Paper prepared for the volume Logical/Informational Dynamics, edited by
Alexandru Baltag and Sonja Smets, Trends in Logic series, Springer.

Chapter 1
Logic and Game Theory

Giacomo Bonanno and Cédric Dégremont

This version: May 7, 2013

Abstract Johan van Benthem has highlighted in his work that many questions arising in the analysis of strategic interaction call for logical and computational analysis. These questions lead to both formal and conceptually illuminating answers, in that they contribute to clarifying some of the underlying assumptions behind certain aspects of game-theoretical reasoning. We focus on the insights of a part of the literature at the interface of game theory and mathematical logic that gravitates around van Benthem's work. We discuss the formal questions raised by the perspective consisting in taking games as models for formal languages, in particular modal languages, and how eliminative reasoning processes and solution algorithms can be analyzed logically as epistemic dynamics and discuss the role played by beliefs in game-theoretical analysis and how they should be modeled from a logical point of view. We give many pointers to the literature throughout the paper.

1.1 Introduction

In the past twenty years the interface between game theory and logic has grown at a fast pace. Johan van Benthem has been taking a very active part in developing this relationship both directly, in terms of his writings, and indirectly, by seeding ideas that have been inspiring developments at the interface between the two fields. While Chapter 10 in this volume (Sandu, 2013) will be concerned with the use of game-theoretical concepts to understand fundamental logical concepts, we are delighted to contribute to this volume in honor of Johan van Benthem by focusing on some of his insights and contributions to game theory from a logical perspective.

Giacomo Bonanno
Department of Economics, University of California at Davis, USA e-mail: gfbonanno@ucdavis.edu

Cédric Dégremont
IRIT, Université Paul Sabatier, Toulouse, France e-mail: cedric.uva@gmail.com

These contributions have both a formal and conceptual aspect. They are formal in that they show how models of strategic interaction and game-theoretical concepts can be embedded in a broader formal context and analyzed from the perspective and with the tools of mathematical logic. They are also conceptual in that they contribute to clarifying some of the underlying assumptions behind certain aspects of game-theoretical reasoning. The late Michael Bacharach was one of the first to appreciate the usefulness of logic in reasoning about games:

> "Game theory is full of deep puzzles, and there is often disagreement about proposed solutions to them. The puzzlement and disagreement are neither empirical nor mathematical but, rather, concern the meanings of fundamental concepts ('solution', 'rational','complete information') and the soundness of certain arguments (that solutions must be Nash equilibria, that rational players defect in Prisoner's Dilemmas, that players should consider what would happen in eventualities which they regard as impossible). Logic appears to be an appropriate tool for game theory both because these conceptual obscurities involve notions such as reasoning, knowledge and counterfactuality which are part of the stock-in-trade of logic, and because it is a prime function of logic to establish the validity or invalidity of disputed arguments" ((Bacharach, 1994), p. 21).

For example, the tools of modal logic have made it possible to give an explicit formulation to concepts that were previously stated either informally or indirectly, such as the notion of rationalizability (Bernheim (1984); Pearce (1984)) as an expression of the notion of common belief in rationality (see Stalnaker (1994) and, for an overview of the epistemic foundations of game-theoretic solution concepts, (Battigalli and Bonanno, 1999a)). The writings of Johan van Benthem have been equally useful in pointing out new insights that modal logic can contribute to the analysis of games.

This chapter is organized as follows. The next section has some background and preliminaries that the reader might consider skipping at a first reading. In Section 1.3, we discuss several ideas put forward by van Benthem: identifying modal languages to reason about extensive games, showing how they can be interpreted and how they can characterize important classes of games and strategically stable strategies. We also show briefly how the approach bridges game theory and computational analysis. Finally we point out how van Benthem's ideas shed new light on the question of under what conditions two games can be considered the same. In Section 1.4 we present the ideas developed by van Benthem (2007b), showing how eliminative reasoning processes and solution algorithms can be analyzed logically as principles of dynamic epistemic logic and under what conditions the convergence of their iteration can be analyzed in fixed-point modal languages. Section 1.5 discusses the role played by revisable beliefs in game-theoretical analysis and how they should be modeled from a logical point of view. Building on this, and following recent results by van Benthem and Gheerbrant we discuss how backward induction can be given different interpretations and, especially, how all can be proven equivalent from the perspective of fixed-point first-order languages.

Johan van Benthem's own views on the relationship between logic and game theory are expressed in (van Benthem, 2007a).

1.2 Preliminaries

1.2.1 Notation

Let S be a set and let A be a finite set and for each $a \in A$, let $R_a \subseteq S \times S$. We let $\wp(S)$ be the power set of S and we let $\#S$ be the cardinality of S. We let R_A^* be the reflexive transitive closure of $\bigcup_{a \in A} R_a$, so that sR_A^*t if and only if either $s = t$ or there is a finite A-path from s to t. We also write R^* for R_A^* when A is clear from the context. ω is the set of natural numbers.

1.2.2 Game theory

We assume some basic familiarity with the concept of a strategic game and of an extensive game with (im)perfect information. We will define concepts formally but a reader completely unfamiliar with game theory might like to consult an introduction to game theory such as Osborne (2004).

1.2.3 Basic modal logic

Let τ be a non-empty countable set. Let PROP be a non-empty countable set (of propositional letters). The basic modal language $\mathrm{ML}(\tau, \text{PROP})$ is recursively defined as follows:

$$\varphi ::= p | \neg \varphi | \varphi \vee \varphi | \langle a \rangle \varphi$$

where p ranges over PROP and a over τ. A model for $\mathrm{ML}(\tau, \text{PROP})$ is a relational structure $\mathbb{M} = \langle W, (R_a)_{a \in \tau}, V \rangle$ where W is a non-empty set, $R_a \subseteq W \times W$ and $V : \text{PROP} \to \wp(W)$. $(W, (R_a)_{a \in \tau})$ is called a τ-frame. We also write $|\mathbb{M}| = W$.

Definition 1 (Semantics of $\mathrm{ML}(\tau, \text{PROP})$). We interpret $\mathrm{ML}(\tau, \text{PROP})$ on pointed relational models as follows:

$$
\begin{array}{ll}
\mathbb{M}, w \models p & \text{iff } w \in V(p) \\
\mathbb{M}, w \models \neg \varphi & \text{iff it is not the case that } \mathbb{M}, w \models \varphi \\
\mathbb{M}, w \models \varphi \vee \psi & \text{iff } \mathbb{M}, w \models \varphi \text{ or } \mathbb{M}, w \models \psi \\
\mathbb{M}, w \models \langle a \rangle \varphi & \text{iff there is some } v \text{ with } wR_a v \text{ and } \mathbb{M}, v \models \varphi \\
\mathbb{M}, w \models [a] \varphi & \text{iff for all } v \text{ such that } wR_a v \text{ we have } \mathbb{M}, v \models \varphi
\end{array}
$$

where $p \in \text{PROP}$, $a \in \tau$. We will write $\top$ for $p \vee \neg p$ and $\bot$ for $\neg \top$. Other connectives $(\wedge, \to, \leftrightarrow)$ are defined in the usual way.

Given a model $\mathbb{M} = \langle W, (R_a)_{a \in \tau}, V \rangle$ and a formula $\varphi \in \mathrm{ML}(\tau, \text{PROP})$ we write $||\varphi||^{\mathbb{M}} := \{w \in W \mid \mathbb{M}, w \models \varphi\}$. Whenever $\mathbb{M}$ is clear from the context, we simply write $||\varphi||$ for $||\varphi||^{\mathbb{M}}$. Given a class C of relational models, we write $C \models \varphi$ whenever for every $\mathbb{M} \in C$ and $w \in |\mathbb{M}|$ we have $\mathbb{M}, w \models \varphi$ and we say that φ is valid on C. The same notion for classes of frames is defined by universally quantifying over the possible valuations of (the relevant) propositional letters. Satisfiability is the dual, existential counterpart to validity, that is, φ is satisfiable over C iff $\neg\varphi$ is not valid over C.

Definition 2 (Bisimulation). A local bisimulation between two pointed relational structures, $(\mathbb{M}, w)$ and $(\mathbb{M}', w')$, with $\mathbb{M} = \langle W, (R_a)_{a\in\tau}, V\rangle$ and $\mathbb{M}' = \langle W', (R'_a)_{a\in\tau}, V'\rangle$ is a binary relation $Z \subseteq W \times W'$ such that wZw' and also for any pair of worlds $(x, x') \in W \times W'$ whenever xZx' then for all $a \in \tau$:

1. x, x' verify the same proposition letters.
2. if $xR_a u$ in $\mathbb{M}$ then there exists $u' \in W'$ with $x'R'_a u'$ and uZu'.
3. if $x'R'_a u'$ in $\mathbb{M}'$ then there exists $u \in W$ with $xR_a u$ and uZu'.

We say that $\mathbb{M} = \langle W, (R_a)_{a\in\tau}, V\rangle$ and $\mathbb{M}' = \langle W', (R'_a)_{a\in\tau}, V'\rangle$ are bisimilar ($\mathbb{M} \underline{\leftrightarrow} \mathbb{M}'$) if there are $w \in W$ and $w' \in W'$ such that $(\mathbb{M}, w)\underline{\leftrightarrow}(\mathbb{M}', w')$.

1.2.4 Epistemic logic

An interesting special case of a modal logic as defined above is epistemic logic. We only briefly recall the basic concepts of epistemic logic. For a more exhaustive introduction to epistemic logic, the reader can consult, e.g., Fagin et al. (1995, ch. 2). Relational structures can compactly represent the information agents have about the world and about the information possessed by the other agents.

Definition 3. An *epistemic model* is a relational structure $(W, N, (\sim_i)_{i\in N}, V)$ where N is a finite set and for each $i \in N$, $\sim_i$ is a binary equivalence relation on W.

To explicitly talk about knowledge one may use the language of basic epistemic logic.

Definition 4 (Syntax of $\mathscr{L}_{\mathrm{EL}}$). The syntax of epistemic language $\mathscr{L}_{\mathrm{EL}}$ is recursively defined as follows:

$$\varphi := \; p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi$$

where $p \in \textsc{prop}$, $i \in N$. We write $\langle i\rangle\varphi$ for $\neg K_i\neg\varphi$.

We also write $\mathscr{L}_{\mathrm{EL}}(N, \textsc{prop})$, when we need to clarify the intended set N and the intended set $\textsc{prop}$. The semantics of $\mathscr{L}_{\mathrm{EL}}$ is as expected and we only give the modal case

$$\mathbb{M}, w \models K_i\varphi \text{ iff for all } v \text{ such that } w \sim_i v \text{ we have } \mathbb{M}, v \models \varphi$$

Standard definitions such as truth sets, satisfiability and validity are of course a special case of the ones given in the previous section. Epistemic logic is fully axiomatized by the axiom system given in Table 1.1.

PL	$\vdash \varphi$ if φ is a substitution instance of a tautology of propositional logic
For $i \in N$,	
Nec	if $\vdash \varphi$, then $\vdash K_i\varphi$
K	$\vdash K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$
T	$\vdash K_i\varphi \to \varphi$
4	$\vdash K_i\varphi \to K_iK_i\varphi$
5	$\vdash \neg K_i\varphi \to K_i\neg K_i\varphi$
MP	if $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$

Table 1.1 EL (also called $S5_N$) axiom system

We write $\mathrm{K}_i[w] := \{v \in W \mid w \sim_i v\}$. For any non-empty group of agents $G \subseteq N$ we write $R^*_G[w] := \{v \in W \mid w \sim^*_G v\}$. Let φ be a formula of epistemic logic. If $R^*_G[w] \subseteq ||\varphi||$ then for any $n \in \omega$ and sequence $i_0, \ldots, i_{n-1}$ with range G, $K_{i_0} \ldots K_{i_{n-1}}\varphi$ holds at w. If the conjunction of all such finite sequences is true at w, it intuitively means that φ is common knowledge at w. But this conjunction is not finitary. We can introduce a new formula $C_G\varphi$, for each $G \subseteq N$, with semantics

$$\mathbb{M}, w \models C_G\varphi \text{ iff for all } v \text{ such that } w \sim^*_G v \text{ we have } \mathbb{M}, v \models \varphi$$

We call the resulting logic $\mathscr{L}_{MEL}$ (for multi-agent epistemic logic). $\mathscr{L}_{MEL}$ is obviously no longer compact, but the resulting logic is still invariant under basic bisimulations (van Benthem, 1997).
Axiomatization. The set of formulas of $\mathscr{L}_{MEL}$ valid over the class of all epistemic models can be axiomatized by extending EL with the axioms in Table 1.2.

$\mathbf{C_G}$**FP**	$\vdash C_G\varphi \to \bigwedge_{i\in G} K_i(\varphi \wedge C_G\varphi)$
$\mathbf{C_G}$**IR**	If $\vdash \varphi \to \bigwedge_{i\in G} K_i(\varphi \wedge \psi)$ then $\vdash \varphi \to C_G\psi$

Table 1.2 Axiom system **MEL**.

Fagin et al. (1995, ch. 2) has a completeness proof for **MEL**.

1.2.5 Model theory

We assume some basic familiarity with first-order logic (FO). For an introduction see, e.g., Ebbinghaus et al. (1994).

Given an operator: $F : \wp(U) \to \wp(U)$ we say that:

1. F is monotone, if for all $X, Y \subseteq U$ whenever $X \subseteq Y$ we have $F(X) \subseteq F(Y)$
2. F is inflationary, if $X \subseteq F(X)$ for all $X \subseteq U$
3. F is deflationary, if $F(X) \subseteq X$ for all $X \subseteq U$

Given an arbitrary operator $F : \wp(U) \to \wp(U)$, let F_{infl} (F_{def}) be the inflationary (respectively deflationary) operator associated with F, defined as follows: $F_{infl}(X) = X \cup F(X)$ (respectively $F_{def}(X) = X \cap F(X)$). Note that for an inflationary (respectively a deflationary) operator F we have $F = F_{infl}$ (respectively $F = F_{def}$). If F has a least (greatest) fixed point, we denote it by $\mathbf{lfp}(F)$ (respectively $\mathbf{gfp}(F)$). Consider the sequence defined by:

$$X^0 = \emptyset; X^\lambda = F_{infl}(\bigcup_{\eta<\lambda} X^\eta)$$

It can be shown that this sequence is inductive and stabilizes at some ordinal $\kappa \leq \#U$ (see Moschovakis (1974)). Call $\mathbf{ifp}(F) = X^\kappa$ the inflationary fixed point of F. The deflationary fixed point of F is defined analogously as the limit of the sequence

$$X^0 = U; X^\lambda = F_{def}(\bigcap_{\eta<\lambda} X^i)$$

Theorem 1 (Knaster-Tarski). *Every monotone operator F has a least fixed point $\mathbf{lfp}(F)$ and a greatest fixed point $\mathbf{gfp}(F)$. Moreover,*

$$\mathbf{lfp}(F) = \bigcap\{X \subseteq U \,|\, F(X) \subseteq X\}$$

$$\mathbf{gfp}(F) = \bigcup\{X \subseteq U \,|\, X \subseteq F(X)\}$$

Let R be an n-ary relation symbol, let $\bar{x}$ be an n-tuple of variables and let $\bar{t}$ be a n-tuple of terms. We say that an occurrence of S is positive, if it is in the scope of an even number of negations. We say that a formula $\varphi(R,\bar{x})$ is positive in R if all occurrences of R are positive.

- FO(LFP) is the extension of FO with least fixed points. Formally it extends FO with the following formation rule: if $\varphi(R,\bar{x})$ is a formula *positive in R*, then $[\mathbf{lfp}_{R,\bar{x}}\varphi(R,\bar{x})](\bar{t})$ is a formula. Where $M \models [\mathbf{lfp}_{R,\bar{x}}\varphi(R,\bar{x})](\bar{a})$ iff $\bar{a} \in \mathbf{lfp}(F_\varphi)$.
- FO(IFP) is the extension of FO with inflationary fixed points. Formally it extends FO with the following formation rule: if $\varphi(R,\bar{x})$ is a formula, then $[\mathbf{ifp}_{R,\bar{x}}\varphi(R,\bar{x})](\bar{t})$ is a formula. Where $M \models [\mathbf{ifp}_{R,\bar{x}}\varphi(R,\bar{x})](\bar{a})$ iff $\bar{a} \in \mathbf{ifp}(F_\varphi)$.

Theorem 2 (Main Theorem of Gurevich and Shelah (1986)). *For every FO(LFP) formula $\varphi(R,\bar{x})$, there is an FO(LFP) formula $\varphi^*(R,\bar{x})$ which is equivalent on all finite structures to $[\mathbf{ifp}_{R,\bar{x}}\varphi(R,\bar{x})](\bar{t})$.*

Corollary 3 (Gurevich and Shelah (1986)) *FO(LFP) = FO(IFP) over finite structures.*

Theorem 4 (Kreutzer (2004)). *For every FO(LFP) formula $\varphi(R,\bar{x})$, there is an FO(LFP) formula $\varphi^\infty(R,\bar{x})$ which is equivalent on all structures to $[\mathbf{ifp}_{R,\bar{x}}\varphi(R,\bar{x})](\bar{t})$.*

Corollary 5 (Kreutzer (2004)) *FO(LFP) = FO(IFP) over all structures.*

1.2.6 Computability and Computational complexity

We assume that the reader is familiar with the concept of a Turing machine. Our introduction will be somewhat informal and the reader is referred to (Papadimitriou, 1994) for a complete presentation of these topics. We refer to a set of (encoding of) inputs as a language. We say that a language L is *recursive* if there exists a Turing machine M that halts on an input w and accepts it whenever $w \in \mathsf{L}$, and halts and rejects the input otherwise. W say that a language L is *recursively enumerable* if there exists a Turing machine M that halts on an input w and accepts it whenever $w \in \mathsf{L}$, and either halts and rejects, or does not halt otherwise.

Besides computability, we are interested in those languages that be recognized by Turing machines using limited number of computation steps or limited amount of working-tape cells. Somewhat informally speaking—for precise definitions, see Papadimitriou (1994)— given a function $f : \omega \to \omega$, let $\mathsf{DTIME}(f)$ (respectively $\mathsf{NTIME}(f)$) be the class of languages which can be decided by a deterministic Turing machine in at most $f(n)$ steps (respectively by a non-deterministic Turing machine M such that all branches in the computation tree of M on x are bounded by $f(n)$) for any input x of size n with $n \geq n_0$ for some $n_0 \in \omega$. $\mathsf{DSPACE}(f)$ (respectively $\mathsf{NSPACE}(f)$) is the class of languages which can be recognized by a deterministic Turing machine using (respectively by a non-deterministic Turing machine M such that, on all branches in the computation tree of M on x, it uses) at most $f(n)$ cells of the working-tape, for inputs of size $n \geq n_0$ for some constant $n_0 \in \omega$. We write

- $\mathsf{PTIME} = \bigcup_{k\in\omega} \mathsf{DTIME}(n^k)$
- $\mathsf{EXPTIME} = \bigcup_{k\in\omega} \mathsf{DTIME}(2^{nk})$
- $\mathsf{NP} = \bigcup_{k\in\omega} \mathsf{NTIME}(n^k)$
- $\mathsf{PSPACE} = \bigcup_{k\in\omega} \mathsf{SPACE}(n^k)$

1.3 Games are process models

Games are process models: in two influential papers van Benthem (2001, 2002) proposes using modal languages to represent the internal structure of dynamic (or extensive-form) games. This starting point comes with important questions such as:

1. in what sense is a game a relational (epistemic, temporal) model for a modal language?
2. what classes of extensive forms can be modally characterized and in what language?
3. what is the computational complexity of the satisfiability problem for logics over epistemic temporal models?
4. what is the right notion of invariance for games?

1.3.1 *Interpreting epistemic-temporal languages over games*

Modal languages can be naturally interpreted over dynamic games. Van Benthem (2001) describes an extensive form as a tuple

$$\left\langle S, I, A, \{R_a\}_{a\in A}, \{\sim_i\}_{i\in I}\right\rangle$$

where S is a set of states (the nodes of the game tree), I is the set of players, A is the set of actions and, for every $a \in A$ and $i \in I$, R_a and $\sim_i$ are binary relations on S. If s is a decision node and $(s,t) \in R_a$ then there is a transition from node s to node t as a consequence of action a being taken by the player assigned to node s. Thus $\bigcup_{a\in A} R_a$ constitutes the game tree and the set of nodes $s \in S$ such that $(s,t) \in R_a$ for some $t \in S$ and $a \in A$ is the set of decision nodes. For every player $i \in I$, $\sim_i$ is an equivalence relation representing the state of information of the player at different stages of the game. As van Benthem notes (van Benthem 2001, p. 229), this is an extension of the traditional definition of extensive-form game where the uncertainty relation of player i is defined only on the set of decision nodes assigned to player i. This issue of specifying the information of a player also at decision nodes that belong to *other* players had earlier been studied in (Battigalli and Bonanno, 1997; Bonanno, 1992b; Quesada, 2001). Finally, adding a valuation V that associates with every propositional letter $p \in \text{PROP}$ the set of nodes at which p is true, yields a *model* of the extensive form.

Among the atomic propositions van Benthem includes sentences such as $turn_i$, which is true precisely at the decision nodes assigned to player i (where it is player i's turn to move). Given a model, one can associate with every uncertainty relation $\sim_i$ a modal operator K_i with the interpretation of $K_i\varphi$ as "player i knows φ" and with the usual semantics (see Section 1.2.4). Similarly, with every transition relation R_a one can associate a modal operator $[a]$ with the interpretation of $[a]\varphi$ as "after action a it is the case that φ" with the expected semantics (cf. Section 1.2.3).

As usual, we can then try and determine which properties of our models can be characterized, at the level of models, but also—and this is naturally where modal logic strength lays—at the level of frames. For the reader unfamiliar with modal logic, let us stress the difference: on the level of models, the modal language can surely not distinguish between a state that has at most one a-successor and a state that has many a-successors, unless these states satisfy different modal formulas. The following result explains this fact:

Theorem 6 (van Benthem 1983). *A formula of first-order logic is equivalent to the translation of a formula of modal logic iff it is invariant under bisimulations.*

However, determinacy of actions can be captured on the level of frames:

$$\langle a\rangle\,\varphi \rightarrow [a]\varphi$$

is valid on a class of frames iff these frames satisfy a-determinacy. In particular $\langle a\rangle\,\varphi \rightarrow [a]\varphi$ is valid a state w in some frame iff w has at most one a-successor in that frame.

1.3.2 Perfect recall and von Neumann extensive forms

As van Benthem points out, game-theoretical assumptions such as: (1) ‘all the nodes in the same information set have the same possible actions’ and (2) ‘a player knows when it his his turn to move’ can be characterized by the formula

$$turn_i \wedge \langle a \rangle \top \ \rightarrow \ K_i(turn_i \wedge \langle a \rangle \top).$$

Of particular interest is van Benthem’s suggestion that the property of ‘perfect recall’ (defined below), which is traditionally incorporated in the definition of extensive form, can be expressed by the formula

$$turn_i \wedge K_i[a]\varphi \ \rightarrow \ [a]K_i\varphi \qquad (vB)$$

which is very appealing, since it based on a simple commutation of the epistemic operator and the dynamic operator.

It turns out that van Benthem’s two suggestions (to extend a player’s uncertainty relation $\sim_i$ beyond player i’s decision nodes and to characterize the property of perfect recall in terms of axiom (vB)) are intimately connected and implicitly identify the subclass of extensive forms known as multi-stage or *von Neumann extensive forms*. Von Neumann extensive forms are defined (see Kuhn (1953), p. 52) by the property that any two decision nodes that belong to the same information set of a player have the same number of predecessors.

In order to make this more precise, we need a few definitions. Let S_i denote the nodes assigned to player i (player i’s decision nodes) and let $R^* = R^*_A$. The property of perfect recall is defined as follows:[1]

> For every player $i \in I$, for all nodes $t, y, y' \in S_i$ and $x \in S$ and for every action a, if tR_ax, xR^*y and $y \sim_i y'$ then there exist nodes $t' \in S_i$ and $x' \in S$ such that $t \sim_i t'$, $t'R_ax'$ and $x'R^*y'$. (PR)

Kuhn (1953) interpreted this property as “equivalent to the assertion that each player is allowed by the rules of the game to remember everything he knew at previous moves and all of his choices at those moves”. Clearly, (PR) implies the following property, which captures the notion that at any of his decision nodes player i remembers what he knew at earlier decision nodes of his:

> If $t, y \in S_i$ and tR^*y, then for every y' such that $y \sim_i y'$ there exists a $t' \in S_i$ such that $t \sim_i t'$ and $t'R^*y'$. (KM)

If, following van Benthem’s suggestion, the uncertainty relation $\sim_i$ of player i is extended from S_i to the entire set S then it is natural to require that the memory property (KM) be preserved by the extension, that is, one would require the extended relation $\sim_i$ to satisfy the following property:

[1] The following definition is Selten’s (Selten, 1975) reformulation of Kuhn’s (Kuhn, 1953) original property which was stated in terms of pure strategies.

If tR^*y and $y \sim_i y'$, then there exists a t' such that $t \sim_i t'$ and $t'R^*y'$. $\quad (KM_{EXT})$

Then we have the following result (see Bonanno (2003), inspired by van Benthem (2001)):

Proposition 1. *Fix an arbitrary extensive form G that satisfies property* (KM)*. Then*
(a) there exists, for every player i, an extension of $\sim_i$ *from* S_i *to S that satisfies* (KM_{EXT}) *if and only if the G is a von Neumann extensive form,*
(b) if G is a von Neumann extensive form then G satisfies (PR) *if and only if axiom* (vB) *is valid in G (relative to an extended relation* $\sim_i$ *that satisfies* (KM_{EXT})*).*[2]

When the extensive form is not von Neumann, then a syntactic characterization of perfect recall is still possible, but it involves a slightly more complex axiom which contains an additional operator (corresponding to the relation R^*: see Bonanno (2003)).

Another line of analysis is concerned with identifying the epistemic-temporal properties characterizing certain types of epistemic updaters. For instance, product updaters (Baltag et al. (1998), see also Chapter 7 in this volume), are typically characterized by a form of perfect recall and a form of uniformity. We refer to Hoshi (2009); van Benthem and Liu (2004); van Benthem et al. (2009) for more details on this line of research.

1.3.3 Backward induction in logic

The preceding modal languages could (indirectly) characterize classes of extensive forms of interest. Putting preferences and strategies into the picture, with corresponding modalities: with $sR_\sigma t$ meaning that t is the continuation of s given that players follow the strategy profile σ, van Benthem et al. (2005) show how 'backward induction' as property of a relation (induced by a profile of strategies) can be modally characterized by a simple PDL (see, e.g., van Eijck and Stokhof (2006)) formula:

Proposition 2 (van Benthem et al. (2005)). *In generic extensive games, the relation* σ *is induced by the unique backward induction profile iff it satisfies the following axiom for all* $i \in N$*:*

$$(turn_i \wedge \langle \sigma^* \rangle (end \wedge p)) \rightarrow [move_i] \langle \sigma^* \rangle (end \wedge \langle \leq_i \rangle p)$$

The formula is essentially saying that a player cannot *unilaterally* deviate from σ at any stage in a way that can make her strictly better off. Hoek and Pauly (2006) has more about similar definability results in the logic literature. In Section 1.5, we put this question in its broader mathematical picture: that of fixed-point logics

[2] A formula is valid in extensive form G if is true at every $s \in S$ in every model based on G.

interpreted on trees (Gheerbrant, 2010; van Benthem and Gheerbrant, 2010) and also discuss backward induction from the perspective of inductive reasoning and inductive belief update.

1.3.4 Existence of extensive games

If we reverse the perspective, and instead of asking whether a given epistemic-temporal property holds of an extensive game with imperfect information, that is if the formula is true at a certain state in a certain epistemic-temporal model, we can ask whether we can construct a strategic situation respecting a collection of constraints. The problem is known in logic, and more generally theoretical computer science, as a satisfiability problem. Let L be a modal language and let C_P be set of extensive temporal models that satisfy a property P. (Note that P could be a collection of such properties.) Formally, the set of validities of L over C_P is the set $\{\varphi \in \mathsf{L} : C_P \models \varphi\}$. The satisfiability problem for a modal language L over a class of models C_P, is to decide given any formula $\varphi \in \mathsf{L}$ whether $\{M, w | M \in C_p, w \in |M|, M, w \models \varphi\} \neq \emptyset$, in which case the answer is positive. The following are important questions at the interface of logic and computer science:

1. Is the set of validities of L over C_P recursively enumerable?
2. If it is, is it recursive?
3. Is the satisfiability problem for L over C_P in $\mathsf{EXPTIME}$? Is it in PSPACE?
4. Is it complete for these classes?

The answer to the first question would be positive, if we could identify a complete finite set of axioms. For a positive answer to the second question, on top of the previous axiomatization, we could show how to construct a model for any finite consistent set $S \subseteq \mathsf{L}$, of size bounded by some $f(|S|)$. Negative answers to these questions can be proved by reduction from acceptance problems for Turing machines or from recurring tiling problems (see Harel (1985)). Interestingly, van Benthem and Pacuit (2006) surveys how—depending on the different assumptions we are making about epistemic-temporal agents (such as, e.g., perfect recall, no learning, synchronicity...)—the satisfiability problem of epistemic-temporal languages will lay on either side of the computability border. One of the most important papers concerned with the assumptions that make the satisfiability problem of epistemic-temporal languages uncomputable is Halpern and Vardi (1989).

The third and fourth questions come, in a sense, second: when one is certain that a problem is algorithmically decidable, one can focus on exactly how many resources (time and working tape space) are required to decide it. Van Benthem and Pacuit (2006) survey also such results and has pointers to the literature. Such results can ultimately be interpreted as describing the computability and the difficulty of deciding whether a list of game-theoretical assumptions are coherent with each other and whether it is possible to find a game satisfying a list of desirable constraints. In that sense the results discussed in the previous sections suggest that modal logic interpreted over game structures are a natural way to allow for the application of computational results to game theory.

Finally note that the satisfiability problem of the basic modal language is already PSPACE-complete hence not considered tractable. By contrast, checking if a formula holds at some state in an epistemic-temporal model is tractable for the types of modal logics we have considered so far: for the most expressive of them, PDL, it can be done in a number of steps polynomial in the size of the formula and of the model. In Section 1.4 we will see that the logical analysis of strategic reasoning calls for more expressive fixed point logics, whose model-checking problem need not be tractable over arbitrary structures.

1.3.5 When are two extensive forms the same?

Besides asking the question "what are appropriate formal languages for games?", van Benthem (2001) also raises the important question "when are two extensive forms the same?". A related question is: when is a transformation of an extensive form "inessential"? These are questions that could be explored further than they have been in the literature, and two immediate approaches come to mind.

For the logician, if the language has been fixed, the question is about finding the right notion of invariance. For modal languages, some adequate notion of bisimulation is usually the answer. For first-order languages, their fixed-point extensions and existential second-order languages, a winning strategy for Duplicator in some form of Ehrenfeucht-Fraïssé game is the answer (Ehrenfeucht, 1961; Fraïssé, 1954). Hence we could have such a game to decide whether the difference between two games is essential or not. Johan van Benthem has mentioned this idea in talks (mentioning also the converse direction: interpreting languages over satisfiability games or over evaluation games, hence we could have a formula describing, indirectly, another formula).

For the game-theorist, the classical approach has been quite different: the issue being to define a different notion of game form, to show that every extensive form can be mapped into the proposed game form and to declare two extensive forms to be equivalent when they are mapped into the same "new" game form. This was done in the literature by mapping extensive forms into reduced normal forms (Dalkey, 1953; Susan and Reny, 1994; Kohlberg and Mertens, 1986; Krentel et al., 1951; Thompson, 1952) or into set-theoretic forms (Bonanno, 1992a). In both cases a corresponding set of "inessential" transformations of extensive forms were identified. As Bonanno (1992a) puts it, these mappings offer a notion of *descriptive*, rather than strategic equivalence.

Interestingly, van Benthem (2001, 2002) proposes also another notion of equivalence based on players' powers. For this purpose one needs to distinguish between terminal nodes and outcomes. Thus to the standard definition of extensive form one would add a set of outcomes W and a map f from the set of terminal nodes Z to W (if we take $Z = W$ and f to be the identity function, then we have the standard definition of extensive form). We can then say that the set $X \subseteq W$ of outcomes belongs to the *powers* of player i if player i has a strategy[3] that forces the play of the game to end at a terminal node associated with an outcome in X. For example, in Figure 1.1 we have two different extensive forms (in particular, they have different sets of terminal nodes) which share the same set of outcomes $W = \{w_1, w_2, w_3\}$ and the same powers for each player (the powers of player 1 are $\{w_1\}$ and $\{w_2, w_3\}$ and the powers of player 2 are $\{w_1, w_2\}$ and $\{w_1, w_3\}$).[4]

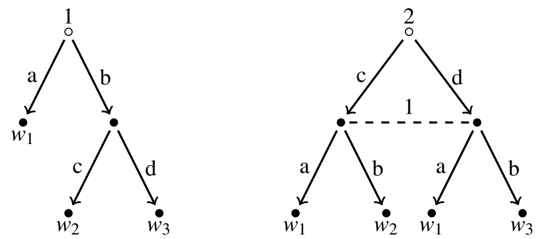

Fig. 1.1 Equivalent powers for 1 and 2 in each game.

Two extensive forms can then be defined to be "the same", or equivalent, if the powers of every player are the same in both. For example, the extensive forms of Figure 1.1 are equivalent. As van Benthem notes (van Benthem (2002), p. 13) this is a notion of equivalence based on the associated "outcome-level normal form". It should also be noted that if instead of extensive forms one considers extensive-form *games* (obtained by associating with every terminal node a payoff for each player) and one identifies outcomes with payoff vectors, then the proposed equivalence coincides with the equivalence based on the reduced normal form (Kohlberg and Mertens, 1986; Thompson, 1952). Indeed, as van Benthem notes (van Benthem (2001), p. 244, Proposition 6), the powers of the players remain the same under the Thompson (Thompson, 1952) transformations.

So far the analysis has been restricted to the powers of the players at the root of the tree. However, one can similarly define the powers of a player at any node: the set $X \subseteq W$ of outcomes belongs to the powers of player i *at node* s if player i has a strategy that forces any play of the game that goes through node s to end at a terminal node associated with an outcome in X. Van Benthem then defines a modal operator $\langle i \rangle$ for every player i with the intended interpretation of $\langle i \rangle \varphi$ as "player i has the power to bring about φ". Given a game G and a model $\mathbb{M}^G$ of G (obtained by adding a valuation that specifies which atomic propositions are true at every node) the validation rule for $\langle i \rangle \varphi$ at node s is thus: $\mathbb{M}^G, s \models \langle i \rangle \varphi$ if $\mathbb{M}^G, t \models \varphi$ for every $t \in X$ where X is one of the powers of player i at s. This approach thus uses the so-called neighborhood semantics for modal logic (Chellas, 1980), whose universal validities

[3] In the game theorist's sense of the word, namely a function that assigns to every information set of player i an action at that information set. Van Benthem calls such objects 'uniform strategies'.

[4] The dotted line in the extensive-form on the right represents an information set of Player 1.

are all the principles of the minimal modal logic except for distribution of $\langle i \rangle \varphi$ over disjunctions. One can then express interesting properties of games by means of this modal language. For example, van Benthem notes that the "consistency" property, according to which if X is a power of player i and Y is a power of player j at some node s then it must be that $X \cap Y \neq \varnothing$, can be characterized by the formula $\langle i \rangle \varphi \rightarrow \neg \langle j \rangle \neg \varphi$.

This is an interesting perspective on extensive-form games that deserves to be studied in more detail. In the logic literature, Coalition Logic (Pauly, 2002), Game Logic (Pauly and Parikh, 2003), Alternating-time temporal logic (Alur et al., 1997), STIT (for "seeing to it that", Belnap et al. (2001)— and NCL (Broersen et al., 2007) have semantics in that spirit. The reader can consult the above references for more information about them and consult Hoek and Pauly (2006) for a survey.

1.4 Reasoning in games: rational dynamics

In order to determine reasonable and/or plausible outcomes for games under given epistemic assumptions, one needs an adequate view of how players will reason from their information to reach a decision. If reasoning is traditionally the object of logic, it is so in an external way: a finite set of valid principles are proven to be everything that agents need to draw all conclusions they need to draw about their environment. If one is interested in the consequences of an agent's current information, this is the relevant level of analysis. In the context of strategic interaction this is generally not enough, since we are also interested in the semantic processes corresponding to how agents update their beliefs when they receive new information, make additional assumptions and draw consequences, and even iterate such processes. In particular we are interested in the convergence of such reasoning processes. This is the subject matter of van Benthem (2007b): a logical approach to such reasoning processes, and the theoretical limits of any such approach. In this section, we follow closely the analysis in this paper, factoring in our own, possibly different, way of looking at the topic.

Let us present the general program. As we have seen in the previous section, a game can be seen as a relational model for a modal language. More generally the epistemic aspects of a strategic situation can be described by a relational model that encodes players' preferences, players' information, and the actions they can take. A modal formula (of some sufficiently expressive language) can encode a notion of *rationality* based on the previous notions. Now assuming the rationality of the players, states in which the formula is not satisfied can be eliminated and the formula can be recursively interpreted in such submodels. To each formula corresponds a mapping on models, whose fixed-points we can hope to define in some fixed-point modal language. Moreover, for any formula and any game we might ask which (profile of) strategies survive, given some assumptions about the epistemic model of the given game. Conversely, we might ask whether there exists some epistemic model satisfying certain properties such that a certain profile of strategies (or a certain strategy) survive the inductive elimination process. All these questions, are both very natural from the point of view of mathematical logic and theoretical computer science, and the implementation of very natural questions in epistemic game theory. We illustrate this with the example of iterative solution algorithm for strategic games.

1.4.1 Epistemic models of games

A strategic game is a formal representation of a multi-agent decision situation, in which two or more agents have to make a decision, independently (rather than sequentially), 'that will influence one another's welfare' (Myerson, 1991).

Definition 5 (Strategic game). A strategic game is a structure of the form

$$\Gamma = \langle N, (A_i)_{i \in N}, (\geq_i)_{i \in N} \rangle.$$

where N is a non-empty finite set of players, for each $i \in N$, A_i is a non-empty finite set of strategies and $\geq_i$ is a total preorder over $A = \times_{j \in N} A_j$.

Note that a strategic game is not by itself a model for some epistemic logic, but it can easily be made so. Let, for example, $\text{PROP}_i = A_i$ and let $\text{PROP}_\Gamma = \bigsqcup_{i \in N} A_i$. In (van Benthem, 2007b) the result is called the full model over Γ and is defined as follows:

Definition 6 (Full epistemic model over Γ). The full epistemic model over

$$\Gamma = \langle N, (A_i)_{i \in N}, (\geq'_i)_{i \in N} \rangle$$

is the multi-agent S5(N) epistemic model

$$M(\Gamma) = \langle W, N, (\sim_i)_{i \in N}, (\geq_i)_{i \in N}, V \rangle$$

expanding Γ with

$$\begin{array}{lll} W & = & \times_{i \in N} A_i \\ ((a_i)_{i \in N}, (b_i)_{i \in N}) \in \sim_i & \text{iff} & a(j) = b(j) \\ (a_i)_{i \in N} \in V(a_j^k) & \text{iff} & a(j) = a_j^k \end{array}$$

In words, the epistemic equivalence relation for j partitions the set of strategy profiles depending on the strategy used by j in that profile. Van Benthem et al. (2011) also advocate the use of a modality for action freedom $[\approx_i]\varphi$ with the box semantics corresponding to the relation $\approx_i$ defined as follows:

$$a \approx_i b \text{ iff } a_{-i} = b_{-i}$$

—and very similar in spirit to (c)stit operators (see e.g. Belnap et al. (2001)) and NCL's $[i]$ operator (Broersen et al., 2007)— and coined by van Benthem et al. 'action freedom modality' after a concept introduced in a talk by Jeremy Seligman.

The full epistemic model has notable non-epistemic properties, such as exactly one strategy is played at each state:

$$\bigwedge_{i \in N} \Big(\bigvee_{p_i \in \text{PROP}_i} p_i \wedge \bigwedge_{p_i, p_j \in \text{PROP}_i, p_i \neq p_j} \neg (p_i \wedge p_j) \Big) \qquad (\psi_\Gamma)$$

Fact 7 *Let $M(\Gamma)$ be the full epistemic model for some strategic game Γ, we have $M(\Gamma) \models (\psi_\Gamma)$*

Such models will also satisfy strategic introspection:

$$\bigwedge_{i \in N} \bigwedge_{p_i \in \text{PROP}_i} (p_i \to K_i p_i) \qquad (\chi_\Gamma)$$

Fact 8 *Let $M(\Gamma)$ be the full epistemic model for some strategic game Γ, for every $p_i \in \text{PROP}_i$, we have $M(\Gamma) \models (\chi_\Gamma)$*

But as far as higher-order knowledge is concerned, agents have very limited information in full epistemic models. To see that, we say that a formula $\varphi \in \mathscr{L}_{\text{MEL}}(N, \text{PROP}_\Gamma)$ is Γ-consistent if it is $\mathbf{MEL}(C_G(\psi_\Gamma), C_G(\chi_\Gamma))$-consistent. The following is a variation on results in van Benthem (1997, 2007b).

Proposition 3. *Let Γ be a strategic game. Any Γ-consistent* existential *formula of the multi-agent epistemic language $\varphi \in \mathscr{L}_{\text{MEL}}(N, \text{PROP}_\Gamma)$ can be satisfied at some state in the full epistemic model $M(\Gamma)$.*

Proof (Sketch of the proof.). Existential formulas are equivalent to disjunctions of path formulas. Such a formula is satisfiable if such a path exists in $M(\Gamma)$ which can be ensured by two conditions: every state on the path should be propositionally satisfiable in $M(\Gamma)$ (this is what consistency with $C_G(\psi_\Gamma)$ ensures) and transitions should respect the epistemic grid structure of the game: *i*-transitions should preserve *i*-atoms (this is what consistency with $C_G(\chi_\Gamma)$ ensures).

Note that the preceding result comes in different flavors: if the language is richer—for example if it can express preferences or the intersection of basic relations—satisfiability will only be guaranteed for sets of formulas that are consistent with formulas of that language corresponding to the structural properties that any game model will satisfy. In other words valid properties on all game models that are invariant under the notion of bisimulation corresponding to a selected language, will need to be accounted for. But the core idea is the same: existential formulas will find a pointed full epistemic model in which they are satisfied.

However, in general, we might be interested in epistemic situations in which agents have non-trivial higher-order information. The looser the notion of an epistemic situation corresponding to a game—that is the less we would like to preserve of the epistemic structure of $M(\Gamma)$—the larger the collection of satisfiable sets of formulas. As an example, in submodels that are preserving the grid-like structure of the game, formulas that are inconsistent with the scheme

$$\langle i\rangle\langle j\rangle\varphi \leftrightarrow \langle j\rangle\langle i\rangle\varphi$$

will not be satisfiable. At the opposite end of the scale, we could work with the class of **KD45**-structures that preserve very minimal structural properties such as coherence of strategies and strategic introspection. Any Γ, **KD45**-consistent set of formulas, would then be satisfiable.

1.4.2 Assuming rationality

Now that we have fixed our models, let us go back to our initial question. How can we model the reasoning processes of agents? Or, simply, to start with, how we can model a single reasoning step? A reasoning should essentially transform an epistemic model into another epistemic model. This is a very general statement. But there is also a very rich diversity of epistemic updates that can be modeled in logics of epistemic dynamics (Baltag and Moss, 2004).

As far as the current analysis—following van Benthem (2007b)—is concerned, we will for now restrict attention to what is arguably the simplest and most natural type of update: relativization. We restrict a model to the set of states that satisfy a certain formula. It is probably in the context of epistemic analysis that relativization is easiest to interpret: it is the result of a public announcement, whose reliability is not challenged by the agents. In this case, the information is "hard information" (van Benthem et al., 2011). Softer types of information would in particular include information from only partially reliable sources, information that the agents consider as revisable (More on this issue in Section 1.5). But more than the result of a single step of eliminative reasoning, it is interesting to know what happens to game models if we recursively iterate such eliminative steps. Before we proceed, we will need a bit of notation. Given a relational model $M = \langle W, (R_a)_{a\in\tau}, V\rangle$ and a set $A \subseteq W$ let

$$M|_A = \langle A, (R'_a)_{a\in\tau}, V'\rangle$$

where $R'_a = R_a \cap (A \times A)$ and $V'(p) = V(p) \cap A$ for each $a \in \tau$ and $p \in$ PROP. We also write $M|_\varphi$ or M^φ for $M|_{||\varphi||^M}$.

Public announcement logic (Plaza, 1989; Gerbrandy, 1999; Baltag and Moss, 2004) is an extension of basic epistemic logic with public announcement operators $\langle\varphi\rangle\psi$ with semantics

$$M, w \models \langle\varphi\rangle\psi \text{ iff } M, w \models \varphi \text{ and } M|_\varphi, w \models \psi$$

Public announcement logic is actually exactly as expressive as basic epistemic logic. Now given a formula φ and a game Γ, inductively define a sequence of models $\sigma(\Gamma,\varphi) = (M_\iota)_{\iota<\gamma}$ as follows:

$$M_0 = M(\Gamma); M_\lambda = (\bigcap_{\eta<\lambda} M_\eta)^\varphi$$

We can ask two questions:

1. is ψ true at some state in M_ι for some $\iota < \gamma$?
2. is ψ true at some state in M_κ for the least ordinal κ such that $M_\kappa = (\bigcap_{\beta<\kappa} M_\beta)^\varphi$?

Here appears another direction of the program of applying logical analysis to games. In Section 1.3, modal logic was used to make explicit subphenomena of importance for strategic interaction, to identify the (model-theoretic, computational) properties of these logics that are relevant for game-theoretic analysis. Here the perspective is somewhat different, we abstract away from solution algorithms to analyze reasoning in games as a special case of reasoning about iterated relativization of relational structures in general.

The first question is concerned with iterated relativization (Miller and Moss, 2005): is it the case that at any stage in the inductive sequence (that is at any step of the reasoning process) ψ holds? (We will treat this question secondly.) The second question is concerned with the limit of iterated relativization. The first important observation to make is the following:

Proposition 4 (van Benthem (2007b)). *Let φ be a modal formula. The limit of iterated φ-relativization is definable in modal iteration calculus, that is, in inflationary fixed-point modal logic.*

Proof (Sketch of the proof.). (van Benthem, 2007b) The idea of the proof is to consider the relativization of φ to a fresh propositional variable X, $(\varphi)^X$. Now $M,w \models (\varphi)^X$ iff $M|_{V(X)},w \models \varphi$. Hence the fixed-point of the deflationary induction for $X \leftarrow (\varphi)^X$ is the limit of iterated φ-relativization.

For arbitrary modal formulas, we cannot, in general, improve on this result and find an equivalent formula in the weaker modal μ-calculus. Consider the modal formula: $\varphi(a,b) := \langle a\rangle\top \vee (r \leftrightarrow [b]\bot)$ and consider labeled transition systems, with labels in $\{a,b\}$.

Proposition 5 (Grädel and Kreutzer (2003)). $(\mathbf{dfp}X \leftarrow (\varphi(a,b))^X)$ *is not equivalent to any MSO-formula.*

Proof (Sketch of the proof.). (Grädel and Kreutzer, 2003) Define $T(n,m)$ to be a tree with two branches at the root: a branch consisting of n a-steps and a branch of m b-steps. Let r be true at the root. The idea of the proof is that the root survives in the deflationary fixed point of $X \leftarrow (\varphi(a,b))^X$ iff $n \geq m$ (see Figure 1.2). But no finite tree automaton can accept $T(n,m)$ iff $n \geq m$. On trees, this is equivalent to the fact that there is no MSO-formula corresponding to $(\mathbf{dfp}X \leftarrow (\varphi(a,b))^X)$.

The undefinability of the limit of iterated relativization then follows from the fact that the modal μ-calculus is a fragment of MSO. However, van Benthem (2007b) shows the following:

Proposition 6 (van Benthem (2007b)). *If φ is existential, the φ-relativization mapping is monotone.*

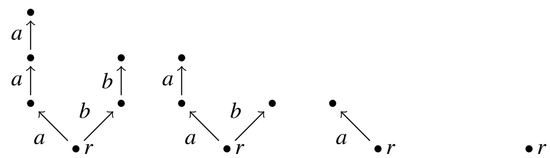

Fig. 1.2 $T_0 = T(3,2)$, $T_1 = T(3,2)|_{\varphi(a,b)}$, $T_2 = T_1|_{\varphi(a,b)}$, $T_3 = T_1|_{\varphi(a,b)} = T_3|_{\varphi(a,b)}$.

Hence, for an existential formula φ, the limit of iterated φ-relativization can be defined in the modal μ-calculus. What does this tell us about games? It tells us that reasoning about the limit of an assumption of rationality is equivalent to model checking a formula of the modal μ-calculus whenever the formula encoding this concept of rationality is existential, and that in general it is equivalent to model checking a formula of the modal iteration calculus. (We refer to Dawar et al. (2004) for a presentation of modal iteration calculus and a comparison with the modal μ-calculus.) An important difference is that, while model-checking problem for the μ-calculus could be tractable Dawar et al. (2004) show that the combined (and expressive) complexity of MIC is PSPACE-complete.

Let us now discuss the definability of iterated relativization. As suggested by van Benthem (2006) iterated relativization is expressible in the modal iteration calculus. Miller and Moss (2005) define a logic of iterated relativization extending public announcement with iterated public announcement operators $\langle\varphi^*\rangle\psi$ with semantics:

$$M, w \models \langle\varphi^*\rangle\psi \text{ iff } M, w \models \langle\varphi\rangle^n\psi \text{ for some } n \in \omega$$

and give a translation from the language of iterated relativization into the modal iteration calculus. Moreover Miller and Moss (2005) show that the satisifiability problem of the logic is highly undecidable (Σ^1_1-complete) by reduction from the tiling problem for recurring domino systems.

1.5 The different faces of Backward induction

Backward induction (henceforth BI) in generic games of perfect information seems at first a very simple solution algorithm with limpid epistemic foundations. If it is common belief between Azazello and Behemoth that they will both play best-responses to their beliefs at every subgame, then in particular Azazello believes that Behemoth will play an action that maximizes his utility in subgames of length 1, hence will play according to the BI solution. Iterating the argument seems to provide us with the conclusion that BI play follows rationality and common belief of rationality. For a formal defense of this claim, the reader should consult Aumann (1995). Even if this was the end of the story, the logical analysis of this correspondence using fixed-point logics (van Benthem and Gheerbrant, 2010; Gheerbrant,

2010) would still be insightful, and we will return to it. But there is more to this story. Consider the game below in Figure 1.3.

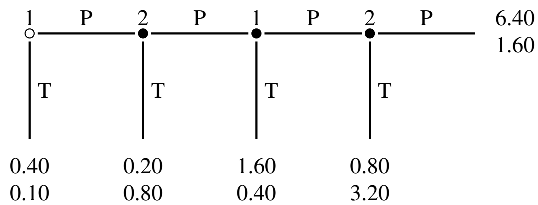

Fig. 1.3 A centipede game.

The BI solution (play T at every decision node) can be justified as follows: if 2 is rational, when the last node will be reached, 2 will play T. Since 1 expects 2 to play rationally, 1 should expect 2 to play T at the last node, hence if 1 is rational, 1 should play T at the penultimate node. The arguments iterates and 1 is argued to have a reason to play T at the first node, on the basis of common belief of rationality. Now, backward induction and any theory of rational behavior in general should be

> immune to deviation from it. That is, it must never be to one's advantage to behave in a manner that the theory deems irrational. But in order to check this, one must be able to evaluate the effect of *not* conforming to the theory. (Reny, 1993)

In particular a theory of rational behavior should have something to say about how rational agents should revise their beliefs when they observe decisions which are incompatible with the theory and how they should make decision after observing such a deviation. Now look back at the example, and assume that 1 deviates from the BI-path and plays P at the first node. What should 2 expect 1 would do in case 2 were to play P? We let the reader decide for herself or himself. Many results have shown sufficient or insufficient assumptions on the belief revision procedure used by the agents to guarantee BI compatible behavior. Our aim is not to survey them (we refer to Perea (2007) for such a discussion). We will also not cover related conceptual questions such as how beliefs should be modeled, what type of beliefs a player can have on her future and current decisions, and, what types of counterfactual reasoning can be involved in strategic reasoning; and refer to Bonanno (2013, Section 4) for an overview in the context of an analysis of the epistemic foundations of BI. Rather, our aim, in this section, is to indicate how the question has triggered logical developments calling for logical analysis of concepts we did not discuss so far: (counterfactual) beliefs—and belief revision.

With this motivation in mind and drawing on both modal logic relational semantics and semantic models developed in the context of AGM (Alchourrón et al., 1985) style belief revision theory (such as Grove (1988) spheres), Board (2004) proposes a modal language interpreted over plausibility-based structures. Independently, several authors in the logic literatures proposed similar models (van Benthem (2007c), Baltag and Smets (2006) and van Ditmarsch (2005)). All these models have to do with the following idea: besides (or instead of) an epistemic relation giving the information of an agent at every state of a model or at every history in a tree (or an extensive form), introduce a plausibility pre-order $\leq_i$: where $x \leq_i y$ means x is at least as plausible as y from the perspective of i. Belief operators can then be interpreted in different ways. Typically, $B_i\varphi$ (read "i believes that φ) can be interpreted as meaning that all $\leq_i$-minimal elements in i's information set are φ-states. As the reader might suspect, some form of well-foundedness of the plausibility relation will then be a desirable feature, for this belief operator to be well-defined. Some authors only require existence of minimal elements in cells of the information partition, which only calls for a local form of well-foundedness. Some authors prefer the $\leq_i$ to be a state-dependent relation and write $y \leq_{i,x} z$ giving them greater generality. This generally calls for additional assumptions if certain forms of positive or negative introspection are desired. But, throughout all these variations, the general idea remains the same.

We don't need to be more formal for now and we will proceed as follows. We start in Section 1.5.1 by illustrating the previous plausibility models with a foundational problem for interactive epistemology (in the sense of Aumann (1999)): agreements and convergence to agreements. We also discuss a non-eliminative revision procedure: radical upgrade. We then return in Section 1.5.3 to backward induction from where we left it earlier and discuss the unifying analysis of BI in fixed-point logics developed in van Benthem and Gheerbrant (2010); Gheerbrant (2010). Section 1.5.4 discusses a sequence of results by van Benthem and Gheerbrant making explicit the link between strategic reasoning as a non-eliminative revision procedure and the backward induction solution.

1.5.1 Plausibility models for the interactive epistemologist

Let us record the definition of an epistemic plausibility models as discussed above.

Definition 7 (Epistemic Plausiblity Model, (Baltag and Smets, 2006)). An *epistemic plausibility model* $\mathbb{M} = \langle W, (\leq_i)_{i\in N}, (\sim_i)_{i\in N}, V\rangle$ has $W \neq \emptyset$, for each $i \in N$, $\leq_i$ is a pre-order on W and $\sim_i$ is a binary equivalence relation on W, and $V : \text{PROP} \to \wp(W)$.

Since we would like to define belief as truth in minimal states in an information cell, we need to make sure that such minimal elements do exist. We call an epistemic-plausibility model $\mathbb{M} = \langle W, (\leq_i)_{i\in N}, (\sim_i)_{i\in N}, V\rangle$ well-founded iff for every subset $X \subseteq W$, X has minimal elements. Clearly this condition is sufficient to guarantee that a belief operator $B_i\varphi$ with semantics:

$$M, w \models B_i\varphi \text{ iff } \min_{\leq_i}\mathrm{K}_i[w] \subseteq ||\varphi||^M$$

is well-defined. In such models the plausibility ordering really encodes prior beliefs, while $\sim_i$ encodes the information of i, hence the above operator is really a posterior belief operator (in the sense of Aumann (1976)'s posteriors). Similarly to the probabilistic case, for any $n \in \omega$, it is possible to construct a pointed epistemic-plausibility model M,w with common prior such that for all sequences σ of length $k \leq n$ over $\{1,2\}$

$$M,w \models K_{\sigma_0} \dots K_{\sigma_k}(B_i p \wedge \neg B_j p)$$

However, again similarly to the probabilistic case, two agents cannot 'agree to disagree'.

Theorem 9 (Dégremont and Roy (2012)). *Common knowledge of disagreement is only possible in a model that is either not well-founded or for which the assumption of a common prior fails.*

Additionally it is possible to see that relativization via beliefs, that is, by public announcement of the beliefs of the agents about some formula φ, will converge to common knowledge of beliefs about φ, and hence, in well-founded models that satisfy common prior, to agreement.

φ-relativization, as we mentioned earlier, is an eliminative type of update. It corresponds to the epistemic event in which all agents accept φ as true information, whose reliability cannot be put into question. Van Benthem (2007c) is concerned with softer types of updates, corresponding to information that can possibly turn out to be wrong. One procedure discussed in van Benthem (2007c) is *radical plausibility upgrade*. It is more easily understood in the context of simple plausibility models (without epistemic relations). Radical plausibility upgrade with φ simply takes every φ-state in the plausibility ordering and puts them above all $\neg\varphi$-states. Within these two classes, the ordering of states remain unchanged. Call the resulting model $M \Uparrow \varphi$. It is possible to introduce a corresponding dynamic operator $[\Uparrow \varphi]\psi$ with semantics

$$M,w \models [\Uparrow \varphi]\psi \text{ iff } M \Uparrow \varphi, w \models \psi$$

The logic can be fully axiomatized by extending the axiomatization of some conditional doxastic logic interpreted over plausibility models with dynamic axioms. We refer to van Benthem (2007c) for details, but let us point out an important difference with the logic of public announcement (PAL), that the reader might expect: PAL validates the following axiom

$$[!p][!\neg p]B_i\bot$$

for propositional letters. This is no longer true for radical upgrade. The opposite is actually true

$$[\Uparrow p][\Uparrow \neg p]\neg B_i\bot$$

given any reasonable semantics of B_i, making radical plausibility upgrade a better candidate for iteration.

1.5.2 Belief revision over time

As we have seen before, one update, one revision is usually not giving the full story. Much of our earlier analyses has a broader impact, now that we have semantics for beliefs and an approach to belief revision. It would take us out of the scope of this paper to discuss such extensions in full details, but let us sketch some important questions that arise now that we are working with belief revision rather than knowledge update. First note that it in the same way that epistemic temporal models can be seen as generalizations (and, from a different perspective, as enrichments) of extensive forms, doxastic temporal models are similarly interestingly related to extensive games. Doxastic temporal logics interpreted over such models were introduced in Friedman and Halpern (1997) and Bonanno (2007), representing time globally as a bundle of possible histories where the beliefs of agents evolve as informational processes unfold.

As a detour, we should note that plausibility models and probability-based approaches are quite different in spirit. The plausibility structure is essentially concerned with different layers the agent can fall back to, in case her initial beliefs are defeated by new information, while probability approaches are concerned with the relative likelihood of different alternatives. The latter offer a rich basis for fine-grained decision-making rules, while the first one is robust to surprising information. There are of course systems at the interface of the two—such as lexicographic probability systems (Blume et al., 1991)—that have proven useful for the analysis of epistemic foundations of solution concepts (see, e.g., Brandenburger et al. (2008)).

Now the dynamic approach discussed in the previous section can also be extended to deal with sequences and repetition of belief revision. Van Benthem and Dégremont (2010); Dégremont (2010) discuss the relation between the temporal and the dynamic approach to belief update, and logics at their interface. Iterated scenarios as discussed in Section 1.4 can be revisited for the more sophisticated type of updates we have just discussed. Baltag and Smets (2009) has some important pioneering results and Baltag et al. (2011), as well Chapter 16 in this volume (Hendricks et al., 2013), show their relevance for a logical approach to learning theory.

1.5.3 Unifying perspectives on backward induction: fixed-point logic on trees

Our earlier analysis ended by mentioning the definability of backward induction over trees in PDL. Again think of backward induction (henceforth BI), as a subset of the successor relation on trees. Van Benthem and Gheerbrant (2010); Gheerbrant (2010) are interested in the unification of characterization of BI using extensions of FO with fixed points. The formula used in the previous result can be shown to correspond to a local concept of rationality expressible in FO with transitive closure for binary relations and the mentioned references have details. Van Benthem and Gheerbrant prefer however a different notion of rationality they call CF2 (for confluence):

$$\texttt{CF2}: \bigwedge_{i\in N} \forall x \forall y((turn_i(x) \wedge \sigma(x,y)) \to (move(x,y) \wedge \forall z(move(x,z) \\ \to \exists u \exists v(end(u) \wedge end(v) \wedge \sigma^*(y,v) \wedge \sigma^*(z,u) \wedge u \leq_i v))))$$

and show that the relation BI—or a more permissive one on arbitrary games, yet equivalent on generic games—can be defined in FO(LFP) as a greatest-fixed point. But first they prove the corresponding semantic result:

Theorem 10 (van Benthem and Gheerbrant (2010); Gheerbrant (2010)). *BI is the largest subrelation of the move relation in a finite game tree satisfying the two properties that (a) the relation has a successor at each non-terminal node, and (b) CF2 holds.*

Let X be a relational symbol *not* in the above vocabulary. Syntactically, the definability of their brand of (the relation) BI in FO(LFP) is as follows:

$$\mathsf{BI}(x,y) = [\mathbf{gfp}_{X,x,y}(move(x,y) \wedge \bigwedge_{i\in N}(turn_i(x) \to \forall z(move(x,z) \\ \to \exists u \exists v(end(u) \wedge end(v) \wedge X^*(y,v) \wedge X^*(z,u) \wedge u \leq_i v))))](x,y)$$

where $X^*(y,v)$ means that there exists an X-path from the interpretation of y to the interpretation of v and is naturally definable in FO(LFP).

Interestingly inductively computing the interpretation of BI(x,y) (van Benthem and Gheerbrant, 2010; Gheerbrant, 2010) in a given game tree is essentially equivalent to inductively computing a backward induction type solution algorithm. This illustrates that both the static of and the dynamic perspective on games can ultimately be unified in the fixed point logic approach to them.

1.5.4 Backward induction and Iterated Plausibility Upgrade

Let us go back to the example represented in Figure 1.3. If you expect players to conform to BI at any stage of the game, you expect in particular that 1 will play T at the first node. In general, the left to right ordering really corresponds to an ordering in terms of plausibility given that you expect players to play according to BI. In van Benthem and Gheerbrant (2010)'s words "Backward Induction creates expectations for players". How it creates them, is something we have yet to see. It is very reasonable to expect the BI procedure would generate such a plausibility ordering inductively.

Before we proceed, we will need a bit of notation and terminology. Assume some finite extensive game. Let $Z(x)$ be the set of terminal nodes that can still be reached from x. Let Z_1, Z_2 be sets of terminal nodes of some finite tree. Given a total ordering

$\leq$ over the terminal nodes and its complement $>$, we write $Z_1 >_{\forall\forall} Z_2$ iff for all $z_1 \in Z_1$ and $z_2 \in Z_2$ we have $z_1 > z_2$. Define $Z_1 \leq_{\forall\forall} Z_2$ similarly. Now call Z_1, Z_2 *ancestor-connected* iff is there is a node x with two-children y and z such that $Z(y) = Z_1$ and $Z(z) = Z_2$. Very roughly speaking, at x, the player who is to move decides between the set Z_1 and the set of Z_2.

Van Benthem and Gheerbrant define a relation of plausibility over leaves of a finite extensive game as a total ordering $\preceq$ of the terminal nodes. Given a set of terminal nodes Z_1, let $\mathrm{B}_i[Z_1] := \min_{\preceq_i} Z_1$. Now consider the following notion of belief-based dominance.

Definition 8 (van Benthem and Gheerbrant (2010); Gheerbrant (2010)). Given a plausibility relation $\preceq_i$, we say that a move to a node x for player i, $\preceq_i$ dominates a move to a sibling y of x in beliefs if $B_i(Z(x)) >_{\forall\forall} B_i(Z(y))$. A move to x is said to be *rational*, if it is not dominated in beliefs by a move to a sibling.

Now inductively evaluate formulas in larger subgames starting from subgames of length 1. Assume the current subgame has a root x with at least two children z and y, if z dominates y in belief z, then upgrade $Z(z)$ over $Z(y)$. Call this procedure iterated radical upgrade of rationality in belief.

Theorem 11 (Gheerbrant (2010)). *On finite games, the* BI *strategy is encoded in the final plausibility ordering at the limit of iterated radical upgrade of rationality in belief.*

A different approach, building on a similar semantic framework, is taken in Baltag et al. (2009). It uses the concept of ‘stable true belief’. Essentially, a belief in φ is a true stable belief, if it is known to be robust to truthful announcements (robust to non-trivializing relativization). Their main result is that common stable true belief of rationality implies the BI outcome. They provide a logical characterization of that result in the sense that that the previous theorem is a validity in some modal language. The reader should consult Baltag et al. (2009) for details. In general, this line of research has the particularity of having a genuinely syntactic dimension to it. The idea of recasting game-theoretic arguments in proof-theoretic terms is something that we have omitted: the interested reader should consult de Bruin (2008); Bonanno (2010); Zvesper (2010).

1.6 Perspectives

We have seen that natural questions arising within game-theoretical analysis call for logical and computational analysis. Van Benthem et al. (2011) argue that the meaning of this is certainly not only that some problems in game theory would require tools from logic, model theory and computational complexity to be solved. Rather, for van Benthem et al. (2011), intelligent interaction, as constituted of informational processes—such as revising beliefs, adjusting strategy, changing goals or preferences—is the object of an emerging "more finely-structured theory of rational agency", that they think of as a "joint off-spring [...] of logic and game theory" and call "theory of play". In that sense, the results and analyses we have discussed in the previous sections, are elements of such a general theory of intelligent interaction, in which game-theoretical, logical and computational methods are simultaneously called for.

1.7 Conclusion

Showing that many problems in strategic interaction are ultimately logical and computational problems is one of the directions of Johan van Benthem's explorations in the past decade. Not only are games natural models for modal logics, allowing for enlightening characterization of classes of games or of relational structures that generalize games. But it is also possible to model the reasoning procedures of agents about games and, the convergence of these procedures. From a logical perspective, the analysis of these problems resides within the expressive power of fixed point logics. The logical analysis projects into computational analysis: from the computational perspective, the latter problems are much more demanding. Unlike properties definable in modal logic, checking if a given fixed-point logic definable property holds about a game is generally not tractable. And if we move to the satisfiability problem, we find that we cross the computability border (if we have not crossed it already by making dangerous assumptions about our models). Hence van Bethem's contribution to this direction of the interface between logic and games has really been two-fold: on the one hand, isolating subphenomena of importance for strategic interaction (such as belief revision) and making their principles explicit by logical analysis; and, on the other hand, putting games into a broader mathematical picture, giving a unifying logical point of view at which the correspondence between static and dynamic approaches to games and their solution naturally appears as two faces of the same mathematical object.

References

Carlos E. Alchourrón, Pater Gärdenfors, and David Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, 1997.

Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

Robert J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1):6–19, 1995.

Robert J. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28(3):263–300, 1999.

Michael Bacharach. The epistemic structure of a theory of a game. *Theory and Decision*, 37:7–48, 1994.

Alexandru Baltag and Lawrence S. Moss. Logics for Epistemic Programs. *Synthese*, 139(2):165–224, 2004.

Alexandru Baltag and Sonja Smets. Dynamic belief revision over multi-agent plausibility models. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory: Proceedings of LOFT'06*, TLG, pages 11–24. AUP, 2006.

Alexandru Baltag and Sonja Smets. Group belief dynamics under iterated revision: fixed points and cycles of joint upgrades. In Aviad Heifetz, editor, *TARK*, pages 41–50, 2009.

Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *TARK '98: Proceedings of the 7th conference on Theoretical Aspects of Rationality and Knowledge*, pages 43–56, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

Alexandru Baltag, Sonja Smets, and Jonathan A. Zvesper. Keep 'hoping' for rationality: a solution to the backward induction paradox. *Synthese*, 169(2):301–333, 2009.

Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. Belief revision as a truth-tracking process. In Krzysztof R. Apt, editor, *TARK*, pages 187–190. ACM, 2011.

Pierpaolo Battigalli and Giacomo Bonanno. Synchronic information, knowledge and common knowledge in extensive games. In M. Bacharach, L.A. Gérard-Varet, P. Mongin, and H. Shin, editors, *Epistemic Logic and the Theory of Games and Decisions*, Theory and decision library, pages 235–263. Kluwer Academic, 1997. Reprinted in (Battigalli and Bonanno, 1999b).

Pierpaolo Battigalli and Giacomo Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2): 149–225, 1999a.

Pierpaolo Battigalli and Giacomo Bonanno. Synchronic information, knowledge and common knowledge in extensive games. *Research in Economics*, 53(1):77–99, 1999b.

Nuel Belnap, Michael Perloff, and Ming Xu. *Facing the future: Agents and choices in our indeterminist world*. Oxford University Press, Oxford, 2001.

Johan van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1983.

Johan van Benthem. *Exploring logical dynamics*. Center for the Study of Language and Information, Stanford, CA, USA, 1997.

Johan van Benthem. Games in dynamic-epistemic logic. *Bulletin of Economic Research*, 53(4):219–48, 2001.

Johan van Benthem. Extensive games as process models. *Journal of Logic, Language and Information*, 11(3):289–313, 2002.

Johan van Benthem. One is a lonely number. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Logic Colloquium '02*, Wellesley MA, 2006. ASL & A.K. Peters.

Johan van Benthem. Five questions on game theory. Automatic Press, 2007a.

Johan van Benthem. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9(1):377–409, 2007b. Erratum reprint, Volume 9(2), 2007, 377–409.

Johan van Benthem. Dynamic logic for belief change. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007c.

Johan van Benthem and Cédric Dégremont. Bridges between dynamic doxastic and doxastic temporal logics. In Bonanno et al. (2010), pages 151–173.

Johan van Benthem and Amélie Gheerbrant. Game solution, epistemic dynamics and fixed-point logics. *Fundamenta Informaticae*, 100(1–4):19–41, 2010.

Johan van Benthem and Fenrong Liu. Diversity of logical agents in games. *Philosophia Scientiae*, 8:163–178, 2004.

Johan van Benthem and Eric Pacuit. The tree of knowledge in action: towards a common perspective. In I. Hodkinson G. Governatori and Y. Venema, editors, *Advances in Modal Logic*, volume 6. College Publications, 2006.

Johan van Benthem, Sieuwert van Otterlo, and Olivier Roy. Preference logic, conditionals and solution concepts in games. In S. Lindström H. Lagerlund and R. Sliwinski, editors, *Modality Matters. Festschrift for Krister Segerberg*, pages 61–76. University of Uppsala, 2005. also available on the ILLC prepublication repository: PP-2005-28.

Johan van Benthem, Jelle Gerbrandy, Tomohiro Hoshi, and Eric Pacuit. Merging frameworks for interaction: DEL and ETL. *Journal of Philosophical Logic*, 2009.

Johan van Benthem, Eric Pacuit, and Olivier Roy. Toward a theory of play: A logical perspective on games and interaction. *Games*, 2(1):52–86, 2011.

B. Douglas Bernheim. Rationalizable strategic behavior. *Econometrica*, 52(4): 1007–1028, 1984.

Lawrence Blume, Adam Brandenburger, and Eddie Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1):61–79, 1991.

Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49:49–80, 2004.

Giacomo Bonanno. Set-theoretic equivalence of extensive-form games. *International Journal of Game Theory*, 20(4):429–447, 1992a.

Giacomo Bonanno. Players' information in extensive games. *Mathematical Social Sciences*, 24(1):35–48, 1992b.

Giacomo Bonanno. A syntactic characterization of perfect recall in extensive games. *Research in Economics*, 57(3):201–217, 2003.

Giacomo Bonanno. Axiomatic characterization of the AGM theory of belief revision in a temporal logic. *Artif. Intell.*, 171(2–3):144–160, 2007.

Giacomo Bonanno. A syntactic approach to rationality in games with ordinal payoffs. In Bonanno et al. (2010), pages 59–86.

Giacomo Bonanno. A dynamic epistemic characterization of backward induction without counterfactuals. *Games and Economic Behavior*, 78:31–43, 2013.

Giacomo Bonanno, Benedikt Löwe, and Wiebe van der Hoek, editors. *Logic and the Foundations of Game and Decision Theory - LOFT 8, 8th International Conference, Amsterdam, The Netherlands, July 3-5, 2008, Revised Selected Papers*, volume 6006 of *Lecture Notes in Computer Science*, 2010. Springer.

Adam Brandenburger, Amanda Friedenberg, and H. Jerome Keisler. Admissibility in games. *Econometrica*, 76(2):307–352, 2008.

Jan Broersen, Andreas Herzig, and Nicolas Troquard. Normal Coalition Logic and its conformant extension. In Dov Samet, editor, *TARK'07*, pages 91–101. PUL, 2007.

Boudewijn de Bruin. Common knowledge of rationality in extensive games. *Notre Dame Journal of Formal Logic*, 49(3):261–280, 2008.

Brian F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

Norman Dalkey. Equivalence of information patterns and essentially determinate games. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the theory of games, vol. II*, Annals of Mathematics Studies, no. 28, pages 217–243. Princeton University Press, Princeton, N. J., 1953.

Anuj Dawar, Erich Grädel, and Stephan Kreutzer. Inflationary fixed points in modal logic. *ACM Transactions on Computational Logic*, 5(2):282–315, 2004.

Cédric Dégremont. *The Temporal Mind. Observations on the logic of belief change in interactive systems*. PhD thesis, ILLC, Universiteit van Amsterdam, 2010. ILLC Dissertation Series DS-2010-03.

Cédric Dégremont and Olivier Roy. Agreement theorems in dynamic-epistemic logic. *J. Philosophical Logic*, 41(4):735–764, 2012.

Hans van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147:229–275, 2005.

Heinz-Dieter Ebbinghaus, Jörg Flum, and Wolfgang Thomas. *Mathematical logic (2. ed.)*. Undergraduate texts in mathematics. Springer, 1994. ISBN 978-3-540-94258-0.

Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1961.

Jan van Eijck and Martin Stokhof. The Gamut of Dynamic Logics. In D. Gabbay and J. Woods, editors, *Handbook of the History of Logic, Volume 6 – Logic and the Modalities in the Twentieth Century*, pages 499–600. Elsevier, 2006.

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, 1995.

Roland Fraïssé. Sur quelques classifications des systèmes de relations. *Publ. Sci. Univ. Alger. Sér. A.*, 1:35–182, 1954.

Nir Friedman and Joseph Y. Halpern. Modeling belief in dynamic systems, part I: foundations. *Artificial Intelligence*, 95(2):257–316, 1997.

Jelle Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, ILLC, Universiteit van Amsterdam, 1999. ILLC Dissertation Series DS-1999-01.

Amélie Gheerbrant. *Fixed-Point Logics on Trees*. PhD thesis, University of Amsterdam, 2010. ILLC Dissertation Series DS-2010-08.

Erich Grädel and Stephan Kreutzer. Will deflation lead to depletion? on non-monotone fixed point inductions. In *LICS*, pages 158–. IEEE Computer Society, 2003. ISBN 0-7695-1884-2.

Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

Yuri Gurevich and Saharon Shelah. Fixed-point extensions of first-order logic. *Annals of Pure and Applied Logic*, 32:265–280, 1986.

Joseph Y. Halpern and Moshe Y. Vardi. The complexity of reasoning about knowledge and time. I. Lower bounds. *Journal of Computer and Systems Science*, 38 (1):195–237, 1989.

David Harel. Recurring dominoes: making the highly undecidable highly understandable. In *Topics in the theory of computation*, pages 51–71, New York, NY, USA, 1985. Elsevier. ISBN 0-444-87647-2.

Vincent Hendricks, Dick de Jongh, and Nina Gierasimczuk. Learning theory and logic. In A. Baltag and S. Smets, editors, *Logical/Informational Dynamics*, number NBR in Trends in Logic, page PAGES. Springer, 2013.

Wiebe van der Hoek and Marc Pauly. Modal logic for games and information. In Patrick Blackburn, Johan van Benthem, and Frank Wolter, editors, *The Handbook of Modal Logic*, pages 1077–1148. Elsevier, 2006.

Tomohiro Hoshi. *Epistemic Dynamics and Protocol Information*. PhD thesis, Stanford, 2009.

Elon Kohlberg and Jean-François Mertens. On the strategic stability of equilibria. *Econometrica*, 54(5):1003–1037, 1986.

W. D. Krentel, J. C. C. McKinsey, and W. V. Quine. A simplification of games in extensive form. *Duke Mathematical Journal*, 18:885–900, 1951.

Stephan Kreutzer. Expressive equivalence of least and inflationary fixed-point logic. *Annals of Pure and Applied Logic*, 130(1–3):61–78, 2004.

Harold W. Kuhn. Extensive games and the problem of information. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the theory of games, vol. II*, Annals of Mathematics Studies, no. 28, pages 193–216. Princeton University Press, Princeton, N. J., 1953.

Joseph S. Miller and Lawrence S. Moss. The undecidability of iterated modal relativization. *Studia Logica*, 79(3):373–407, 2005.

Yiannis M. Moschovakis. *Elementary Induction on Abstract Structures*. North-Holland, New York, 1974.

Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

Martin Osborne. *An introduction to game theory*. Oxford University Press, New York, 2004.

Christos H. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.

Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

Marc Pauly and Rohit Parikh. Game logic - an overview. *Studia Logica*, 75(2): 165–182, 2003.

David G Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 52(4):1029–1050, 1984.

Andrés Perea. Epistemic foundations for backward induction: an overview. In Johan van Benthem, Dov Gabbay, and Benedikt Löwe, editors, *Interactive logic. Proceedings of the 7th Augustus de Morgan Workshop*, volume 1 of *Texts in Logic and Games*, pages 159–193. Amsterdam University Press, 2007.

Jan A. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216. Oak Ridge National Laboratory, 1989.

Antonio Quesada. On expressing maximum information in extensive games. *Mathematical Social Sciences*, 42(2):161–167, 2001.

Philip J. Reny. Common belief and the theory of games with perfect information. *Journal of Economic Theory*, 59(2):257–274, 1993.

Gabriel Sandu. Learning theory and logic. In A. Baltag and S. Smets, editors, *Game semantics in philosophical logic*, number NBR in Trends in Logic, page PAGES. Springer, 2013.

Reinhardt Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55, 1975.

Robert Stalnaker. On the evaluation of solution concepts. *Theory and Decision*, 37: 49–73, 1994.

Elmes Susan and Philip J. Reny. On the strategic equivalence of extensive form games. *Journal of Economic Theory*, 62(1):1–23, 1994.

F.B. Thompson. Equivalence of games in extensive form. RAND Research memorandum 759, The RAND Corporation, 1952.

Jonathan A. Zvesper. *Playing with information*. PhD thesis, ILLC, Universiteit van Amsterdam, 2010. ILLC Dissertation Series DS-2010-02.