LOGIC IN GAMES

JOHAN VAN BENTHEM

Logic in Games

Logic in Games

Johan van Benthem

The MIT Press
Cambridge, Massachusetts
London, England

©2014 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu.

This book was set in the LaTeX programming language by Fernando Velázquez Quesada and the author. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Benthem, Johan van, 1949–
Logic in games / Johan van Benthem.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-01990-3 (hardcover : alk. paper)
1. Game theory. 2. Logic. I. Title.
QA269.B36 2014
519.3—dc23

2013015059

10 9 8 7 6 5 4 3 2 1

Contents

Preface

This book is about encounters between logic and games. My interest in this interface started in my student days when I read the classic *Games and Decisions* (Luce & Raiffa 1957). It was reinforced by teaching my first course on philosophical logic in 1975, where the exciting, and exasperating, intricacies of Lorenzen dialogue games were a key theme. Later on, Hintikka's evaluation games entered my radar as a natural companion. Even so, my first systematic search was only in 1988, when I wrote a literature survey on all contacts I could find between logic and games for the Fraunhofer Foundation, commissioned for the then large amount of 2,000 German marks. I found an amazing number of interesting lines, even though there was nothing like a systematic field. Then in the 1990s, the TARK conferences on reasoning about knowledge brought real game theorists into my world who were transforming the area of epistemic logic. One person who convinced me that there was much to learn for logicians here was my Stanford colleague Yoav Shoham. As a result, my Spinoza project "Logic in Action" (1996–2001) had games as a main line, and we organized a number of meetings with game theorists and computer scientists working on the foundations of interaction. This theme has persisted at the ILLC in Amsterdam, with highlights such as the lively Marie Curie Center "Gloriclass" (2006–2010), that produced some 12 dissertations related to games in mathematics, computer science, linguistics, and the social sciences.

The origins for this book are lecture notes for the course *Logic in Games* taught in the years 1999–2002 in Amsterdam, Stanford, and elsewhere, for students coming from philosophy, mathematics, computer science, and economics. It was my way of exploring the area, with bits and pieces of established theory, and a lot of suggestions and hunches, many of them since taken up by students in papers and dissertations. Now, 10 years later, I still do not have a stable view of the subject: things keep shifting. But there is enough of substance to put up for public scrutiny.

This book has two entangled strands, connected by many bridges. First, it fits in my program of Logical Dynamics, putting information-driven agency at a center place in logic. Thus, it forms a natural sequel to the earlier books *Exploring Logical Dynamics* (van Benthem 1996) and *Logical Dynamics of Information and Interaction* (van Benthem 2011e). While this earlier work emphasized process structure and social informational events, this book adds the theme of multi-agent strategic interaction. This logical dynamics perspective is particularly clear with the first main strand in this book, the notion of a Theory of Play emerging from the combination of logic and game theory. It occupies roughly half of the book, and is prominent in Parts I and especially II, while Part III (and to some extent Part V) provide natural continuations to more global views of games.

The book also has a serious second strand that is not about the logical dynamics eschatology. The "in" of the title *Logic in Games* is meant to be ambiguous between two directions. The first is "logic *of* games," the use of logic to understand games, resulting in systems that are often called "game logics." But there is also a second direction of "logic *as* games," the use of games to understand basic notions of logic such as truth or proof. This is explained in Part IV on "logic games," of which there exists a wide variety, and it is also the spirit of various sorts of game semantics for logical systems. I find these two directions equally fundamental, and Part VI explores a number of ways in which they interact, even though the precise duality still escapes me. I believe that this interplay of the words "of" and "in" is not particular to logic and games, but that it is in fact a major feature of logical theory anywhere. Eventually, this may also throw new light on logical dynamics, as we will see in the Conclusion of this book.

Like some students and many colleagues, readers may find these perspective changes confusing. Therefore, the material has been arranged in a way that allows for several independent paths. The sequence of Parts I, II, III, and V forms an introduction to logics of games, addressing basic themes such as levels of game equivalence or strategic reasoning, with the Theory of Play as a major highlight, integrating game logics with the dynamic-epistemic logic of informational events. Part IV is a freestanding introduction to logic games, while Part V can be read as a natural continuation crossing over to general game logics. Part VI then extends this interface between our two main directions.

Read in whatever way, this book is meant to open up an area, not to close it. Its way of arranging the material brings to light quite a few open research problems, listed throughout, extending an earlier list in my survey paper *Open Problems in Logic and Games* (van Benthem 2005b).

While this book is not primarily technical, placing its main emphasis on exploring ideas, it is not a self-contained introduction to all the logics that will be linked with games. The reader is supposed to know the basics of logic, including modal logic and its computational interfaces. Many textbooks will serve this purpose. For instance, van Benthem (2010b) covers most of the basics that will be needed in what follows. Also, it will help if the reader has had prior exposure to game theory of the sort that can be achieved with many excellent available resources in that field.

What remains is the pleasant duty of mentioning some important names. As usual, I have learned a lot from supervising Ph.D. students working on dissertations in this area, in particular, Boudewijn de Bruin, Cédric Dégrémont, Amélie Gheerbrant, Nina Gierasimczuk, Lena Kurzen, Sieuwert van Otterloo, Marc Pauly, Merlijn Sevenster, and Jonathan Zvesper. I also thank my co-authors on several papers that went into the making of this book: Thomas Ågotnes, Cédric Dégremont, Hans van Ditmarsch, Amélie Gheerbrant, Sujata Ghosh, Fenrong Liu, Ştefan Minică, Sieuwert van Otterloo, Eric Pacuit, Olivier Roy, and Fernando Velázquez Quesada. And of course, many colleagues and students have been inspirational in several ways, of whom I would like to mention Krzysztof Apt, Sergei Artemov, Alexandru Baltag, Dietmar Berwanger, Giacomo Bonanno, Adam Brandenburger, Robin Clark, Jianying Cui, Paul Dekker, Nic Dimitri, Jan van Eijck, Peter van Emde Boas, Valentin Goranko, Erich Grädel, Davide Grossi, Paul Harrenstein, Jaakko Hintikka, Wilfrid Hodges, Wiebe van der Hoek, Guifei Jiang, Benedikt Löwe, Rohit Parikh, Ramaswamy Ramanujam, Robert van Rooij, Ariel Rubinstein, Tomasz Sadzik, Gabriel Sandu, Jeremy Seligman, Sonja Smets, Wolfgang Thomas, Paolo Turrini, Yde Venema, Rineke Verbrugge, Mike Wooldridge, and especially, Samson Abramsky. Of course, this is just a register of debts, not a list of endorsements. I also thank the readers who sent detailed comments on this text, the three anonymous reviewers for the MIT Press, and, especially, Giacomo Bonanno. In addition, Fernando Velázquez Quesada provided indispensable help with the physical production of this book. Finally, many of the acknowledgments that were stated in my preceding book *Logical Dynamics of Information and Interaction* (van Benthem 2011e) remain just as valid here, since there are no airtight seals between the compartments in my research. Thanks to all.

Johan van Benthem
Bloemendaal, The Netherlands
December 2012

Introduction
Exploring the Realm of Logic in Games

There are many valid points of entry to the interface zone between logic and games. This Introduction explains briefly why the interface is natural, and then takes the reader on a leisurely, somewhat rambling walk along different sites linking logic and games in a number of ways. In the course of this excursion, many general themes of this book will emerge that will be taken up more systematically later on. Readers who have no time for leisurely strolls can skip straight ahead to Chapter 1.

1 Encounters between logic and games

The appeal of games Games are a long-standing and ubiquitous practice, forming a characteristic ingredient of human culture (Huizinga 1938, Hesse 1943). Further, to the theorist of human interaction, games provide a rich model for cognitive processes, carrying vivid intuitions. The two perspectives merge naturally: a stream of ever new games offers a free cognitive laboratory where theory meets practice. Not surprisingly then, games occur in many disciplines: economics, philosophy, linguistics, computer science, cognitive science, and the social sciences. In this book, we will focus on connections between games and the field of logic. In this Introduction, we will show by means of a number of examples how this is a very natural contact. Many themes lightly touched upon here will then return in the more technical chapters of the book.

Logic of games and logic as games In what follows, we will encounter two aspects of the title of this book *Logic in Games*. Its connective 'in' is deliberately ambiguous. First, there is logic *of* games, the study of general game structure, which will lead us to contacts between logic, game theory, and also computer science and philosophy. This study employs standard techniques of the field: "game logics" capture essential aspects of reasoning about, or inside, games. But next, there is also logic *as* games, the study of logic by means of games, with "logic games" capturing basic reasoning activities and suggesting new ways of understanding what logic is. Thus, we have a cycle

$$\begin{tikzcd} \mathit{Logic} \arrow[r, bend left] & \mathit{Games} \arrow[l, bend left] \end{tikzcd}$$

Moreover, cycles invite spinning round in a spiral, or a carousel, and one can look at game logics via associated logic games, or at logic games in terms of matching game logics. Some students find this dual view confusing, preferring one direction while ignoring the other. But in this book, both sides will be present, even though we are far from understanding fully how they are intertwined. Our main focus will be on logic of games throughout, but logic as games remains an essential complementary viewpoint.

This Introduction is an informal tour of this interface. First, we introduce some simple logic games, showing how these naturally give rise to more general questions about games. This brings us to the topic of defining games in general, their analogies with processes in computer science, and their analysis by means of standard process logics. Next, we consider game theory, an area with its own motivations and concerns. We discuss a few basic themes, note their logical import, and suggest some contours of an interface between logic and game theory. What typically emerges in this mix is an analysis of players, leading to what may be called a "Theory of Play" involving many standard topics from philosophical logic. Finally, we explain what this book contains.

2 Logical games

Argumentation games While the origins of logic in antiquity are not well understood, reflection on legal, political, or philosophical debate seems a key factor in its emergence in the Greek, Indian, and Chinese traditions. Consider that debate clearly resembles a game. A person may win an argument, upholding a claim against an opponent, but there is also the bitter taste of defeat. In this process, well-timed responses are crucial. This discourse aspect has persisted in the history of logic, although descriptive aspects have also shaped its course. After the mathematical turn of the field, exact models of dialogue emerged (Lorenzen 1955). As an illustration, consider the well-known inference

$$\text{from premises } \neg A,\ A \vee B \text{ to conclusion } B$$

In the descriptive view of logic, this inference spells out what the world is like given the data at our disposal. However, we can also view it in discourse mode, as a basic subroutine in an argumentation game. There is a proponent ***P*** defending the claim B against an opponent ***O*** who has already committed to the premises $\neg A$, $A \vee B$. The procedure lets each player speak in turn. We record some possible moves:

1 ***O*** starts by challenging ***P*** to produce a defense of B.
2 ***P*** now presses ***O*** on the commitment $A \vee B$, demanding a choice.
3 ***O*** must respond to this, having nothing else to say.

There are two options here that we list separately:

$3'$ ***O*** commits to A.
$4'$ ***P*** now points at ***O***'s commitment to $\neg A$,

and ***P*** wins because of ***O***'s self-contradiction.

$3''$ ***O*** commits to B.
$4''$ Now ***P*** uses this concession to make a defense to 1.

O has nothing further to say, and loses.

One crucial feature emerges right here. Player ***P*** has a *winning strategy*: whatever ***O*** does, ***P*** can counter to win the game. This reflects an idea of logical validity as safety in debate. An inference is valid if the proponent has a winning strategy for the conclusion against an opponent granting the premises. This pragmatic view is on a par with semantic validity as the preservation of truth or syntactic validity as derivability. Valid arguments are those that can always be won in debate, as long as the moves are chosen well.

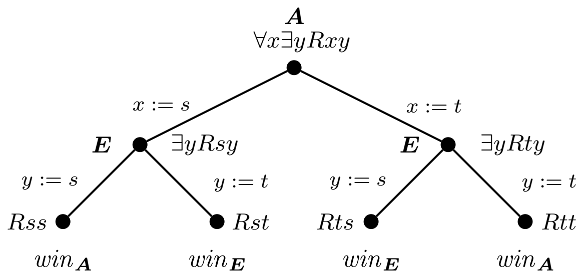
Note to the game-theoretic reader The above scenario can easily be cast as an extensive game in the standard sense of game theory. Likewise, the notion of a strategy employed here is the standard one of a function that prescribes an action for a player at each turn, without constraining what other players do at their turns. Extensive games and strategies will return throughout this book, and many precise connections between logical and game-theoretic views will unfold as we proceed.

Consistency games Argumentative dialogue is one way in which logic involves games for different actors. The other side of the coin is maintaining consistency. Player $\boldsymbol{O}$ has a positive purpose, claiming that the set $\{\neg A, A \vee B, \neg B\}$ is consistent. Indeed, maintaining consistency is an important feature of ordinary communication. Medieval logic had an "Obligatio Game" testing debating skills by requiring a student to maintain consistency while responding to challenges issued by a teacher:

> A number of rounds n is chosen, the severity of the exam. The teacher gives abstract assertions $P_1, \ldots, P_n$ that the student has to accept or reject as they are put forward. In the former case, P_i is added to the student's current commitments, and otherwise, the negation $\neg P_i$ is added. The student passes by maintaining consistency throughout.

This presentation is from Hamblin (1970). For more historical detail and accuracy, see Dutilh-Novaes (2007), Uckelman (2009). In principle, the student always has a winning strategy, by choosing some model $\boldsymbol{M}$ for the complete language of the teacher's assertions and then committing according to whether a statement is true or false in $\boldsymbol{M}$. But the realities of the game are of course much richer than this simple procedure.

Evaluation games Another famous logic game arises with understanding assertions. Consider two people discussing a quantifier statement $\forall x \exists y \varphi(x, y)$ about numbers. One player, $\boldsymbol{A}$, chooses a number x, and the other, $\boldsymbol{E}$, must come up with some number y making $\varphi(x, y)$ true. Intuitively, $\boldsymbol{A}$ challenges the initial assertion, while $\boldsymbol{E}$ defends it. To make this more concrete, consider a simple model with two objects s and t, and a relation $R = \{\langle s, t\rangle, \langle t, s\rangle\}$ (a so-called 2-cycle). An evaluation game for the assertion $\forall x \exists y Rxy$ can be pictured as a tree whose leaves are wins for the player who is right about the atomic statement reached there:



Game theorists will recognize a simple extensive form game here with four histories.

The obvious fact that $\forall x \exists y Rxy$ is true in the 2-cycle model is again reflected in a game-theoretical feature. Player $\boldsymbol{E}$ has a winning strategy in this evaluation game, which may be stated as: "choose the object different from that mentioned by $\boldsymbol{A}$." Thus, as in the preceding examples, a logical notion (this time, truth) corresponds to the existence of a winning strategy in a suitable game. This fact can be made precise by providing a general definition of evaluation games $\boldsymbol{game}(\varphi, \boldsymbol{M}, s)$ for arbitrary first-order formulas φ, models $\boldsymbol{M}$, and variable assignments s (cf. Chapter 14). Here a player called verifier $\boldsymbol{V}$ claims that the formula is true, while a falsifier $\boldsymbol{F}$ claims that it is false.

FACT The following two assertions are equivalent for all models, assignments, and formulas: (a) $\boldsymbol{M}, s \models \varphi$, (b) $\boldsymbol{V}$ has a winning strategy in $\boldsymbol{game}(\varphi, \boldsymbol{M}, s)$.

Logic games By now there are logic games for a wide variety of tasks (Hodges 2001). Much of modern logic can be usefully cast in the form of games for model checking (Hintikka 1973), argumentation and dialogue (Lorenzen 1955), comparing models for similarity (Fraïssé 1954, Ehrenfeucht 1961), or constructing models for given formulas (Hodges 2006, Hirsch & Hodkinson 2002). We will introduce the major varieties in Part IV of this book, suggesting that any significant logical task can be "gamified" once we find a natural way of pulling apart roles and purposes. Whatever the technical benefits of this shift in perspective, it is an intriguing step in reconceptualizing logic, away from lonesome thinkers and provers, to interactive tasks for several actors.

3 From logic games to general game structure

Now we take a step toward the other direction of this book. As a subspecies of games, logic games are very specialized activities. Nevertheless, they involve various broader game-theoretical issues. Nice examples can be found with the above evaluation games. We now present short previews of three fundamental issues, determinacy, game equivalence, and game operations.

Determinacy The above evaluation games have a simple but striking feature: either verifier or falsifier has a winning strategy. The reason is the logical law of the excluded middle. In any semantic model, either the given formula φ is true or its negation is. Thus, either $\boldsymbol{V}$ has a winning strategy in the game for φ, or $\boldsymbol{V}$ has a winning strategy in the game for the negation $\neg\varphi$, an operation that triggers a role switch between $\boldsymbol{V}$ and $\boldsymbol{F}$ in the game for φ. Equivalently, in the latter case, there is a winning strategy for $\boldsymbol{F}$ in the game for φ. Two-player games in which some player has a winning strategy are called *determined*.

The general game-theoretic background of our observation is a result in Zermelo (1913); see also Euwe (1929). We state this background here for two-person games whose players $\boldsymbol{A}$ and $\boldsymbol{E}$ can only win or lose, and where there is a fixed finite bound on the length of all runs.

THEOREM All zero-sum two-player games of fixed finite depth are determined.

Proof The proof is a style of solving games that will return at many places in this book. We provide a simple bottom-up algorithm determining the player having the winning strategy at any given node of a game tree of this finite sort. First, color those end nodes black that are wins for player $\boldsymbol{A}$, and color the other end nodes white, being the wins for $\boldsymbol{E}$. Then extend this coloring stepwise as follows. If all children of node n have been colored already, do one of the following:

(a) If player $\boldsymbol{A}$ is to move, and at least one child is black: color n black; if all children are white, color n white,

(b) If player $\boldsymbol{E}$ is to move, and at least one child is white: color n white; if all children are black, color n black.

This procedure colors all nodes black where player $\boldsymbol{A}$ has a winning strategy, while coloring those where $\boldsymbol{E}$ has a winning strategy white. The key to the adequacy of the coloring can be proved by induction: a player has a winning strategy at a turn iff this player can make a move to at least one daughter node where there is again a winning strategy. ■

This algorithm stands at a watershed of game theory and computer science. It points to the game solution method of *Backward Induction* that we will discuss at many places in this book. Used as a computational device, sophisticated modern versions have solved real games such as Checkers, as well as central tasks in Artificial Intelligence (Schaeffer & van den Herik 2002). Zermelo and Euwe were concerned with Chess, an old interest in computer science and cognitive science, which also allows draws. Here the theorem implies that one of the players has a non-losing strategy. Today, it is still unknown which one, as the game tree is so huge.

REMARK Infinite games
Infinite two-player games of winning and losing need not be determined. Never-ending infinite games are of independent interest, and they raise logical issues of their own that we will study in Chapters 5 and 20.

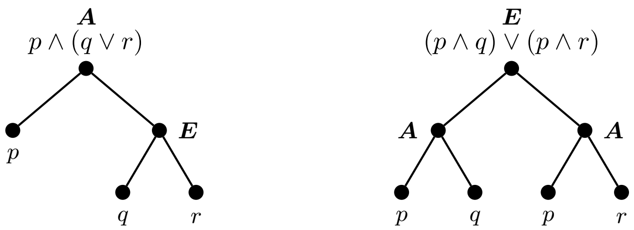
The main point of interest for this book is how close a basic game-theoretic fact can be to a standard logical law. In fact, one way of proving Zermelo's Theorem is merely by unpacking the two cases of the excluded middle for finite iterated quantified assertions "For every move of $\boldsymbol{A}$, there is a move for $\boldsymbol{E}$ (and so forth) such that $\boldsymbol{E}$ wins" (cf. Chapter 1).

Game equivalence Determinacy is important, but it is just a special property of some simple games. Logic also raises basic issues concerning arbitrary games.

EXAMPLE Propositional distribution
Consider the propositional law of distribution for conjunction over disjunction:

$$p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$$

The two finite trees in the following figure correspond to evaluation games for the two propositional formulas involved, letting $\boldsymbol{A}$ stand for falsifier and $\boldsymbol{E}$ for verifier.



This picture raises the following intuitive question: "When are two games the same?" In particular, does the logical validity of distribution mean that the pictured games are the same in some natural sense? Game equivalence is a fundamental issue, which has been studied in game theory (Thompson 1952 is a famous early source), and it will be investigated in more detail in Chapter 1. For now, intuitively, if we focus on *turns and moves*, the two games are not equivalent: they differ in etiquette (who gets to play first) and in choice structure.

This is one natural level for looking at games, involving details of the fun of playing. But if one's focus is on *achievable outcomes* only, the verdict changes.

Both players have the same powers of achieving outcomes in both games: $\boldsymbol{A}$ can force the outcome to fall in the sets $\{p\}$, $\{q, r\}$, $\boldsymbol{E}$ can force the outcome to fall in the sets $\{p, q\}$, $\{p, r\}$. Here, a player's "powers" are those sets U of outcomes for which the player has a strategy making sure the game will end inside U, no matter what all the others do. On the left, $\boldsymbol{A}$ has two strategies, *left* and *right*, yielding powers $\{p\}$ and $\{q, r\}$, and $\boldsymbol{E}$ has strategies yielding powers $\{p, q\}$, $\{p, r\}$. On the right, player $\boldsymbol{E}$ has strategies *left* and *right* giving $\boldsymbol{E}$ the same powers as on the left. By contrast, player $\boldsymbol{A}$ now has four strategies:

left: L, *right*: L, *left*: L, right: R, *left*: R, *right*: L, *left*: R, *right*: R

The first and fourth give the same powers for $\boldsymbol{A}$ as on the left, while the second and third strategy produce merely weaker powers subsumed by the former. ■

We will see later what game equivalences make sense for what purpose. For now, we note that distribution is an attractive principle about safe scheduling shifts that leave players' powers intact. Thus, once more, familiar logical laws encode significant game-theoretic content.

Game operations It is not just logical laws that have game-theoretic content. The same holds for the logical constants that make predicate logic tick. Evaluation games give a new game-theoretical take on the basic logical operations:

(a) conjunction and disjunction are choices $G \wedge H$, $G \vee H$

(b) negation is role switch, also called dual $\neg G$, or G^d

Clearly, choice and switch are completely general operations forming new games out of old. Here is another such operation that operates inside evaluation games. Consider the earlier rule for an existentially quantified formula $\exists x\psi(x)$:

$\boldsymbol{V}$ picks an object d in $\boldsymbol{M}$, and play then continues with $\psi(d)$

Perhaps surprisingly, the existential quantifier $\exists x$ does not serve as a game operation here: it clearly denotes an atomic game of object picking by verifier. The general operation in this clause hides behind the phrase "continues," which signals

(c) sequential composition of games $G\,;H$

Still, these are just a few of the natural operations that form new games out of old. Here is one more. So far we have two forms of game conjunction. The Boolean $G \wedge H$ forces a choice at the start, and the game not chosen is not played at all. Sequential composition $G\,;H$ may lead to play of both games, but only if the first ever gets completed. Now consider two basic games, "family" and "career." Neither Boolean choice nor sequential composition seems the right conjunction, and most of us try to cope with the following operation:

(d) parallel composition of games $G\,||H$

We play a stretch in one game, then switch to the other, and so on.

We will study game operations much more systematically in Part V, including connections with several systems of logic.

4 Games as interactive processes

Toward real games We have now seen how games for logical tasks have a general structure that makes sense for all games. Let us now go all the way toward real games, in economic or social behavior, sports, or war. All of these involve rule-governed action by intelligent players. We switch perspectives here, using logic as a general tool for analyzing these games. In this broader realm, logic clarifies process structure, but also the mechanics of deliberation and action by players as the game proceeds. For a start, we consider the first strand on its own, using a perspective from computer science.

Extensive games as processes Games are an enriched form of computational process, having participants with possibly different goals. Thus, games have started replacing single machines as a realistic model of distributed computation, complex computer systems, and the Internet today. An *extensive game* is a tree consisting of possible histories of moves by players taking turns indicated at nodes, while outcomes are either marked by numerical utility values for all players, or ordered by qualitative preference relations for all players. Without the preference relations, one has an extensive game form. Chapter 1 has more formal definitions of these notions, and the further chapters in Parts I and II will add more as matters are discussed in greater depth.

Trees with admissible runs such as this are very familiar from a logical point of view. They occur as "labeled transition systems" in computer science, and also as standard models for modal or temporal logics, being of the general form

$$\boldsymbol{M} = (S, \{R_a\}_{a \in A}, V)$$

where S is a universe of states or worlds, the R_a are binary transition relations on S for atomic state-changing actions a in some given set A, while the valuation V marks, for each atomic property p in some given base vocabulary, at which states p holds. There is a large logical literature on these process graphs (cf. Blackburn et al. 2001, or Chapter 1), that may be viewed either as abstract machines, or when unraveled to tree structures, as spaces of all possible executions of a process.

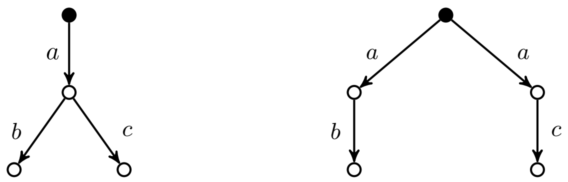
Specialized to extensive games, the states are a domain of action stages related by transitions for available moves for players, and decorated with special predicates:

$$\boldsymbol{M} = (NODES, MOVES, \boldsymbol{turn}, \boldsymbol{end}, VAL)$$

Non-final nodes are turns for players with outgoing transitions for available moves. Special proposition letters $\boldsymbol{turn}_i$ mark turns for player i, and ***end*** marks final points. The valuation *VAL* may also interpret other predicates at nodes, such as utility values for players. In this book, we will mainly use extensive game trees, although much of what we say applies to process graphs in general.

Process equivalences Now the earlier topic of structural equivalence returns. A basic concern in computer science is the level at which one wants to model processes.

Example Levels of equivalence
A well-known case of comparison in the computational literature involves the following two machines (or one-player games):

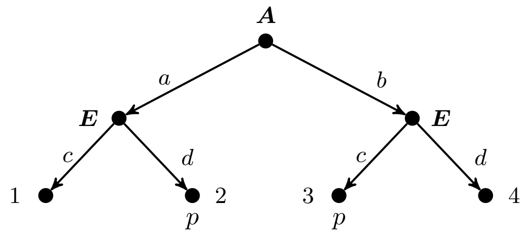


Do these diagrams represent the same process? Both produce the same finite sequences of observable actions $\{ab, ac\}$, although the first machine starts deterministically, and the other with an internal choice. In terms of external input-output behavior then, the machines are the same, given their "finite trace equivalence." But if one is also interested in internal control, in particular, available choices, then a better measure is a well-known finer structural comparison tracking choices, called "bisimulation." Indeed, the two machines have no bisimulation, as we will see in Chapter 1, which will present precise definitions for the relevant notions. ■

In the field of computation, there is a hierarchy of process equivalences, from coarser finite trace equivalence to finer ones such as bisimulation. No best level exists: it depends on the purpose. The same is true for games. Extensive games go well with bisimulation, but the earlier power level is natural, too, being an input-output view closer to trace equivalence.

Games and process logics The ladder of simulations also has a syntactic counterpart. The finer the process equivalence, the more expressive a matching *language* defining the relevant process properties. In particular, bisimulation is correlated with the use of *modal logic*, which will be one of the main working languages for games in this book.

Example Games as modal process graphs
Consider a simple two-step game between two players $\boldsymbol{A}$ and $\boldsymbol{E}$, with end nodes indicated by numbers, and one distinguished proposition letter p:



Clearly, $\boldsymbol{E}$ has a strategy making sure that a state is reached where p holds. This power is expressed by the following typical modal formula that is true at the root:

$$[a]\langle d\rangle p \wedge [b]\langle c\rangle p$$

The left-hand conjunct of this formula says that after every execution of a (marked by the universal modality $[\ ]$) there exists an execution of d (marked by the existential modality $\langle\ \rangle$) that results in a state satisfying p. The right-hand conjunct is similar. Since all actions are unique, the difference between every and some is slight here, but its intent becomes clearer in a related $[\ ]\langle\ \rangle$-type claim about the game with actions involving choice. The following response pattern will return at many places in our logical analysis of players' strategic powers:

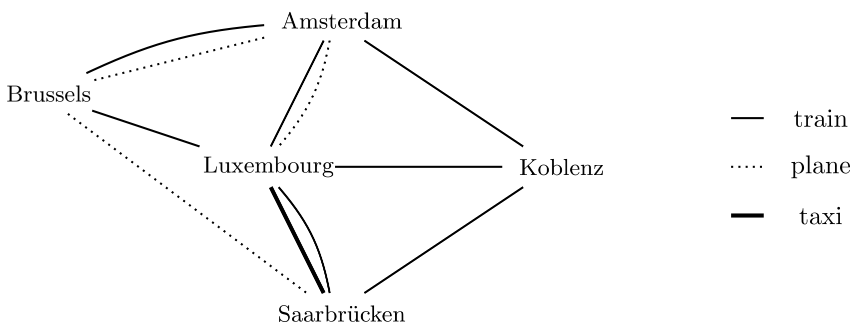
$$[a \cup b]\langle c \cup d\rangle p$$

where $a \cup b$ stands for the choice program of executing either a or b at the agent's discretion, a notion from modal logics of actions (see Chapter 1 for details). ∎

This is just a start, and the structure of games supports many other logical languages, as we will see later in this book.

From algorithms to games The link between processes and games is not just theory. Computational tasks turn into games quite easily. Consider the key search problem of graph reachability: "Given two nodes s and t in a graph, is there a sequence of successive arrows leading from s to t?" There are fast *Ptime* algorithms finding such a path if one exists (Papadimitriou 1994). But what if there is a disturbance, a reality in travel?

Example Sabotage games
The following network links two European centers of logic and computation:



Let us focus on two nodes s and t, namely, Amsterdam and Saarbrücken. It is easy to plan trips either way. But what if the usual transportation system breaks down, and a malevolent demon starts canceling connections, anywhere in the network? Let us say that, at every stage of our trip, the demon first takes out one connection. Now we have a genuine two-player game, and the question is who can win where.

From Saarbrücken to Amsterdam, a German colleague has a winning strategy. The demon's opening move may block Brussels or Koblenz, but the player gets to Luxembourg in the first round, and to Amsterdam in the next. The demon may also cut a link between Amsterdam and a city in the middle, but the player can then go to at least one place with two intact roads. But from the Dutch side, the demon has the winning strategy. It first cuts a link between Saarbrücken and Luxembourg. If the Dutch player goes to any city in the middle, the demon has time in the next rounds to cut the last link to Saarbrücken. ■

One can gamify any algorithmic task into a "sabotage game" with obstructing players. In general, the solution complexity will go up, as we will see in Chapter 23. By now, sabotage games have been used for quite different tasks as well, such as scenarios for learning.

Interactive computation as games Sabotage games exemplify a more general phenomenon, related to our earlier dichotomy between logic of games and logic as games. In addition to providing notions and tools for analyzing games, modern computer science has also started using games themselves as models for interactive computation where systems react to each other and their environment. Some recent paradigms exemplifying this perspective will be discussed in Chapters 18 and 20.

Process logics and game theory Many process calculi coexist in computer science, including modal, dynamic, and temporal logics. These will return in Parts I and II of this book, in describing fine structure of general games. As for coarser views, we will study logics of strategic powers in Part III, and matching global game operations in Part V, presenting two relevant calculi, dynamic game logic and linear game logic.

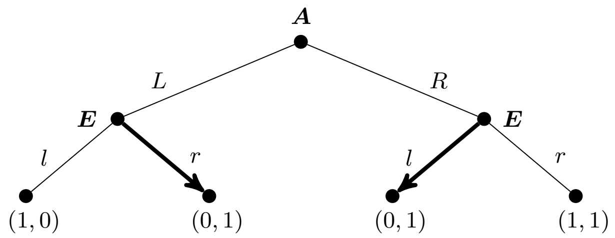
It is important to note that fine or coarse are not cultural qualifications here. This diversity of perspectives reflects two legitimate uses of logical methods in any area, namely, providing different levels of *zoom*. Sometimes, logic is used to zoom in on a topic, providing finer details of formulation and reasoning that were left implicit before. But sometimes also, logical calculi provide a higher-level abstraction, zooming out from details of a given reasoning practice to make general patterns visible. This book will provide recurrent instances of both of these zoom functions.

5 Logic meets game theory

Real games are not just about actions and information. They also crucially involve players' evaluation of outcomes, encoded in utility values or qualitative preferences. It is the balance between information, action, and evaluation that drives rational behavior. We now explore how this affects the earlier style of thinking, using logic to analyze basic assumptions about how players align their information and evaluation.

Preference, Backward Induction, and rationality How can we find, not just any, but a best course of action in a game? Assuming players to be rational, how can theorists predict behavior, or make sense of play once observed? Game theorists are after equilibria that show a stability making deviation unprofitable, although off-equilibrium behavior can sometimes be important, too: see Schelling (1978). In finite extensive games, the basic procedure of Backward Induction extends the Zermelo coloring to find such equilibria.

EXAMPLE Predicting behavior in the presence of preferences
Consider an earlier game, with players' view of outcomes displayed in ordered pairs ($\boldsymbol{A}$-value, $\boldsymbol{E}$-value):



In the earlier game of just winning and losing, $\boldsymbol{E}$ had a winning strategy marked by the black arrows, and it did not matter what $\boldsymbol{A}$ did. But now suppose that $\boldsymbol{A}$ has a slight preference between the two sites for $\boldsymbol{A}$'s defeat, being the end nodes with values $(0, 1)$. The defeat to the left takes place on an undistinguished beach, but that to the right on a picturesque mountaintop, and bards might well sings ballads about $\boldsymbol{A}$'s last stand for centuries. The new utility values for the outcomes might then be as follows, with a tiny positive number ε for the mountaintop:

$$(1,0) \qquad (0,1) \qquad (\varepsilon,1) \qquad (1,0)$$

With these preferences, $\boldsymbol{A}$ goes right at the start, and then $\boldsymbol{E}$ goes left. ∎

The algorithm computing this course of action finds values for each node in the game tree for each player, representing the best outcome value that the player can guarantee through best possible further play (as far as within the player's power).

DEFINITION Backward Induction algorithm
Here is a more precise description of the preceding numerical calculation:

> Suppose $\boldsymbol{E}$ is to move, and all values for daughter nodes are known. The $\boldsymbol{E}$-value is the maximum of all the $\boldsymbol{E}$-values on the daughters, and the $\boldsymbol{A}$-value is the minimum of the $\boldsymbol{A}$-values at all $\boldsymbol{E}$-best daughters. The dual case for $\boldsymbol{A}$'s turns is completely analogous.

Different assumptions yield modified algorithms. For instance, with a benevolent opponent, in case of ties, one might reasonably expect to get the maximal outcome from among the opponent's best moves. ∎

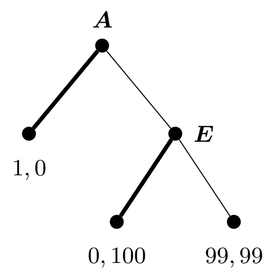
Nash equilibrium The general game-theoretic notion here is this. We state it for two players, but it works for any number (cf. Osborne & Rubinstein 1994). Any two strategies σ and τ for two players ***1*** and ***2***, respectively, determine a unique outcome $[\sigma, \tau]$ of the game, obtained by playing σ and τ against each other. This outcome can be evaluated by both players. Now we say that

> a pair of strategies σ, τ is a *Nash equilibrium* if, for no $\sigma' \neq \sigma$, $[\sigma', \tau] >_{\boldsymbol{1}} [\sigma, \tau]$, and similarly for player ***2*** with τ.

That is, neither player can improve its own outcome by changing strategies while the other sticks to the one given. Backward Induction yields strategies that are in equilibrium, even "subgame perfect equilibrium": best strategies at nodes remain best when restricted to lower nodes heading sublimes underneath.

Backward Induction is an attractive style of analysis that often makes sense. Even so, it also has instances that may give one pause.

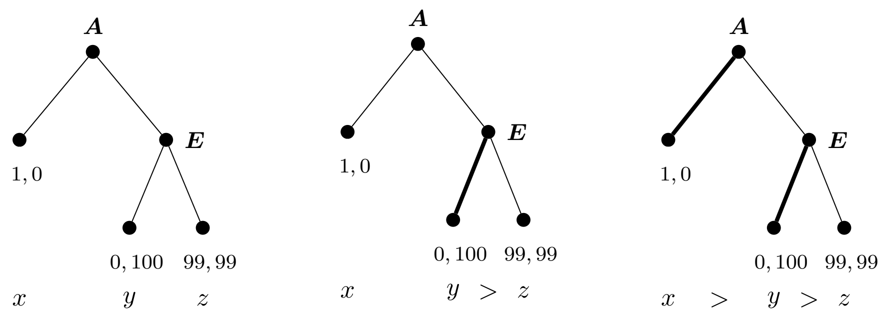
Example A debatable equilibrium
In the following game, the algorithm tells player ***E*** to turn left at the relevant turn, which then gives player ***A*** a belief that this will happen, and so, based on this belief about the other player, ***A*** should turn left at the start:



As a result, both players are worse off than in the outcome $(99, 99)$. ■

This perhaps surprising outcome raises the question of why players should act this way, and whether a more cooperative behavior could be stable. Indeed, the example is reminiscent of game-theoretic discussions of non-self-enforcing equilibria (Aumann 1990) or of so-called assurance games (Skyrms 2004). Our aim here is not to either criticize or improve game theory. The example is intended as a reminder of the subtlety of even the simplest social scenarios and the choice points that we have in understanding them. In particular, the reasoning leading to the above outcome is a mixture of many long-standing interests of logicians, including action, preference, belief, and counterfactuals. We will return to it at several places in Part I of this book, probing the logical structure of rationality. In Part II, we will go further, and analyze backward induction as a dynamic deliberation procedure that transforms a given game by successive announcements of rationality of players that gradually create a pattern of expectations. Here is an illustration in our particular case.

EXAMPLE Building up expectations in stages
Think of expectations as ordering the histories of a game by relative plausibility. In the picture below, this is marked by symbols $>$. Order appears as we keep announcing that players are "rational-in-beliefs," never playing a dominated move whose outcomes they believe to be worse than those of some other available move:



If all this is still too cryptic now, things will become fully clear in Chapter 8. ■

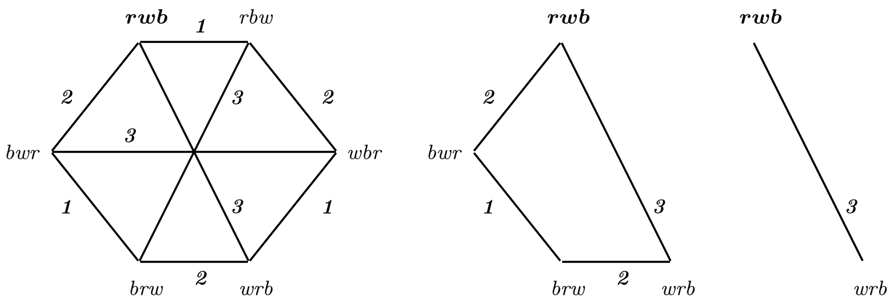
Imperfect information So far, we have looked at fully transparent games, where players' only uncertainty is about the future. But many social scenarios have further kinds of uncertainty.

EXAMPLE A card mini-game
Simple card games are an excellent setting for studying communication. Three cards red, white, and blue are given to three players: ***1*** gets red, ***2*** white, and ***3*** blue. All players see their own cards, but not the others. Now ***2*** asks ***1*** "Do you have the blue card?" and the truthful answer is forthcoming: "No." Who knows what?

If the question is genuine, player ***1*** will know the cards after it was asked. After the answer, player ***2*** knows, too, while ***3*** still does not. But there is also knowledge about others involved. At the end, all three players know that ***1*** and ***2***, but not ***3***, have learned the cards, and this fact is common knowledge between them. ■

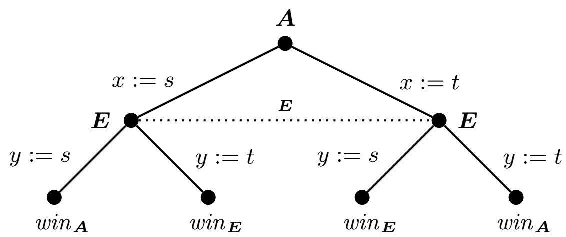
This way of understanding the scenario presupposes that questions are sincere, as seems reasonable with children. The reader will find it interesting to analyze the case with a possibly insincere question in similar terms (our later methods in this book will cover both).

Iterated knowledge about others or common knowledge in groups are crucial to games, as they help keep social behavior stable. We will study these notions in Part II of this book using epistemic logics with update mechanisms explaining the information flow in our example. As a preview, the following sequence of diagrams (it helps to play it in one's mind as a video) shows how successive information updates shrink an initial model with six possible deals of the cards. The question rules out the two worlds where ***2*** has the blue card, the answer rules out the two worlds where ***1*** has the blue card:



Let us now look at such games of "imperfect information" more abstractly, as general processes. Here is one more illustration of the sort of information structure that then emerges.

Example Evaluating formulas under imperfect information
Consider the earlier evaluation game for the formula $\forall x \exists y Rxy$, but now assuming that verifier is ignorant of the object chosen by falsifier in the opening move. The new game tree looks as follows, with a dotted line indicating $\boldsymbol{E}$'s uncertainty:



This game is quite different from the earlier one. In particular, allowing only strategies that can be played without resolving the uncertainty, as seems reasonable, player $\boldsymbol{E}$ has only two of the original four strategies left in this game: *left* and *right*. We easily see that determinacy is lost: neither player has a winning strategy! ■

Logics of imperfect information games The richer structure in these games invites a modal process language with an epistemic modality $K\varphi$ for knowledge, defined as truth of φ in all states one cannot tell apart from the current one.

Let us explore how this formalism works in the above setting. The following logical formulas describe player $\boldsymbol{E}$'s plight in the central nodes of the above game:

(a) $K_{\boldsymbol{E}}\big(\langle y := t\rangle win_{\boldsymbol{E}} \vee \langle y := s\rangle win_{\boldsymbol{E}}\big)$
Here $\boldsymbol{E}$ knows that some move will force a win for $\boldsymbol{E}$, picking either s or t.

(b) $\neg K_{\boldsymbol{E}}\langle y := t\rangle win_{\boldsymbol{E}} \;\wedge\; \neg K_{\boldsymbol{E}}\langle y := s\rangle win_{\boldsymbol{E}}$
Now, there is no particular move of which $\boldsymbol{E}$ knows that it will force a win.

This is the well-known "de re, de dicto" distinction from philosophical logic. A person may know that the ideal partner is out there (de dicto), without ever finding out who is the one (de re). The epistemic logic of imperfect information games encodes interesting properties of players, such as perfect recall or bounded memory. We will take this up in Chapter 3, and again in the study of types of players and styles of play in Part II of this book.

We have now seen at least two ways in which knowledge enters games: "forward ignorance" of the future course of play, and "sideways ignorance" about where we are in the game. We will discuss such different forms of information in much greater detail later on, especially in Chapters 5 and 6 on various models for games.

Remark Logic about logic
If you recall that the preceding example was itself a logical game, the modal logic introduced here is a logic about logic. If these sudden shifts in perspective are bothersome, please realize that dizzying self-reflective flights occur quite often in logic, and if that does not help, just take a gulp of fresh air before reading on.

Strategic form games While extensive games are appealing, game theory also has a quite different, and more widely used format, depicted in familiar matrix pictures. A *strategic game* consists of (a) a finite set N of players, (b) for each player $i \in N$ a non-empty set A_i of actions available to the player, and (c) for each player $i \in N$ a preference relation $\geq_i$ on $A = \Pi_{j\in N} A_j$.

This encodes global strategies as atomic actions, with the tuples in A standing for the total outcomes of the game that can be evaluated by the players. This level of structure is like the earlier power view, although the precise analogy is somewhat delicate (cf. Chapter 12). Strategic games are often depicted as matrices that encode basic social scenarios.

EXAMPLE Matrix games
A simple example of a strategic game is Hawk versus Dove. In many settings, agents can choose between two behaviors: aggressive or meek. Here are some matching preferences, annotating outcomes in the order ($\boldsymbol{A}$-value, $\boldsymbol{E}$-value):

		E	
		dove	*hawk*
A	*dove*	3, 3	1, 4
	hawk	4, 1	0, 0

The understanding here is different from the earlier extensive game trees, in that players now choose their actions simultaneously, independently from each other. What is optimal behavior in this scenario? The two straightforward Nash equilibria of this game are (*hawk, dove*) and (*dove, hawk*).

Another evergreen of this format is the famous Prisoner's Dilemma. Consider two countries caught in the following situation:

		E	
		arm	*disarm*
A	*arm*	1, 1	4, 0
	disarm	0, 4	3, 3

Here the only Nash equilibrium (*arm, arm*) is suboptimal in that disarming would benefit both players. The Prisoner's Dilemma raises deep issues about social cooperation (Axelrod 1984), but we have nothing to add to these in this book. ■

Not all matrix games have Nash equilibria with just the pure strategies as stated. A counterexample is the common game Matching Pennies, discussed in Chapters 3 and 21. However, one can increase the space of behaviors by adding "mixed strategies," probabilistic mixtures of pure strategies. Strategic games then bring their own theory beyond the Zermelo tradition, starting from results going back to the work of Borel, von Neumann, and Nash (cf. Osborne & Rubinstein 1994).

THEOREM All finite strategic games have Nash equilibria in mixed strategies.

Logics of strategic games Like extensive games, strategic games can be seen as models for logical languages of action, preference, and knowledge (van Benthem et al. 2011). In particular, defining Nash equilibria in strategic games has long been a benchmark problem for logics in game theory (cf. van der Hoek & Pauly 2006). In this book, strategic games will be mostly a side topic, since rational action and reasoning are often better studied at the level of detail offered by extensive games. Still, Chapters 12 and 13 will take a closer look at the modal logic of matrix games, while including some well-known conceptual issues in modeling simultaneous action.

Solving games by logical announcements Strategic games have their own solution procedures, but again these invite connections with logic, viewing matrices themselves as models for logical languages of knowledge and action. A classical method is *iterated removal of strictly dominated strategies (SD^ω)*, where one strategy dominates another for a player i if its outcomes are always better for i. The following example explores how this works.

EXAMPLE Pruning games by removing strictly dominated strategies
Consider the following matrix, with the legend (***A***-value, ***E***-value) for pairs:

			E	
		a	b	c
	d	$2, 3$	$2, 2$	$1, 1$
A	e	$0, 2$	$4, 0$	$1, 0$
	f	$0, 1$	$1, 4$	$2, 0$

First remove the dominated right-hand column: ***E***'s action c. After that, the bottom row for ***A***'s action f is strictly dominated, and after its removal, ***E***'s action b becomes strictly dominated, and then ***A***'s action e. The successive removals leave just the Nash equilibrium (d, a). ■

There is an extensive game-theoretic literature analyzing game solution in terms of players' knowledge and epistemic logic (cf. de Bruin 2010), defining optimal profiles in terms of common knowledge or belief of rationality. But there is a simple dynamic take as before, using an analogy to the well-known puzzle of the Muddy Children (Fagin et al. 1995), where iterated announcement of the children's ignorance leads to a stable solution in the limit. The above matrices may be viewed as epistemic models where players have decided on their own action, but are yet in ignorance of what others have chosen. Now, these models may change as further information comes in, say through deliberation, and as with our card game, models will then get smaller. For SD^ω, a statement of rationality that drives a matching deliberation procedure is this: "No one plays a strategy that one knows to be worse than another option."

As this gets repeated, information flows, and the game matrix shrinks in successive steps until a first fixed point is reached, much in the style of information flow that we saw already with the earlier three cards example:

$$\boxed{\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array}} \quad \boxed{\begin{array}{cc} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{array}} \quad \boxed{\begin{array}{cc} 1 & 2 \\ 4 & 5 \end{array}} \quad \boxed{\begin{array}{c} 1 \\ 4 \end{array}} \quad \boxed{\begin{array}{c} 1 \end{array}}$$

Each box acts as an epistemic model as described above: for instance, ***E***'s ranges of ignorance are the vertical columns. Each successive announcement increases players' knowledge, until the first fixed point is reached, where rationality has become common knowledge.

Thus, as with Backward Induction, there is a natural logic to solving strategic games, and we will investigate its details in Chapter 13.

Probability and mixed strategies A crucial aspect of real game theory is its use of probabilities. As we noted, all finite strategic games have equilibria in probabilistic mixed strategies. The notion of equilibrium is the same in this larger strategy space, with outcomes of profiles computed as expected values in the obvious sense. The following illustration is from the newspaper column Savant (2002), although there may also be official game-theoretical versions.

EXAMPLE The Library Puzzle

"A stranger offers you a game. You both show heads or tails. If both show heads, she pays you \$1, if both tails, she pays \$3, while you must pay her \$2 in case you show different things. Is this game fair, with expected value $1/4\times(+1)+1/4\times(+3)+1/2\times(-2) = 0$?" In her commentary, Vos Savant said the game was unfair to you with repeated play. The stranger can play heads two-thirds of the time, which gives you an average payoff of $2/3\times(1/2\times(+1)+1/2\times(-2))+1/3\times(1/2\times(+3)+1/2\times(-2) = -1/6)$. But what if you choose to play a different strategy, namely, "heads all the time"? Then the expected value is $2/3 \times (+1) + 1/3 \times (-2) = 0$. So, what is the fair value of this game?

It is easy to see that, if a strategy pair (σ, τ) is in equilibrium, each pure strategy occurring in the mixed strategy σ is also a best response for player ***1*** to τ. Then one can analyze the library game as follows. In equilibrium, let the stranger play heads with probability p and tails with $1 - p$. You play heads with probability q

and tails with probability $1-q$. Now, your expected outcome against the p-strategy must be the same whether you play heads all the time, or tails all the time. That is: $p \times 1 + (1-p) \times -2 = p \times -2 + (1-p) \times 3$, which works out to $p = 5/8$. By a similar computation, q equals 5/8 as well. The expected value for you is $-1/8$. Thus, the library game is indeed unfavorable to you, although not for exactly the reason given by the author. ■

There are several ways of interpreting what it means to play a mixed strategy (Osborne & Rubinstein 1994). For instance, besides its equilibria in pure strategies, the earlier Hawk versus Dove game has an equilibrium with each player choosing *hawk* and *dove* 50% of the time. This can be interpreted biologically in terms of stable populations having this mixture of types of individual. But it can also be interpreted in terms of degrees of belief for players produced by learning methods for patterns of behavior in evolutionary games (cf. Leyton-Brown & Shoham 2008, Hutegger & Skyrms 2012).

Logic and probability will not be a major theme in this book, although many of the logical systems that we will study have natural probabilistic extensions.

Infinite games and evolutionary game theory Here is the last topic in our tour of relevant topics in game theory. While all games so far were finite, infinite games arise naturally as well. Infinite computational processes are as basic as finite ones. Programs for standard tasks aim for termination, but equally important programs such as operating systems are meant to run forever, facilitating the running of finite tasks. Likewise, while specific conversations aim for termination, the overarching game of discourse is in principle unbounded. These are metaphors, but there is substance behind them (see Lewis 1969 and Benz et al. 2005 on the use of signaling games in understanding the conventions of natural language).

Infinite games have been used to model the emergence of cooperation (Axelrod 1984) by infinite sequences of Prisoner's Dilemma games. In such games, Backward Induction fails, but new strategies emerge exploiting the temporal structure. A key example is the Tit for Tat strategy: "copy your opponent's last choice in the next game," with immediate rewards and punishments. Tit for Tat is in Nash equilibrium with itself, making cooperation a stable option in the long run. It will find logical uses in Chapters 4 and 20.

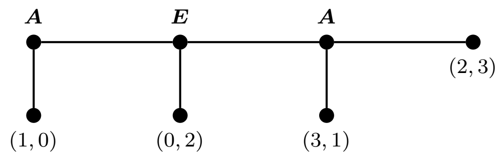
Many new notions arise in this setting, such as "evolutionary stability" for strategies against mutant invaders (Maynard-Smith 1982). Evolutionary game theory is a branch of dynamical systems theory (Hofbauer & Sigmund 1998). It has not been studied very much by logicians so far (but see Kooistra 2012 on defining evolutionary stable equilibria), and this book will have little to say about it, except in Chapters 5 and 12. In the realm of infinite evolutionary games, much of our later analysis in this book may need rethinking.

6 From logic and game theory to Theory of Play

The preceding discussion might suggest that all is well in the foundations of games, and that logic just serves to celebrate that. But things are more complex, as there are serious issues concerning the interpretation and logical structure of games.

What justifies Backward Induction? A good point of entry is Backward Induction. Its solutions have been criticized for not making coherent sense in some cases, as the following example shows.

EXAMPLE The paradox of Backward Induction
Consider the following game (longer versions are sometimes called "centipedes"):



Backward Induction computes the value $(1, 0)$ for the initial node on the left, i.e., $\boldsymbol{A}$ plays down. Now this is strange, as both players, even when competitive, can see that they would be better off going across to the end, where $\boldsymbol{A}$ is sure to get more than 1, and $\boldsymbol{E}$ gets more than 0. ■

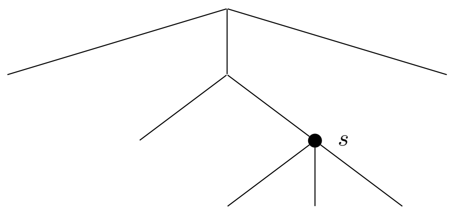
Motivating this solution leads to epistemic-logic based foundations of game theory, with insights such as the following result from Aumann (1995).

THEOREM The Backward Induction solution must obtain in extensive games whose players act while having common knowledge of rationality.

Even so, questions remain about the intuitive interpretation of the off-equilibrium path, which is crucial to the counterfactual reasoning keeping the behavior in place. Is the standard intuitive story behind Backward Induction coherent, when one really thinks it through?

Suppose that ***A*** moves right, deviating from Backward Induction, would ***E*** really stick to the old reasoning? ***E*** may think that ***A*** is now shown to be a different type of player whose behavior at the end might differ from initial expectations.

Making the players explicit The general issue here is the transition from a priori deliberation to analyzing actual real-time play of a game, where crucial information may become available about other participants. Backward Induction looks toward the future of the tree only, and hence deviations are not informative: they can be seen as mistakes without further effects for the rest of the game. This is one extreme line to take, maybe most appropriate to pregame analysis, and not so much to actual behavior during play. In the latter context, other lines of reasoning exist, where the past is informative. Given that players have come to the current node, what are the most plausible expectations about their future behavior? In an appealing picture, game trees will now mark a distinguished point s that indicates how far play has progressed, and players' accessibility relations for knowledge or belief can depend on that stage s:



Thus, players know what has happened, and let their behavior depend on two things: what the remaining game looks like, and what happened so far in the larger game. Now the space of reasoning gets much larger, and in making recommendations or predictions, we need to know more about the agents, their belief revision policies, and their styles of play (cf. Stalnaker 1999). For algorithmic alternatives to Backward Induction in this line, see the game-theoretic literature on Forward Induction: for instance, Perea (2012).

We will explore this broader program in Parts I and II of this book, expressed in the equation

$$\text{Logic} + \text{Game Theory} = \text{Theory of Play}$$

Theory of Play involves input from various areas of logic. Process structure provides the playground for games, and as we have seen, tools here come from computational logic. But play also involves agency, and hence we move to philosophical logic, and current multi-agent systems. Reasoning about social scenarios involves philosophical themes such as knowledge, belief, strategies, preferences, and goals of agents. All this is crucial to games, where players deliberate, plan, and interact over time. This calls for enriched process logics of various sorts: epistemic, counterfactual, and preference-based. We will encounter many such combinations in the course of this book. To a logician, this is gratifying, since it shows how games provide unity to a fast-expanding field.

7 Conclusion

The examples in this Introduction offer a first look at the interface of logic and games. They raise a perhaps bewildering variety of issues at first glance. Let us emphasize once more that their purpose at this stage is merely to prove that there is an interface. We hope to have shown how almost every aspect of logic has some game behind it, while vice versa, about every feature of games has some logical ramifications. The subsequent chapters of this book will present a more systematic theory behind all this.

The main duality The examples offered here came in two flavors: "logic as games," and "logic of games." The reader will have noticed that the latter line gained the upper hand as we proceeded. This book is indeed largely about the logical study of games, using notions and results from computational logic and philosophical logic. Still, as we said at the outset, this strand is not our exclusive focus. We will also devote serious attention to logical core tasks cast in the form of games in Part IV, and as our finale, the interplay of game logics and logic games will be at center stage in Part VI. This reflects our view that the duality shown here is also crucial to logic in general.

A meeting of disciplines Our examples have shown that many disciplines meet in their interest in games, often with a role for logic as an analytical tool. First of all there is game theory itself, with topics ranging from abstract mathematical analysis to empirical studies of social behavior. Entangled with this, we encountered basic themes from computer science where new flavors of game theory are emerging that incorporate ideas from process theory, automata theory, and studies of agency. Our introduction also touched on uses of games in mathematics, philosophy, and linguistics. In this book, we will develop an integral logical perspective on games, while freely using relevant ideas from all of these areas. This is not to say that our treatment contains or supersedes these other approaches. They have their own achievements and style that we cannot do full justice to. For a better perspective on what we do, and do not do, in this book, the reader can consult several excellent recent texts written with different aims, and in different styles. We will list a few at the end of this Introduction.

Do games change logic as we know it? Now let us return to what we do cover in this book. Our final point for this Introduction is about the historical thrust of the topics raised here. Logic in games, in both its senses, proposes a significant extension of classical agendas for logic, and also in some parts, a radical departure from established ways of viewing the basic logical notions. Still, it will be clear to any reader of this book that the methodology employed throughout is the standard mathematical modus operandi of the field, running from formal modeling to the design of formal systems and their meta-theoretical properties. Mathematics is neutral as to our view of logic. This is historically interesting. The famous criticism in Toulmin (1958) was that over the centuries, logic had developed an obsession with form, and accordingly, with abstract mathematical concepts that miss the essence of reasoning. In contrast, a competing alternative was offered of "formalities," the procedure that forms the heart of specialized skills such as legal argumentation, or of debate in general. The perspective on games pursued in this book shows that this is an entirely false opposition. Formalities have form. There is a surprising amount of logical structure to activities, procedures, and intelligent interaction, and it is brought out by the tools that logicians have known and loved ever since the mathematical turn of the field in the 19th century.

8 Contents of this book

The chapters of this book develop the interface of logic and games using both existing and new material. Each chapter has a clearly indicated purpose, but its nature may differ. Some are about established lines of research, others propose a new perspective and raise new questions, while still others are mainly meant to make a connection between different fields. Likewise, quite a few chapters contain original results by the author, but some merely give a didactic presentation of known techniques.

In Part I of this book, we will look at games as rich interactive processes, pursuing logics of games. Chapter 1 is about the bare process structure of extensive games, and its analysis by means of modal logics of action, systematically related to thinking in terms of bisimulations for game equivalence. Chapter 2 adds preference structure, and shows how the crucial notions of rationality operative in Backward

Induction and other game solution methods can be defined in modal logics of action and preference, or, zooming in on details of their computation, in more expressive fixed point logics. Chapter 3 adds considerations of information flow in games of imperfect information, showing how these support epistemic extensions of the preceding logics. Combining action and knowledge allows us to analyze well-behaved types of players, including those with perfect memory, in terms of special axioms. Chapter 4 then turns to the topic of strategies that players use for achieving global goals in a game, often neglected as objects of study in their own right, showing how these can be defined in propositional dynamic logic and related formalisms. Chapter 5 is a brief discussion of infinite games and temporal logics, extending the earlier modal analysis of mostly finite games. Finally, Chapter 6 is a systematic discussion of successively richer "models of games" that are needed when we want to encode more information about players.

Part II of the book then moves toward the logical dynamics of the many kinds of action that occur in and around games. Chapter 7 is a brief introduction to relevant ideas and techniques from dynamic-epistemic logic, although we will assume that the reader has access to more substantial sources beyond this book. Chapter 8 analyzes pre-play processes of deliberation, connecting Backward Induction with dynamic procedures of iterated hard or soft update with rationality, and through these, to the field of belief revision and learning scenarios. Chapter 9 then considers the dynamic logic of the many processes that may occur in actual real-time game play: observing moves, receiving other kinds of relevant information, or even changes in players' preferences. The chapter also shows how these same techniques can analyze post-play rationalizations of a game, and even what happens under actual changes in a current game. These technical results suggest the emergence of something more general, a Theory of Play whose contours and prospects are discussed in Chapter 10.

Part III of the book investigates more global perspectives on games, still with a view to the earlier concerns in Parts I and II. Chapter 11 shows how our earlier modal techniques generalize to players' powers for influencing outcomes. Chapter 12 investigates strategic form games and shows how modal logics still make good sense for analyzing information, action, and freedom. Chapter 13 then analyzes solution algorithms for strategic games in dynamic-epistemic logics of deliberation scenarios, suggesting a dynamic-epistemic foundation for game theory.

Next, Part IV makes a turn to logic games, introducing the major varieties of logic as games. Chapter 14 has evaluation games for several logical languages (first-order, modal, fixed point logics). Chapter 15 deals with comparison games between models that provide fine structure for earlier notions of structural equivalence. Chapter 16 has games for model construction in tasks of maintaining consistency, and Chapter 17 has related games for dialogue and proof, the other side of this coin. Finally, Chapter 18 draws some general lines running through and connecting different logic games, and shows how similar ideas play in the foundations of computation.

Part V mixes ideas from game logics and logic games in two major calculi of game-forming operations. Chapter 19 discusses dynamic game logic, a generalization of dynamic logic of programs to neighborhood models for players' powers. Chapter 20 has linear logics for parallel game constructions, linking up with infinite games viewed as models for interactive computation involving different systems.

Part VI of the book then draws together logic of games and logic as games, not in one grand unification, but in the form of a number of productive interfaces and merges. Chapter 21 is mainly about logical evaluation games with imperfect information, Chapter 22 discusses recent knowledge games played over epistemic models, and Chapter 23 presents sabotage games as a model of generalized computation that has features of both game logics and logic games. The final chapters raise more theoretical issues. Chapter 24 shows how logic games can serve as a complete representation for certain kinds of general game algebra, and Chapter 25 presents further general themes, including hybrid, but natural systems merging features of logic games and game logics.

The book ends with a Conclusion drawing together the main strands, identifying the major lacunae in what we have so far, and pointing at prospects for the way ahead. In particular, what the reader can learn from our material, besides the grand view as such, are many techniques of general interest. These include game equivalences, logics for internal and external analysis of game structure, algebras of game operations, and dynamic logics for information flow. Moreover, there are conceptual repercussions beyond techniques, including changes in our way of viewing games, and logic.

Further sources Our book is just one pass through the field, without any pretense at achieving completeness. As we have said earlier, the subject of games can be approached from many angles, starting in many disciplines. Here are a few sources. Hodges (2001) and van der Hoek & Pauly (2006) are compact surveys of logic and games; Osborne & Rubinstein (1994), Binmore (2008), Gintis (2000), Perea (2012), and Brandenburger et al. (2013) are lively presentations of modern game theory; the parallel volumes Leyton-Brown & Shoham (2008) and Shoham & Leyton-Brown (2008) present game theory with many links to agency in computer science; Grädel et al. (2001) presents the powerful theory of games as a model for reactive systems that is emerging in contacts with computational logic and automata theory, while Apt & Grädel (2011) adds contacts with standard game theory; Abramsky (2008) gives game semantics for programming languages and interactive computation; Mann et al. (2011) presents game-theoretic logics in philosophy and linguistics; Gintis (2008) connects game theory to the big issues in social epistemology; Pacuit & Roy (2013) present epistemic game theory for logicians and philosophers; de Bruin (2010) is a critical philosophical study of this same interface; Clark (2011) gives a broad canvas of classical and evolutionary games applied to logic, linguistics, and empirical cognitive science; Hodges (2006) and Väänänen (2011) are elegant mathematical treatises showing games at work in mathematical logic; and Kechris (1994) has sophisticated game methods in the set-theoretical foundations of mathematics.

Two major current interface conferences between logic and game theory are TARK `http://www.tark.org/` on reasoning about rationality and knowledge, and LOFT `http://www.econ.ucdavis.edu/faculty/bonanno/loft.html` on foundations of games and decision theory. Many computer science conferences on logical foundations of computation and agency also have important games-related material. A website listing relevant publications and events is `http://www.loriweb.org`.