avowed purpose is to bypass negotiations, to create a more effective and efficient process that eliminates politics and persuasion,[4] and instead

[4] "Efficiency" as used here refers to process efficiency, taking less time, rather than the economists' notion of outcome efficiency or Pareto optimality, leaving no gains on the table.

pulls a rabbit out of the computer, given appropriate inputs. The attractiveness of their solutions—as seen in the current wave of fair division proposals—is that they avoid the messiness of strategic thinking and its cost in time and friendships, by confronting conflicting parties with the implications of their preferences. Rather like the debate over justice in negotiation [27], their inconclusiveness is shown by the fact that there are so many solutions, each purporting to produce the fairest division, so that in the end the outcome requires another return to negotiation.

The attractive quality of such models is their precision. This makes them useful *ex post* and even in *media res* as illustrations for showing implications and target solutions, much as models *in* negotiation are used. But it is this very precision that makes such models inapplicable to negotiation as a process. As Raith shows very well, process is different from procedure; fair-decision models, for example, have a procedure of their own, of little relevance or similarity to the political process of negotiation. This is a difficulty shared by other economic treatments of negotiation such as concession/convergence theories of bilateral monopoly developed in the classical work of Edgeworth [6], Zeuthen [29], and Hicks [10], all tight theories based on the assumption of a specific starting point with fixed incremental moves. The problem is not the assumption of fixed points per se but the fact that the assumption either inhibits acts of creativity, such as packaging and unpackaging, prospecting, and reframing, and that it further assumes that all the processes of creativity have already been accomplished in a sterile—that is, a political—atmosphere.

Models *for* negotiation are Boulwarism, named after the tactic of the General Electric executive who presented labor with "fair" proposals for wages that would obviate the time-wasting process of negotiation. But his methods ignored the need for ownership that negotiation meets and so were roundly denounced by labor, even though they produced as good results as—and perhaps better than—those obtained by negotiation. Just as elections are an expensive and time-consuming process of picking the best candidate, so are negotiations less efficient but more satisfying.[5]

The same criteria for the models' attractiveness to practitioners—transparency and the ability to generate insights—apply to models *for* negotiation. In many of these models, once one starts to take apart the black box, one opens the process to negotiation rather than obviating it. If the workings of the black box are accepted, their outcome can be useful as a measure against which negotiations can be judged. A third scenario would be to turn to the model after negotiations have become deadlocked and to use its outcome as a goal and spur to the political process. In all these uses the models become adjuncts to bargaining, entering somewhere in the process much as models *in* negotiation do, providing external information to the enlightenment of the bargainers.

The fourth scenario would simply use the models as intended, to replace the messy inefficiency of bargaining. Such use is much more likely in personal

[5] The notion that technical solutions obviate politics is an error common both to law and economics, as seen inter alia in the right of self-determination which encourages the use of the violence that it is designed to supplant, and in the principle of equitable representation that encourages the gerrymandering it seeks to replace.

encounters where the basis of trust and familiarity is present, even if not in evidence at the time. Inheritance divisions are often cited. Raiffa, with his avuncular manner as a vehicle for confidence building, claims great success in avoiding costly negotiations for his consultees through fair-division models. It is doubtful if avuncularity would carry the day in international conferences or even in collective bargaining, and in the two known instances where multivariate utility (MAU) analysis was used in international negotiation, it was rejected as redundant in one (Panama Canal negotiations) and used as an adjunct in the other (Philippines base negotiations), and the experiment has not been repeated in the decades since [21].

The workshop discussions raised the case of negotiations in the European Union (EU) as a setting where formal models of the *for* type would be helpful, where parameters that were the basis for initial negotiation positions changed but options were assumed to be known beforehand. In this dynamic situation formal models could have helped to quickly analyze new options. Since the EU expects unanimity on most issues, states have to analyze fall-back positions before the start of the negotiations and, given the number of possibilities, formal models could help to organize positions.

The problem is a two-level game of transparency. The model *for* negotiations faces an even greater problem of acceptance by negotiators, as it either replaces them entirely with a "machine" or—as in the case of models *in* negotiations—provides important information. It must therefore inspire enormous trust for its procedure if it is to replace the regular negotiation process, a trust that itself will require some degree of process somewhere down the line. As noted, that degree of trust is conceivable in disputes over technical matters between close partners who share a deep relationship and similar status and culture. Negotiations in water-allocation disputes among otherwise close and similar partners may provide an example. Power inequities, cultural differences, and high stakes all act to reduce the acceptability of a technical procedure for negotiators.

Added to these problems is the "reentry problem," as Raith notes, already a great challenge to negotiators returning home to explain what they have gained and what they have paid. It is difficult enough for negotiators to sell their political compromises to a waiting audience; it is even more difficult to explain that these compromises were made by a computer. Model solutions are axiomatic, specifically excluding strategic interaction and process ownership and relying on the fairness argument that each party has done the best it could, given that the other party has done its best too. That is win–win at its best, and the best payoff of models *for* negotiation may be that they instill the idea of joint gains fairly distributed. But as their acceptance depends on the prior acceptance of that idea, there is a daunting circularity that limits their use.

In sum, models *of* negotiation are inherently more useful to modelers than to negotiators, explicating the process in their chosen terms of analysis and thereby strengthening that analysis for further understanding of the negotiation process. Models *in* negotiations provide a useful input into the process, combining insight with transparency in a self-reinforcing relationship. Models *for* negotiation are most helpful when used *for* negotiations rather than replacing them, as a reference outcome and a stimulus to creative thinking. Under some conditions, such models may even replace negotiations, much as regimes (norms, rules, principles, regulations, expectations) in anarchic systems and laws in sovereign systems now do on some issues and in some sectors [26]. Until we reach that state the world around, models need to find their place alongside the inefficient politics and strategic interaction of negotiation that gives parties a direct hand in shaping their fate. While developing finer machines, we need also train a finer hand.

6 Lessons for Modelers

Finally, there are lessons for formally trained modelers who want to help solve real-world problems. Such lessons emerged from the previous chapters and from the workshop discussions around the other chapters of this book. Six principal lessons are presented in the order in which they would arise in the modeling process.

Lesson 1: Understand the problem

The modeler, whether he is undertaking a technical presentation of an effect used in negotiation, a descriptive analysis *of* some past international negotiation that is of interest to a historian, or a normative study *for* an ongoing or future negotiation about which a diplomat or politician is asking the modeler, must above all understand the problem. Not only does he have to know those elements that enter into his model, such as parties, strategies, and payoffs, he also has to know the background of the issue to be negotiated, its hidden incentives, the alternatives to a negotiated agreement, and the short- and long-term consequences of potential results.

Even though the final result may be structurally simple, experience shows that in general there is no way to get quick and direct information that is relevant to the model from the practitioner because the practitioner does not know what is or is not relevant. He knows what he wants to know, but he does not know what information is needed for the answer. The modeler must spend time and needs patience to learn about matters that otherwise would either be not interesting to him or directly relevant to his approach.

When a formal modeler publishes his results, he spends little time justifying them in terms of real applications, and the examples in this publication are no exception. The same thing is found in published proofs of mathematical theorems that are stated in their final elegant and parsimonious form, without reference to the twisting paths that led to their discovery. Occasionally, cases can be found in the literature that provide more information; an example is Bueno de Mesquita's [5] formal analysis *of* the negotiation of the Concordat of Worms in the year 1122 where it is demonstrated how a detailed and elaborate description of the history leads to a concise, if not simple, model of the negotiations.

Lesson 2: Justify assumptions

Once the first draft of some model is formulated, its underlying assumptions have to be discussed with the practitioner, whether a historian or a diplomat, and such a discussion may lead to considerable revisions as assumptions that are important for the model may be difficult to justify from the practitioner's point of view.

An important example is found in several chapters: as discussed by Avenhaus and Krieger, any game-theoretic model is based on the assumption that the tree or normal form, that is, moves, process, structure, and payoffs is known to all parties. In many real cases this may not be true, as noted in Schüssler's chapter, where even the possibility of unintentionally false representation of preferences is discussed. If one party does not know or has only partial knowledge of the preferences or payoffs of the other, this fact has to be formulated explicitly, but here again this limited knowledge has to be common knowledge. An example of such a case of partially incomplete knowledge is Güner's analysis *of* the Greece–Turkey Aegean waters conflict.

When assumptions have been clarified and their consequences analyzed, the result may often look trivial—in one of our internal discussions, one of us exclaimed, "But your conclusions just repeat your assumptions!" When models become more complicated this will happen less frequently; however, more trust is needed from the side of the client of the model. Thus, it may be wise for the theorist to start with models that may look trivial in order to build up the confidence needed for his more elaborate versions (assuming he is able to keep the practitioner's attention during the trivial phase).

Lesson 3: Find the method

As has been noted many times in this book and elsewhere, formal models cannot predict negotiation processes. They cannot even describe the real courses of negotiations that depend among many other factors on the personalities and their sympathies and antipathies at the negotiation table as well as on unforeseeable outside events that cannot be modeled formally. Their usefulness depends on the choice of the method and within it the choice of the relevant degree of sophistication of the method to be used. This choice will depend on the questions that are expected to be answered. The model has to be tailored to the questions; the old saying "Don't crack nuts with a sledgehammer" holds. In other words, if one develops an ambitious model with many alternatives and a complicated information structure, one should have good reasons to justify the complexity. Major parts of Rudnianski's chapter are devoted to this basic modeling problem, as seen in his discussion of the assumptions for the analysis *of* the Spratly Islands conflict.

Lesson 4: Explain the method

It is difficult to explain mathematical methods to practical people who have little time and less patience. While there is no point in explaining abstract formulations, long derivations of theorems, or computational details, the practitioner needs a basic understanding of the kind of reasoning that makes up formal models. The modeler may start with a simple model, as, for example, an extensive form game with two parties and only two alternatives for each party, but he runs the danger that the listener will dismiss the whole enterprise as trivial. In fact, in simple cases the analysis can be performed purely verbally—the model is just a simplified way of presenting the problem. More parties, more moves, even some information sets will no longer be considered trivial if the listener has the patience and motivation is listen.

In addition, the idea of a Nash equilibrium in pure strategies may be explained and understood in general terms, but it is already more difficult to explain the concept of mixed strategies and of Nash equilibria in those mixed strategies. And it is nearly impossible to explain the problem of multiple equilibria and to convey an understanding of the fact that this need not be a deficiency of the theory. Equilibrium selection is still a major issue in the development of the theory itself, but whatever progress is made, the fact remains that sometimes multiple equilibria are just a reflection of real-life situations. On the other hand, it is easier to discuss whether payoffs to parties in the negotiation can be represented as scalar ones than whether vector payoffs have to be used, making the analysis much more complicated.

There is no royal path to the solution of these problems, as Raiffa has often noted. The modeler has to recognize that there is a problem that he has to face and that its solution demands considerable pedagogical effort and ingenuity.

Lesson 5: Explaining results

Nobel Prize Laureate Selten [23] describes his experience in the 1960s when he worked for the United States Arms Control and Disarmament Agency (ACDA) on international relations: "I had the impression that the practitioners did not really listen to anything applied. They felt that they knew best how to solve their problems. They were much more pleased by qualitative interpretations of abstract game-theoretical results." The later Nobel Prize Laureate Aumann gave an excellent talk of this kind, explains under what conditions it is advantageous to reveal or not to reveal information by strategic actions in repeated games with incomplete information. We were told that, after Aumann's presentation, a senior official said to his entourage: "We know these things . . . but we like to be reminded." He adds, however, that he did not feel discouraged but continued his work in international relations. Together with his Israeli friend, Perlmutter, he developed the so-called Scenario Bundle Analysis of constructing simple game models on the basis of expert judgments and applied it first to the Persian Gulf Conflict and later, together with Reiter, to the Kosovo conflict (see [22]).

There is the problem that the practitioner may consider the results either to be trivial or known or not interesting. What can be done? As always, the answer depends on the specific problem. As Wierzbicki pointed out already, and as illustrated by Avenhaus and Krieger in their discussion of the Rambouillet negotiations, vector payoffs present a special problem as they usually generate multiple equilibria, undermining the possibility of predictive or normative analysis. Yet these results need not be useless as they sometimes tell us what can *not* happen and why.

A rather general solution to the problem of explaining results is any kind of interactive procedure, already contained in the second and fourth lesson and leading to the sixth and seventh: instead of just explaining assumptions, method, and results, one should try to teach the practitioner to use the model himself, perhaps by computer, perhaps with appropriate assistance. If he can change basic elements of the model himself and observe the results, he may get a feeling for the approach and thus develop the confidence needed for constructive cooperation between theory and practice. Güner has stated this very convincingly at the end of his analysis *of* the Greek–Turkey Aegean waters conflict.

Lesson 6: Revise and validate

Some of the formal models presented here contain a method for revision as a constitutive element, like the fair-division procedure. Any interactive procedure foresees the revision of assumptions once the results are known, Selten's Scenario Bundle Analysis provides a good example. There may even be a danger that assumptions are changed when the results that the practitioner would like to see have not yet been obtained. Intentional false representation of one's own preferences may even happen, as Schüssler points out in his analysis *for* a Camp David scenario. Nevertheless, a procedure for revision in one way or another is mandatory for any serious modeling effort.

Validation is more complicated, and we have to distinguish between the different purposes of formal analyses described in the introduction to this volume. What does validation mean in the case of descriptive studies *of* past negotiations whose outcomes are already known? Here, as we understand the problem, validation means validation of hypotheses, and hypotheses are assumptions that may or may not lead to the modeler's results as confirmed or not by history.

In the case of a predictive analysis, validation is simple if there is enough time to wait for the outcome of the negotiation. Okada in his chapter presents an elaborate analysis *of* the international negotiations of the Kyoto Protocol, but he sees the importance of the model not only in its predictive capacity but also in the structural insight it provides. The author is modest, although we are convinced that his model could also be used in negotiation.

In a normative analysis *of* and *for* negotiations, the situation is again complicated: Was the analysis successful because the envisaged results were reached, or was it unsuccessful because the results were not reached but successful in showing results that should/could have been reached? Perhaps at present these are artificial questions, as in the near future it will rarely happen that at an important international negotiation a modeler's binding advice, based on his formal model, will be asked for and used. Much more realistic is the question of how models *in* negotiations are validated. With the chapters on the use of IIASA's RAINS model in the negotiations on Amann's Convention on Long-range Transboundary Air Pollution and on the Massachusetts Institute of Technology (MIT) model of Antrim used in the negotiations during UNCLOS, two valuable examples emerge: both models were considered as tools that provided guidance to the negotiators in searching for acceptable compromise solutions, and here the answer to the validation question is very simple: both models were validated by being accepted as negotiation tools by all parties to the negotiations.

7 Conclusion to the Conclusion

This review has taken a comprehensive view of the relationship between formal modeling and negotiation in an effort to ascertain their mutual values. Modeling has its own justifications that are not necessarily dependent on its practical value, and its application to negotiation has an impact on its own development (which has not been greatly explored in this publication). What has interested us more is the effect of modeling on the analysis and practice of negotiation. The time may come when modeling can grasp all the complexities of the negotiation process, with all the side values—including the three "ships" and education—that those complexities contain. For the moment it does not, but the effort to model realistically and to provide satisfying alternatives to the inefficient process of negotiation may in time produce a methodology that understands and replaces the negotiation process. Models *of* negotiation have their self-justifying value in this effort. Only by sticking to that effort can fully effective results be obtained.

There is no doubt about the usefulness of models *in* negotiation, and one can only hope that further models can be developed to help the education process that negotiation involves. Not every subject can be modeled in this way but more can than are currently being presented. Models of processes, costs, and benefits could conceivably have been of use in such disparate negotiations as those over civil conflict in southern Sudan and the Ivory Coast, additions to the Kyoto Protocol to the Framework Convention on Climate Change (as discussed in part by Okada), rectifications of the International Treaty on Endangered Species (ITES), and the many ramifications of the Doha Round of the World Trade Organization (WTO) negotiations.

It has been frequently noted (by O'Neill among others) that the real benefit of modeling is to develop a way of thinking, and it is certain that anything that encourages better reasoning and creativity is a welcome contribution to the negotiation process. But formalizing that contribution has its costs. The way of thinking involves a series of binary or multiple choices among fixed alternatives. While an explicitly formalized process can improve reasoning, it can also stifle creativity. If the given alternatives are too constraining, the name of the game is to create new ones, not to choose among those already given. No formal process could have helped the process of finding a solution to the Peru–Ecuador border dispute in the mid-1990s; only the introduction of an entirely new formula, replacing competing legal claims with a common development purpose, made agreement possible at the end of the decade ([24], [25]). Moreover, models in their explicitness and parsimony hang on a single factor or parameter (public acceptance for Fearon ([8], [9]), test of control for Kydd and Walter [13]), providing a synecdochic analysis. To the extent that their aim is to highlight one effect among many, it is insightful; to the extent that it takes the part for the whole, it is restrictive.

But the burden of proof of usefulness is on the modeler, not on the practitioner. This means that the modeler must provide useful alternatives to unmodeled negotiation but must also make his results intelligible to the diplomat. The procedural challenge doubles the substantive one and makes it even more difficult. In one sense, this is the same challenge faced by conceptual analysts of negotiation, who often face diplomats who believe that only experience and a fingertips feel for the process can inform diplomacy, unaware that what analysts study and formulate into concepts is simply what practitioners do when they do it well. The practitioner may be put off by the different language and by "heathen" interference in his sacred domain, but it is up to the analyst—formal or informal modeler—to make his intrusion useful. It is true that the practitioner ignores conceptualization at his peril and that the analyst—formal or informal—can always retire to do his own thing without suffering any loss. But for the greater good of both, and for the improvement of the process and its outcome, the diplomat should make an effort to listen and learn. However, the modeler—formal or informal —must also make an effort to climb out of his prophylactic world and make his message understandable. This book has tried to further that process.

References

1. Avenhaus R (2001) Game theory and the negotiation process. PINPoints Newsletter 16:4–5, International Institute for Applied Systems Analysis, Laxenburg, Austria
2. Avenhaus R, Beetz J, Kilgour DM (2005) Quantitative models for armament escalation and negotiations. In: Zartman IW, Faure GO (eds) Escalation and negotiation. Cambridge University Press, Cambridge, UK
3. Axelrod R (1984) The Evolution of cooperation. BasicBooks, New York, USA
4. Brams S, Doherty A, Weidner M (1994) Game theory: focusing on the players, decisions and agreements. In: Zartman IW (ed) International multilateral negotiation. Jossey-Bass, San Francisco, CA, USA
5. Bueno de Mesquita B (2002) Predicting politics. The Ohio State University Press, Columbus, OH, USA
6. Edgeworth FY (1881/1967) Mathematical physics. Kegan, Paul, London, UK
7. Faure GO (ed 2003) How people negotiate. Resolving disputes in different cultures. Kluwer Academic Publishers, Dordrecht, The Netherlands
8. Fearon J (1994) Domestic political audiences and the escalation of international disputes. American Political Science Review LXXXVIII 3:577–92 (September)
9. Fearon J (1997) Signaling foreign policy interests: tying hands versus sinking costs. Journal of Conflict Resolution XXXI:68–90
10. Hicks J (1932) The theory of wages, Macmillan, New York, USA
11. Kydd A (2000) Trust, reassurance and cooperation. International Organization, LIV 3:325–57
12. Kydd A (2001) Trust building, trust breaking: the dilemma of NATO enlargement. International Organization 55:808–28
13. Kydd J, Walter B (2002) Sabotaging the peace: the politics of extremist violence. International Organization LVI 2:263–96
14. Lax D, Sebenius J (1986) The manager as negotiator, Free Press, New York, USA
15. Morrow JD (1994) Game theory for political scientists. Princeton University Press, Princeton, NJ, USA
16. Muthoo A (1999) Bargaining theory with applications. Cambridge University Press, Cambridge, UK
17. Nash J (1950) The bargaining problem. Econometrica 18:155–122. Reprinted in: Young O (ed) (1975) Bargaining. University of Illinois Press, Urbana, IL, USA

18. Nash J (1953) Two-person cooperative games. Econometrica XXI 1:28–40. Reprinted in: Young O (ed) (1975) Bargaining. University of Illinois Press, Urbana, IL, USA
19. O'Neill B (1999) Honor symbols and war. University of Michigan Press, Ann Arbor, MI, USA
20. Putnam R (1988) Diplomacy and domestic politics: the logic of two-level games. International Organization XXXXII 3:427–60
21. Raiffa H, Richardson J, Metcalfe D (2002) Negotiation analysis: the science and art of collaborative decision-making. Harvard Belknap Press, Cambridge, MA, USA
22. Reiter E, Selten R (eds) 2003, Zur Lösung des Kosovo-Konfliktes. Die Anwendung der Szenario-Bündelanalyse im Konfliktmanagement. Nomos, Baden Baden, Germany
23. Selten R (1999) Game theory and economic behavior. Vol I. 291–325, Edward Elgar, Cheltenham, UK
24. Simmons BA (1999) Territorial disputes and their resolution. United States Institute of Peace (USIP) Press, Washington DC, USA
25. Simmons BA (2005) The Peru–Ecuador border dispute. In: Zartman IW, Kremenyuk V (eds) Peace vs. justice: negotiating forward- vs. backward-looking outcomes. Rowman and Littlefield, Bolder, CO, USA
26. Spector BI, Zartman IW (eds 2004) Getting it done: post-agreement negotiation and international regimes. United States Institute for Peace (USIP) Press, Washington DC, USA
27. Zartman IW (1996) Negotiation as a search for justice. International Negotiation I 1:79–98
28. Zartman IW, Faure GO (eds) (2005) Escalation and international negotiation. Cambridge University Press, London, UK
29. Zeuthen F (1930) Problems of monopoly and economic warfare. Routledge and Sons, London, UK. Excerpts in: Young O (ed) Bargaining. University of Illinois Press, Urbana, IL, USA

Index

About the Processes of International Negotiation (PIN) Program at the International Institute for Applied Systems Analysis (IIASA)

Since 1988, the PIN Network at IIASA in Laxenburg, Austria, has been conducted by an international Steering Committee of scholars, meeting three times a year to develop and propagate new knowledge about the processes of negotiation. The Committee conducts one to two workshops every year devoted to the current collective publication project and involving scholars from a wide spectrum of countries, in order to tap a broad range of international expertise and to support scholarship on aspects of negotiation. It also offers mini-conferences on international negotiations in order to disseminate and encourage research on the subject. Such "Road Shows" have been held at the Argentine Council for International Relations, Buenos Aires; Beijing University, Beijing; the Center for Conflict Resolution, Haifa; the Center for the Study of Contemporary Japanese Culture, Kyoto; the School of International Relations, Tehran; the Netherlands Institute of International Relations, Clingendael, The Hague; the Swedish Institute of International Affairs, Stockholm; the University of Cairo; the Centre for the Study of Civil War at PRIO in Oslo; University Hassan II, Casablanca; the University of Helsinki; the Diplomatic Academy in Vienna; the University of Bayreuth; the UN University for Peace, San José; and The Johns Hopkins University in Washington, Bologna, and Nanjing. The PIN Network publishes a semiannual newsletter, *PINPoints*, and sponsors a network of over 4,000 researchers and practitioners in negotiation. The Network has been supported by the William and Flora Hewlett Foundation and the United States Institute of Peace. Contact: pin@iiasa.ac.at.

Members of the PIN Steering Committee

Rudolf Avenhaus
The German Armed Forces University Munich

Franz Cede
Austrian Ambassador to Belgium and NATO

Guy Olivier Faure
University of Paris V–Sorbonne

Victor Kremenyuk
The Russian Academy of Sciences

Paul Meerts
The Netherlands Institute of International Relations, Clingendael

Gunnar Sjöstedt
The Swedish Institute of International Affairs

I. William Zartman
The Johns Hopkins University

Selected Publications of the PIN Program

Escalation and Negotiation in International Conflicts, I.W. Zartman, G.O. Faure, editors, 2005, Cambridge University Press, Cambridge, UK.

Peace versus Justice: Negotiating Backward- and Forward-Looking Outcomes, I.W. Zartman, V. Kremenyuk, editors, 2005, Rowman & Littlefield Publishers, Inc., Lanham, MD, USA.

Negotiating European Union, P.W. Meerts, F. Cede, editors, 2004, Palgrave Macmillan, Basingstoke, UK.

Getting It Done: Post-Agreement Negotiations and International Regimes, B.I. Spector, I.W. Zartman, editors, 2003, United States Institute of Peace Press, Washington DC, USA.

How People Negotiate: Resolving Disputes in Different Cultures, G.O. Faure, editor, 2003, Kluwer Academic Publishers, Dordrecht, Netherlands.

Professional Cultures in International Negotiation: Bridge or Rift?, G. Sjöstedt, editor, 2003, Lexington Books, Lanham, MD, USA.

Containing the Atom: International Negotiations on Nuclear Security and Safety, R. Avenhaus, V.A. Kremenyuk, G. Sjöstedt, editors, 2002, Lexington Books, Lanham, MD, USA.

International Negotiation: Analysis, Approaches, Issues, 2nd Edition, V.A. Kremenyuk, editor, 2002 Jossey-Bass Inc. Publishers, San Francisco, CA, USA.

Preventive Negotiation: Avoiding Conflict Escalation, I.W. Zartman, editor, 2001, Rowman and Littlefield Publishers, Inc., Lanham, MD, USA.

Power and Negotiation, I.W. Zartman, J.Z. Rubin, editors, 2000, The University of Michigan Press, Ann Arbor, MI, USA.

International Economic Negotiation. Models versus Reality, V.A. Kremenyuk, G. Sjöstedt, editors, 2000, Edward Elgar Publishing Limited, Cheltenham, UK.

Negotiating International Regimes: Lessons Learned from the United Nations Conference on Environment and Development (UNCED), B.I. Spector, G. Sjöstedt, I.W. Zartman, editors, 1994, Graham & Trotman Limited, London, UK.

International Multilateral Negotiation: Approaches to the Management of Complexity, I.W. Zartman, editor, 1994, Jossey-Bass Inc. Publishers, San Francisco, CA, USA.

International Environmental Negotiation, G. Sjöstedt, editor, 1993, Sage Publications, Newbury Park, CA, USA.

Culture and Negotiation. The Resolution of Water Disputes, G.O. Faure, J.Z. Rubin, editors, 1993, Sage Publications, Inc., Newbury Park, CA, USA.

Processes of International Negotiations, F. Mautner-Markhof, editor, 1989, Westview Press Inc., Boulder, CO, USA.