BRANCHING TIME LOGIC, PERFECT INFORMATION GAMES AND BACKWARD INDUCTION

Giacomo Bonanno
Department of Economics,
University of California,
Davis, CA 95616-8578 USA
e-mail: gfbonanno@ucdavis.edu
www.econ.ucdavis.edu/~bonanno

First draft: November 1998.
Comments welcome

Keywords: tense logic, branching time,
prediction, backward induction, internal consistency.

Abstract

The logical foundations of game-theoretic solution concepts have so far been explored within the confines of epistemic logic. In this paper we turn to a different branch of modal logic, namely temporal logic, and propose to view the solution of a game as a complete prediction about future play. We extend the branching time framework by adding agents and by defining the notion of prediction. We show that perfect information games are a special case of extended branching time frames and that the backward-induction solution is a prediction. We also provide a characterization of backward induction in terms of the property of internal consistency of prediction.

1. Introduction

The logical foundations of game theory have been the object of a recent and growing literature.[1] Most papers in this area make use (directly or indirectly) of epistemic modal logic, that is, the logic of knowledge and belief, and try to determine what assumptions on the beliefs and reasoning of the players are implicit in various solution concepts. The task of this research program is to identify for any game the strategies that might be chosen by rational and intelligent players who know the structure of the game and the preferences of their opponents and who recognize each other's rationality and reasoning abilities.

In this paper we turn to a different branch of modal logic, namely temporal logic[2], and propose to view the solution of a game as a *prediction* about future play.

The focus of this paper is on extensive games with perfect information[3], which are modeled in a natural way within the framework of branching time logic. In the next section we extend the semantics of branching time by adding agents and by defining the notion of prediction. A prediction can be thought of as a belief about the future[4] and Sections 2 and 3 are devoted to the analysis of what logical properties one should attribute to predictions in general.[5] In Section 4 we show that extensive games with perfect information are a special case of branching time frames and that the backward-induction solution of such games can indeed be viewed as a prediction (that is, it satisfies the logic of prediction developed in Section 3). In Section 5 we provide a syntactic characterization of backward induction in terms of *internal consistency* of prediction, in the following sense. If at some predicted future time Player i's payoff is q then, no matter what action Player i takes, it will always be the case that *if* Player i's payoff is, or is predicted to be, r then r is not greater than q.[6]

[1] Extensive surveys of this literature are given in Battigalli and Bonanno (1998) and Dekel and Gul (1997). These two papers provide a fairly comprehensive list of references.

[2] See, for example, van Benthem (1991), Burgess (1984), Goldblatt (1992) and Øhrstrøm and Hasle (1995).

[3] See, for example, Fudenberg and Tirole (1991) or Osborne and Rubinstein (1994).

[4] When we make a non-trivial prediction about the future we select, among the conceivable future descriptions of the world, those that appear to us to be most likely.

[5] The analysis presented in Sections 3 and 4 extends that of Bonanno (1998) where the logic of prediction was first studied.

[6] It we think of the prediction as a "recommendation" to the players, then internal consistency says that if the recommendation is that (the game be played in such a way that) Player i get a payoff of q then it is not possible for Player i to take an action after which her payoff is greater

This notion of internal consistency (or stability) of a solution is not new: it was first introduced within cooperative game theory by von Neumann and Morgenstern (1947) and subsequently applied by Joseph Greenberg (1990) in his all-encompassing theory of social situation. The novelty of this paper lies in the interpretation of a solution as a prediction within the framework of branching-time logic and in the proof that the implicit logic behind the backward induction solution is that of an internally consistent prediction. As far as we know this is also the first time that the tools of temporal logic have been used to analyze game theoretic concepts.[7]

2. Agents in branching time

Definition 2.1. *A branching-time frame with agents (BTA frame for short) is a tuple* $\left\langle T, \prec, N, A, \{R_{ia}\}_{(i,a)\in N\times A}\right\rangle$ *where*

- T is a set of moments or points in time or states
- $\prec$ is a binary relation on T (representing the ordering of time) satisfying the following properties:

(P.0) antisymmetry: if $t_1 \prec t_2$ then $t_2 \nprec t_1$.
(P.1) transitivity: if $t_1 \prec t_2$ and $t_2 \prec t_3$ then $t_1 \prec t_3$.
(P.2) backward linearity: if $t_1 \prec t_3$ and $t_2 \prec t_3$ then either $t_1 = t_2$ or $t_1 \prec t_2$ or $t_2 \prec t_1$.

- $N = \{1, ..., n\}$ is a *finite* set of agents
- A is a *finite* set of actions

than q or the recommendation is that (the game be played in such a way that) Player i get a payoff greater than q.

[7] The logic of agency in branching time has been studied extensively in the philosophical literature: see, for example, Belnap and Perloff (1988), Chellas (1992), Horty and Belnap (1995), Horty (1996) and references therein. These papers, however, focus on philosophical issues concerning the notion of action or "seeing to it that" and there is no explicit consideration of game theoretic issues. Furthermore, while we make use of view of standard tense logic, those papers rely on the more complex "Ockhamist" semantics, where the truth of a formula is not evaluated at a single point in time, but at a pair consisting of a time point and a branch or history through it; the future operator then refers to time points in this branch only and, therefore, the resulting logic is that of linear time. A further operator is then added to capture the notion of historical necessity and contingency.

- for every $(i, a) \in N \times A$, R_{ia} is a binary relation on T satisfying the following property:

  (P.3) R_{ia} subrelation of $\prec$: if $t_1 R_{ia} t_2$ then $t_1 \prec t_2$.

Properties (P.0)-(P.2) constitute the definition of *branching time* in temporal logic.[8] In particular, (P.2) expresses the notion that, while a given moment may have different possible futures, its past is settled. The interpretation of $t_1 R_{ia} t_2$ is that at time t_1 agent i has available action a which leads from t_1 to t_2. Property (P.3) expresses the notion that actions can only affect the future. It is possible that for some i and t, $R_{ia}(t) \stackrel{def}{=} \{t' \in T : tR_{ia}t'\}$ is empty for all $a \in A$. In such a case agent i does not have any actions available at time t. [9]

For every $i \in N$, let $R_i \stackrel{def}{=} \bigcup_{a \in A} R_{ia}$. Thus tR_it' if and only if agent i has available *some* action at t that leads from t to t'. [10]

Example 2.2. *The following is a BTA frame:* $T = \{t_1, t_2, ..., t_7\}$, $N = \{1, 2\}$, $\prec = \{(t_1, t_2), (t_1, t_4), (t_1, t_5), (t_2, t_4), (t_2, t_5), (t_1, t_3), (t_1, t_6), (t_1, t_7), (t_3, t_6), (t_3, t_7)\}$, $A = \{a, b\}$, $R_{1a} = \{(t_1, t_2)\}$, $R_{1b} = \{(t_1, t_3)\}$, $R_{2a} = \{(t_2, t_4), (t_3, t_6)\}$, $R_{2b} = \{(t_2, t_5), (t_3, t_7)\}$. *This frame is shown in Figure 1 where an arrow from* t *to* t' *indicates that* $t \prec t'$ *and all the arrows due to transitivity are deleted (thus Figure 1 is the Hasse diagram of* $\langle T, \prec \rangle$*); furthermore the label* i, a *is assigned to the arrow from* t *to* t' *if and only if* $(t, t') \in R_{ia}$.

[8] See, for example, Burgess (1984), Halpin (1988), Øhrstrøm and Hasle (1995).

[9] One could require actions to be deterministic, by imposing that if $tR_{ia}t'$ and $tR_{ia}t''$ then $t' = t''$. However, in general the effect of an action may depend on external factors. For example, the action of opening the window may lead from state t where the window is closed to either state t' where the window is open and it rains or to state t'' where the window is open and it does not rain. Requiring actions to be deterministic would then make it necessary to add "Nature" to the set of agents (in the example, Nature would choose between rain and no rain).

Another possible requirement is that actions be "time-deterministc" in the sense that if $tR_{ia}t'$ and $tR_{ia}t''$ then $t' \nprec t''$ and $t'' \nprec t'$.

[10] Note that simultaneous actions are not ruled out, that is, it is possible that, for some t and some i and j with $i \neq j$, both $R_i(t) \stackrel{def}{=} \{t' \in T : tR_it'\}$ and $R_j(t)$ are non-empty. In this case restrictions need to be imposed to guarantee that the actions of different agents are compatible with each other. For the purpose of this paper simultaneity of actions can be ignored.

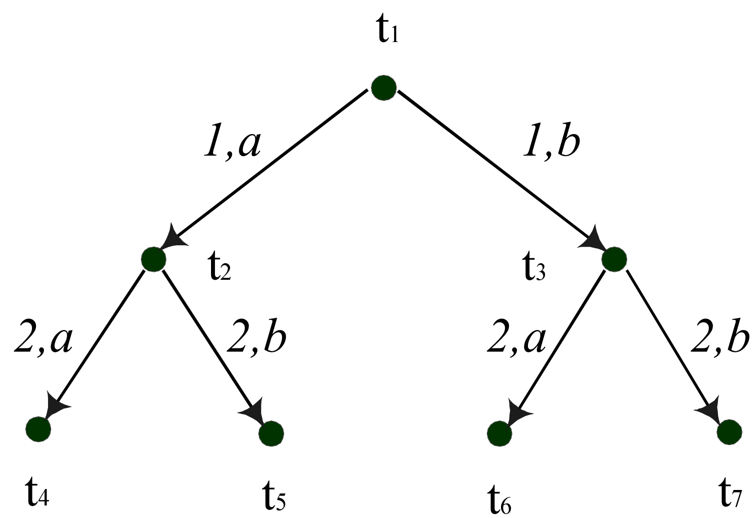

Figure 1

Every $t \in T$ should be thought of as a complete description of the world at time t, and sets of dates represent propositions. In order to establish this interpretation one needs to introduce a formal language and the notion of a model based on a frame. This will be done in Section 3.

It will be shown in Section 4 that extensive forms with perfect information are a special case of BTA frames.

Definition 2.3. *Given a BTA frame, a prediction for it is a binary relation $\prec_p$ on T satisfying the following properties:*

(P.4) $\prec_p$ subrelation of $\prec$: if $t_1 \prec_p t_2$ then $t_1 \prec t_2$.
(P.5) transitivity: if $t_1 \prec_p t_2$ and $t_2 \prec_p t_3$ then $t_1 \prec_p t_3$.
(P.6) $\prec_p$ is serial if $\prec$ is: $\forall t \in T$, if $\exists t_1$ s. t. $t \prec t_1$, then $\exists t_2$ s. t. $t \prec_p t_2$.
(P.7) time consistency: if $t_1 \prec t_2$, $t_2 \prec t_3$ and $t_1 \prec_p t_3$ then $t_1 \prec_p t_2$ and $t_2 \prec_p t_3$.

(P.4) expresses the notion that predicting the future consists in selecting a subset of the conceivable future states (those that are believed to be most likely).

The interpretation of $\prec_p$ in terms of prediction (i.e. belief about the future) makes (P.5) a natural requirement: it can be viewed as incorporating a principle of coherence of belief close in spirit to van Fraassen's Reflection Principle (van Frassen, 1984). (P.6) requires that a prediction be *complete*, in the sense that a prediction be made whenever possible: if there is a conceivable future of t (that is, if $\prec$ is serial at t) then there must be a predicted future of t (that is, $\prec_p$ is serial at t). Property (P.7) says the following. Suppose that at time t_1 a conceivable future development is represented by the path $t_1 t_2 t_3$ (that is, $t_1 \prec t_2$ and $t_2 \prec t_3$): this is shown in Figure 2, where, as before, a continuous arrow from t to t' denotes that $t \prec t'$. Suppose also that t_3 lies in the predicted future of t_1 (that is, $t_1 \prec_p t_3$): this is shown in Figure 2 by a dotted arrow from t_1 to t_3. Then (P.7) imposes the following requirements:

(a) since reaching t_3 requires going through t_2, t_2 should lie in the predicted future of t_1 (that is, $t_1 \prec_p t_2$), and

(b) since reaching t_2 is consistent with (is a partial realization of) the prediction that t_3 will be reached, the prediction should continue to hold at t_2, that is, t_3 should be in the predicted future of t_2 ($t_2 \prec_p t_3$).

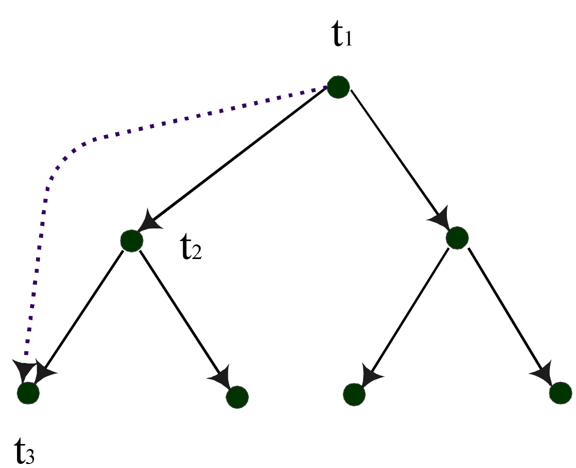

Figure 2

Example 2.4. *For the BTA frame of Example 2.2 (cf. Figure 1) the following is a prediction according to Definition 2.3:* $\prec_p = \{(t_1, t_3), (t_1, t_6), (t_3, t_6), (t_2, t_5)\}$.

This is represented in Figure 3 by a dotted line next to an arrow that belongs to both $\prec$ and $\prec_p$, omitting dotted lines that can be obtained by transitivity (thus the dotted lines alone represent the Hasse diagram of $\langle T, \prec_p \rangle$).

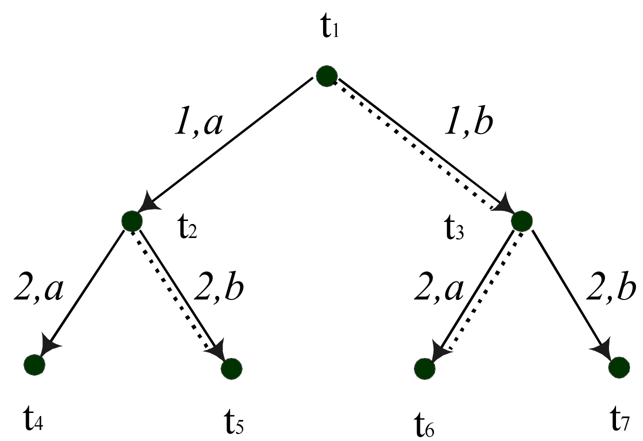

Figure 3

It will be shown in Section 4 that the backward induction solution of an extensive game with perfect information is a prediction in the sense of Definition 2.3.

Definition 2.5. *An augmented BTA frame is a BTA frame together with a prediction for it.*

3. Syntax

We consider a propositional language with several modal operators:

- Tense and prediction operators: G, H, G_p and H_p. The intended interpretation is as follows:

$G\phi$: "it is going to be the case at every future date that ϕ"
$H\phi$: "it has always been the case that ϕ"
$G_p\phi$: "it is going to be the case in every *predicted* future that ϕ"
$H_p\phi$: "it has always been the case at every past date at which today was predicted that ϕ"

- Action operators: $\Box_{ia}$ (for every $(i,a) \in N \times A$), whose intended interpretation is:

$\Box_{ia}\phi$: "after agent i takes action a, it is the case that ϕ"

The formal language is built in the familiar way from the following components: a countable set S of sentence letters (representing atomic propositions), the connectives $\neg$ and $\vee$ (from which the other connectives $\wedge$, $\rightarrow$ and $\leftrightarrow$ are defined as usual) and the above modal operators.[11] Let $F\phi \overset{def}{=} \neg G\neg\phi$, $P\phi \overset{def}{=} \neg H\neg\phi$, $F_p\phi \overset{def}{=} \neg G_p\neg\phi$, $P_p\phi \overset{def}{=} \neg H_p\neg\phi$ and, for every $i \in N$, $\Box_i\phi \overset{def}{=} \bigwedge_{a\in A} \Box_{ia}\phi$. Thus the intended interpretation is:

$F\phi$: "at some future date it will be the case that ϕ "

$P\phi$: "at some past date it was the case that ϕ "

$F_p\phi$: "at some *predicted* future date it will be the case that ϕ "

$P_p\phi$: "at some past date at which today was predicted it was the case that ϕ "

$\Box_i\phi$: "no matter what action agent i takes, it will be the case that ϕ".

Remark 1. *The notion that agent i has the power to bring about that ϕ or has control over ϕ can be expressed by the conjunction*

$$\left(\bigvee_{a\in A} \Box_{ia}\phi\right) \wedge \neg\Box_i\phi$$

(that is, the agent can bring about that ϕ with some action and it is not the case that ϕ holds no matter what the agent does).[12]

[11] The set Φ of formulae is thus obtained from the sentence letters by closing with respect to negation, disjunction and the operators G, H, G_p, H_p and $\Box_{ia}$: (i) for every $p \in S$, $(p) \in \Phi$, (ii) if $\phi, \psi \in \Phi$ then all of the following belong to Φ: $(\neg\phi)$, $(\phi \vee \psi)$, $G\phi$, $H\phi$, $G_p\phi$, $H_p\phi$ and $\Box_{ia}\phi$.

[12] Thus we side with Chellas (1992) in finding it more desirable not to include a negative condition in the definition of "agent i brings about (or sees to it) that ϕ". See, in particular, the discussion in Horty and Belnap (1995, pp. 599-600).

Given an augmented BTA frame one obtains a *model* $\mathcal{M}$ *based on it* by adding a function $V : S \to 2^T$ (where 2^T denotes the set of subsets of T) that associates with every sentence letter p the set of dates at which p is true. For non-modal formulae truth at a point in a model is defined as usual.[13] Validation for modal formulae is as follows:[14]

$\mathcal{M}, t \models G\phi$ iff $\mathcal{M}, t' \models \phi$ for all t' such that $t \prec t'$.

$\mathcal{M}, t \models H\phi$ iff $\mathcal{M}, t'' \models \phi$ for all t'' such that $t'' \prec t$.

$\mathcal{M}, t \models G_p\phi$ iff $\mathcal{M}, t' \models \phi$ for all t' such that $t \prec_p t'$.

$\mathcal{M}, t \models H_p\phi$ iff $\mathcal{M}, t'' \models \phi$ for all t'' such that $t'' \prec_p t$.

$\mathcal{M}, t \models \Box_{ia}\phi$ iff $\mathcal{M}, t' \models \phi$ for all t' such that $tR_{ia}t'$.

It follows from the definitions of $\Box_i$ and R_i that

$\mathcal{M}, t \models \Box_i\phi$ iff $\mathcal{M}, t' \models \phi$ for all t' such that tR_it'.

A formula ϕ is *valid in model* $\mathcal{M}$ if $\mathcal{M}, t \models \phi$ for all $t \in T$; it is *valid on a frame* if it is valid in every model based on it.

The semantics of augmented BTA frames can be axiomatized as follows. Denote by $\mathbb{L}_0$ the basic system specified by the following axiom schemata and rules of inference.

Axiom schemata: all the classical tautologies as well as the following

(A.0a) $G(\phi \to \psi) \to (G\phi \to G\psi)$	(A.0b) $G_p(\phi \to \psi) \to (G_p\phi \to G_p\psi)$
(A.0c) $H(\phi \to \psi) \to (H\phi \to H\psi)$	(A.0d) $H_p(\phi \to \psi) \to (H_p\phi \to H_p\psi)$
(A.0e) $\Box_{ia}(\phi \to \psi) \to (\Box_{ia}\phi \to \Box_{ia}\psi)$	
(A.0f) $\phi \to GP\phi$	(A.0g) $\phi \to G_pP_p\phi$
(A.0h) $\phi \to HF\phi$	(A.0i) $\phi \to H_pF_p\phi$

[13] $\mathcal{M}, t \models \phi$ denotes that ϕ *is true at time* t *in model* $\mathcal{M}$ and $\mathcal{M}, t \not\models \phi$ denotes that ϕ is false at t. For a sentence letter p, $\mathcal{M}, t \models p$ iff $t \in V(p)$; furthermore, $\mathcal{M}, t \models \neg\phi$ iff $\mathcal{M}, t \not\models \phi$ and $\mathcal{M}, t \models (\phi \vee \psi)$ iff either $\mathcal{M}, t \models \phi$ or $\mathcal{M}, t \models \psi$. It follows that $\mathcal{M}, t \models (\phi \wedge \psi)$ iff $\mathcal{M}, t \models \phi$ and $\mathcal{M}, t \models \psi$, and $\mathcal{M}, t \models (\phi \to \psi)$ iff $\mathcal{M}, t \models \phi$ implies $\mathcal{M}, t \models \psi$.

[14] Thus
$\mathcal{M}, t \models F\phi$ iff $\mathcal{M}, t' \models \phi$ for some t' with $t \prec t'$
$\mathcal{M}, t \models P\phi$ iff $\mathcal{M}, t'' \models \phi$ for some t'' with $t'' \prec t$
$\mathcal{M}, t \models F_p\phi$ iff $\mathcal{M}, t' \models \phi$ for some t' with $t \prec_p t'$
$\mathcal{M}, t \models P_p\phi$ iff $\mathcal{M}, t'' \models \phi$ for some t'' with $t'' \prec_p t$.

Rules of inference:

Modus Ponens: from ϕ and $\phi \to \psi$ to infer ψ,

Necessitation: from ϕ to infer $G\phi$, $H\phi$, $G_p\phi$, $H_p\phi$ and $\Box_{ia}\phi$.

Let $\mathbb{L}_1$ be the extension of $\mathbb{L}_0$ obtained by adding the following axiom schemata:[15]

(A.1) $G\phi \to GG\phi$

(A.2) $P\phi \wedge P\psi \to P(\phi \wedge \psi) \vee P(\phi \wedge P\psi) \vee P(P\phi \wedge \psi)$

(A.3) $G\phi \to \Box_{ia}\phi$

(A.4) $G\phi \to G_p\phi$

(A.5) $G_p\phi \to G_pG_p\phi$

(A.6) $G_p\phi \wedge F\phi \to F_p\phi$

(A.7a) $P_p\phi \wedge P_p\psi \to P_p(\phi \wedge \psi) \vee P_p(\phi \wedge P_p\psi) \vee P_p(P_p\phi \wedge \psi)$

(A.7b) $P_p\phi \wedge P\psi \to P_p(\phi \wedge \psi) \vee P_p(\phi \wedge P\psi) \vee P_p(P\phi \wedge \psi)$

A formula ϕ is a *theorem* of $\mathbb{L}_1$ iff it can be obtained in a finite number of steps from the axioms using the rules of inference, that is, iff there is a sequence $\langle \phi_1, ..., \phi_m \rangle$ such that (i) $\phi_m = \phi$ and (ii) each ϕ_j is either an axiom or is obtained from one or more ϕ_k with $k < j$ by using a rule of inference.

Proposition 3.1. *(Soundness and Completeness). The following are equivalent:*

(1) ϕ is a theorem of $\mathbb{L}_1$,

(2) ϕ is valid on every augmented BTA frame.

We omit the proof of Proposition 3.1. Bonanno (1998) proves soundness and completeness for the system without agents (thus, on the semantic side, for frames without the relations R_{ia} and, on the syntactic side, for a logic without the operators $\Box_{ia}$ and, therefore, without axioms (A.0e) and (A.3)). The proof of Proposition 3.1 is an extension of that result.

[15] The axioms have been numbered so as to correspond to the properties of frames. Thus, for $0 \leq j \leq 6$, axiom (A.j) corresponds to property (P.j) in the sense that a frame satisfies property (P.j) if and only if axiom (A.j) is valid on that frame. Similarly, as shown in Bonanno (1998), property (P.7) corresponds to the conjunction of (A.7a) and (A.7b).

4. Extensive games with perfect information

In this section we show that an extensive game with perfect information is a special case of a BTA frame and that the backward induction solution is a special case of a prediction. In Section 5 we provide a characterization of backward induction.

Recall that a *rooted tree* is a pair $\langle T, \rightarrowtail \rangle$ where T is a set of *nodes* and $\rightarrowtail$ is a binary relation on T (if $t \rightarrowtail t'$ we say that t *immediately precedes* t' or that t' *immediately succeeds* t) satisfying the following properties:

1. there is a unique node t_0 with no immediate predecessors; it is called the *root*,

2. for every node $t \in T \backslash \{t_0\}$ there is a unique path from t_0 to t, that is, there is a unique sequence $\langle x_1, ..., x_m \rangle$ in T with $x_1 = t_0$, $x_m = t$, and, for every $j = 1, ...m - 1$, $x_j \rightarrowtail x_{j+1}$.

Given a rooted tree $\langle T, \rightarrowtail \rangle$, a *terminal node* is a $t \in T$ which has no immediate successors. Let $Z \subseteq T$ denote the set of terminal nodes. It is easy to see that if T is finite then $Z \neq \emptyset$.

Definition 4.1. *A finite extensive form with perfect information is a tuple* $\langle T, \rightarrowtail, N, \iota \rangle$ *where* $\langle T, \rightarrowtail \rangle$ *is a finite rooted tree,* $N = \{1, ..., n\}$ *is a set of players and* $\iota : T \backslash Z \to N$ *is a function that associates with every non-terminal or decision node the player who moves at that node. If* $i = \iota(t)$ *and* $t \rightarrowtail t'$ *we say that the pair* (t, t') *is a choice of player* i *at node* t.

Figure 4 below shows an example of an extensive form with perfect information (ignoring payoffs).

Lemma 4.2. *A finite extensive form with perfect information is a special case of a BTA frame (cf. Definition 2.1).*

Proof. Let $\prec$ be the transitive closure of $\rightarrowtail$, that is, $t \prec t'$ iff there is a path from t to t'. It is straightforward to show that $\prec$ satisfies properties (P.0)-(P.3) of Definition 2.1. Furthermore, let A be a set of labels which is in one-to-one correspondence with $\rightarrowtail$ (viewed as a set of ordered pairs).[16] Given an arbitrary $(t, t') \in \rightarrowtail$, if $a \in A$ is the corresponding label and $i = \iota(t)$, then (1) $R_{ia} = \{(t, t')\}$ and (2) for every $j \neq i$, $R_{ja} = \emptyset$. It is obvious that property (P.3) is satisfied. ■

[16] There are other ways in which the set A of actions could be defined, e.g. one could take a set of labels with cardinality equal to the maximum outdegree among the nodes in T.

Definition 4.3. *Given a finite extensive form with perfect information one obtains a perfect information game by adding, for every $i \in N$, a payoff or utility function $u_i : Z \to \mathbb{Q}$ (where Z is the set of terminal nodes and $\mathbb{Q}$ is the set of rational numbers).*

Figure 4 shows a perfect information game with three players. The vector (x_1, x_2, x_3) written next to a terminal node z represents the payoff vector $(u_1(z), u_2(z), u_3(z))$ and there is an arrow from t to t' if and only if $t \rightarrowtail t'$. For every decision node t, the corresponding player $\iota(t)$ is written next to it.

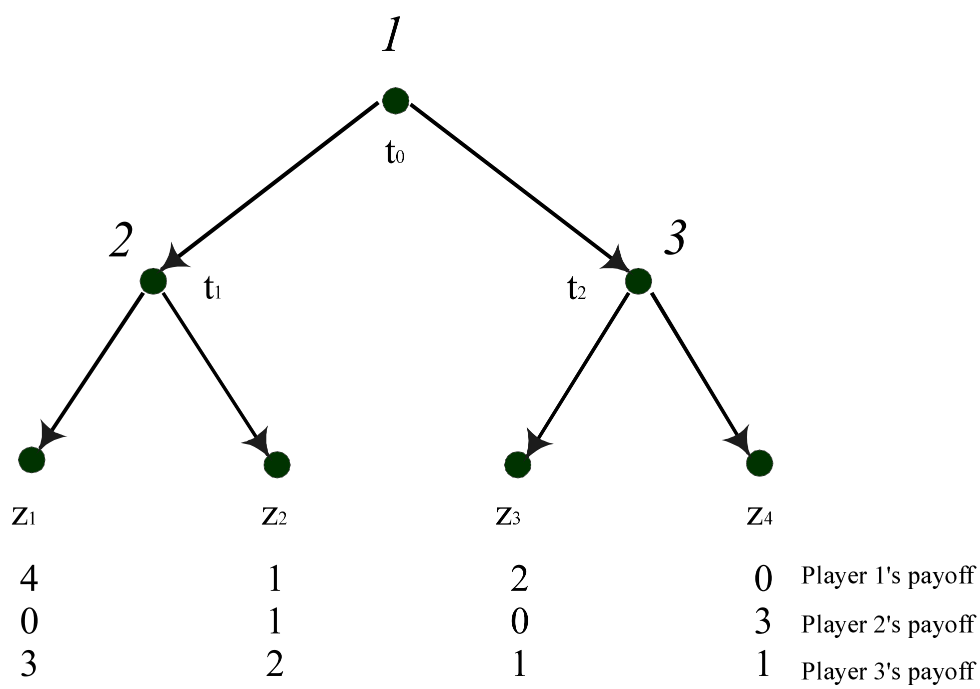

Figure 4

A well-known procedure for solving a perfect information game is the *backward induction* algorithm first used by Zermelo (1913) for the game of chess. The algorithm starts at the end of the game and proceeds backwards towards the root:

1. Start from a decision node t whose immediate successors are only terminal nodes (e.g. node t_1 in Figure 4) and select one choice that maximizes the utility of player $\iota(t)$ (in the example of Figure 4, at t_1 player 2 would make the choice that leads to node z_2 since it gives her a payoff of 1 rather than 0, which is the payoff that she would get if the play proceeded to node z_1). Delete the immediate successors of t and assign to t the payoff vector associated with the selected choice.

2. Repeat step 1 until all the decision nodes have been exhausted.

Figure 5 shows a possible outcome of the backward induction algorithm for the game of Figure 4. The choices selected by the algorithm are shown as dotted lines next to the corresponding arrows.

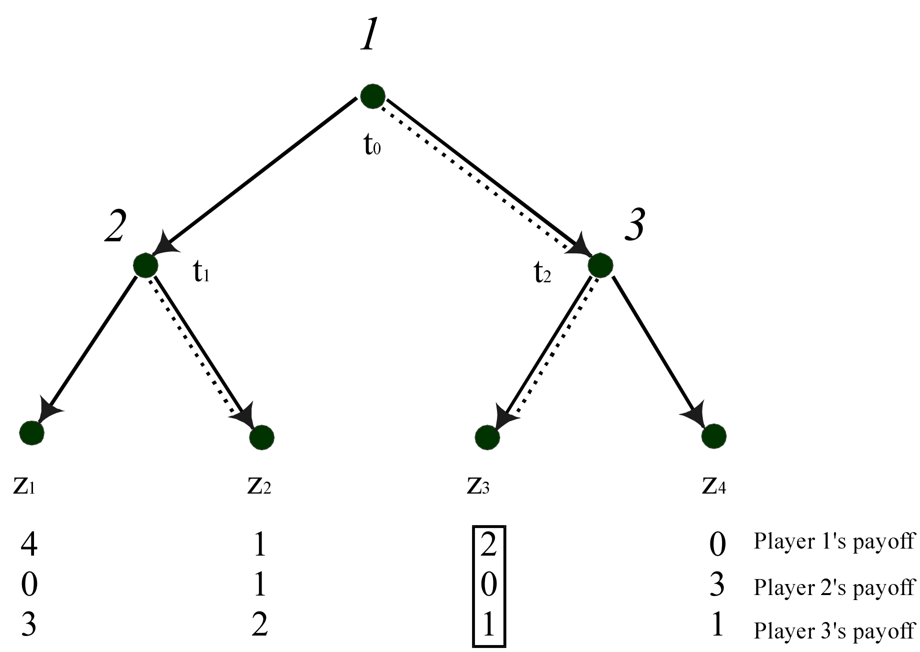

Figure 5

Note that the backward induction algorithm may yield more than one solution. Multiplicity may arise if there are players who have more than one utility-maximizing choice. For example, in the game of Figure 4 at t_2 both choices are optimal for Player 3. The selection of choice (t_2, z_3) leads to the solution shown in Figure 5, while the selection of choice (t_2, z_4) leads to a different solution shown in Figure 6.

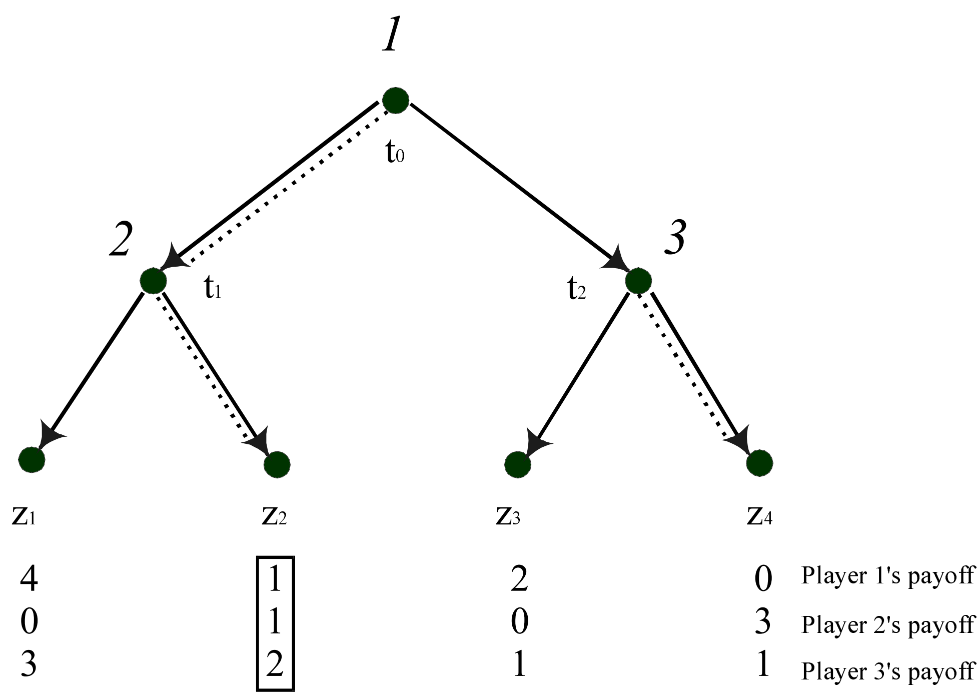

Figure 6

Definition 4.4. *A perfect information game is generic if no player is indifferent between any two terminal nodes, that is, if* $\forall i \in N$, $\forall z, z' \in Z$ *if* $u_i(z) = u_i(z')$ *then* $z = z'$.

Remark 2. *In a generic game the backward induction algorithm yields a unique solution.*

The above examples suggest a similarity between solutions obtained using the backward induction algorithm and the notion of prediction given in Definition 2.3.

We now show that indeed a backward-induction solution is a prediction. To do this we need to give a more precise definition of backward-induction.

Definition 4.5. *Given a finite perfect information game* $\langle T, \rightarrowtail, N, \iota, \{u_i\}_{i \in N} \rangle$, *the set* $T_k \subseteq T$ *of level* k *nodes (with* $k \geq 0$*) is defined recursively as follows:*

(1) $T_0 = Z$ *(that is, level 0 nodes are all and only the terminal nodes),*

(2) for $k \geq 1$, $t \in T_k$ *iff (a)* $t \in T \backslash Z$, *(b) every immediate successor of* t *is a node of level not greater than* $k-1$, *and (c) at least one immediate successor of* t *is of level* $k-1$.

We denote by $\ell(t)$ the level of node t (thus $t \in T_{\ell(t)}$). Note that a node t is of level k iff k is the length of the maximal path from t to a terminal node, as illustrated in Figure 7.

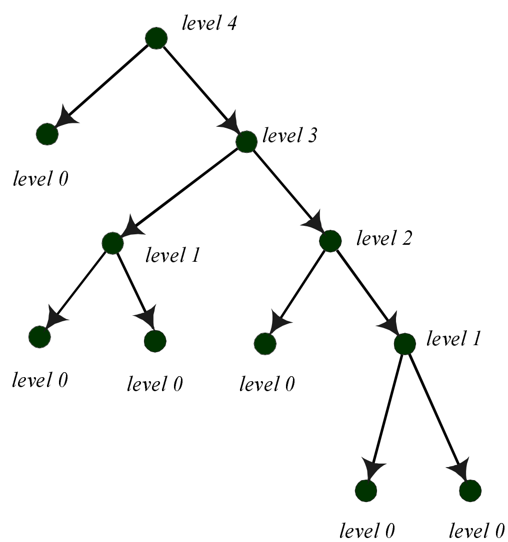

Figure 7

Definition 4.6. *Given a finite perfect information game* $\langle T, \rightarrowtail, N, \iota, \{u_i\}_{i \in N} \rangle$ *define, for* $k \geq 1$, *a binary relation* $\rightarrowtail_{BI}^k$ *on* T *and, for every* $i \in N$, *the function* $u_i^k : T_k \to \mathbb{Q}$ *recursively as follows:*

- *definition of* $\rightarrowtail_{BI}^{1}$:

  (1) if $t \rightarrowtail_{BI}^{1} t'$ *then (a)* $t \in T_1$ *(that is, t is a level-1 node) and* $t \rightarrowtail t'$, *(b)* $u_{\iota(t)}(t') \geq u_{\iota(t)}(t'')$ *for all* t'' *such that* $t \rightarrowtail t''$, *(c) if* $t \rightarrowtail_{BI}^{1} t'$ *and* $t \rightarrowtail_{BI}^{1} t''$ *then* $t' = t''$ *and*

  (2) $t \rightarrowtail_{BI}^{1} t'$ *for some* t';[17]

- *definition of* $u_i^1 : T_1 \to \mathbb{Q}$: $u_i^1(t) = u_i^0(t')$ *where* $u_i^0 = u_i$ *and* t' *is the unique node such that* $t \rightarrowtail_{BI}^{1} t'$;[18]

- *definition of* $\rightarrowtail_{BI}^{k}$ *for* $k > 1$:

  (1) if $t \rightarrowtail_{BI}^{k} t'$ *then (a)* $t \in T_k$ *(that is, t is a level-k node) and* $t \rightarrowtail t'$, *(b)* $u_{\iota(t)}^{\ell(t')}(t') \geq u_{\iota(t)}^{\ell(t'')}(t'')$ *for all* t'' *such that* $t \rightarrowtail t''$, *(c) if* $t \rightarrowtail_{BI}^{k} t'$ *and* $t \rightarrowtail_{BI}^{k} t''$ *then* $t' = t''$ *and*

  (2) $t \rightarrowtail_{BI}^{k} t'$ *for some* t';

- *definition of* $u_i^k : T_k \to \mathbb{Q}$: $u_i^k(t) = u_i^{\ell(t')}(t')$ *where* t' *is the unique node such that* $t \rightarrowtail_{BI}^{k} t'$.

For the example of Figure 6 above, we have: $\rightarrowtail_{BI}^{1}= \{(t_1, z_2), (t_2, z_4)\}$, $\rightarrowtail_{BI}^{2}= \{(t_0, t_1)\}$, $(u_1^1(t_2), u_2^1(t_2), u_3^1(t_2)) = (0, 3, 1)$, $(u_1^1(t_1), u_2^1(t_1), u_3^1(t_1)) = (1, 1, 2)$, $(u_1^2(t_0), u_2^2(t_0), u_3^2(t_0)) = (1, 1, 2)$.

Definition 4.7. *Given a finite perfect information game* $\langle T, \rightarrowtail, N, \iota, \{u_i\}_{i \in N} \rangle$ *a binary relation* $\rightarrowtail_{BI}$ *on* T *is called a backward induction relation if*

$$\rightarrowtail_{BI} = \bigcup_{k=1}^{\ell(t_0)} \rightarrowtail_{BI}^{k}$$

where the relations $\rightarrowtail_{BI}^{k}$ *are obtained according to Definition 4.6.*

[17] Thus $\rightarrowtail_{BI}^{1}$ mimics the first step of the backward induction algorithm: for every "last decision node" t, $\rightarrowtail_{BI}^{1}$ associates with t a unique immediate successor t' which maximizes the payoff of the player assigned to node t.

[18] Thus, for every player $i \in N$, u_i^1 associates with a level-1 decision node t the payoff associated with the terminal node t' selected by $\rightarrowtail_{BI}^{1}$. This definition corresponds to the step in the backward-induction algorithm of pruning the tree and making t a terminal node with the payoff vector associated with the terminal node that follows the choice selected at t.

Thus, for the example of Figure 6, $\rightarrowtail_{BI}= \{(t_0, t_1), (t_1, z_2), (t_2, z_4)\}$.[19]Note that a given perfect information game might have more that one backward-induction relation. For example, for the game of Figure 4, one backward induction relation is the one just described, which is illustrated in Figure 6, and a different one is $\rightarrowtail_{BI}= \{(t_0, t_2), (t_1, z_2), (t_2, z_3)\}$, which is illustrated in Figure 5.

The next lemma shows that a backward-induction relation of a perfect information game can be viewed as a prediction according to Definition 2.3.

Lemma 4.8. *Let $\langle T, \rightarrowtail, N, \iota, \{u_i\}_{i\in N}\rangle$ be a finite perfect information game and $\rightarrowtail_{BI}$ a backward induction relation for it. Let $\prec_p$ be the transitive closure of $\rightarrowtail_{BI}$. Then $\prec_p$ is a prediction in the sense of Definition 2.3.*

Proof. We need to show that $\prec_p$ satisfies properties (P.4)-(P.7) of Definition 2.3. First of all, it is clear from Definition 4.6 that $\prec_p$ is a subrelation of $\prec$ (the transitive closure of $\rightarrowtail$: see Lemma 4.2). By construction, $\prec_p$ is transitive. It is easy to see from Definition 4.6 that t is such that there is no t' with $t \prec_p t'$ only if t is a terminal node (which is also the only case where there is no t' with $t \prec t'$); thus property (P.6) is satisfied. Finally, if $t_1 \prec_p t_3$ and $t_1 \prec t_2$ and $t_2 \prec t_3$ then: (1) by definition of $\prec$, there is a $\rightarrowtail$-path from t_1 to t_3 through t_2, (2) by definition of $\prec_p$, there is a $\rightarrowtail_{BI}$-path from t_1 to t_3, which, since $\rightarrowtail_{BI}$ is a subrelation of $\rightarrowtail$, is also a $\rightarrowtail$-path from t_1 to t_3. By definition of tree, the $\rightarrowtail$-path from t_1 to t_3 is unique; hence the $\rightarrowtail_{BI}$-path from t_1 to t_3 goes through t_2 . Thus, by definition of $\prec_p$, we have that $t_1 \prec_p t_2$ and $t_2 \prec_p t_3$, that is, property (P.7) is satisfied. ■

Definition 4.9. *Given a perfect information game $\langle T, \rightarrowtail, N, \iota, \{u_i\}_{i\in N}\rangle$, a relation $\prec_p$ on T is called a backward induction prediction if $\prec_p$ is the transitive closure of a backward-induction relation $\rightarrowtail_{BI}$ for that game.*

Remark 3. *Every finite perfect information game has at least one backward-induction prediction, although, as noted above, it may have more than one. However, in generic games (cf. Definition 4.4) the is a unique backward-induction prediction.*

Remark 4. *It is clear from Definitions 4.6, 4.7 and 4.9 that,*

[19] In game-theoretic terms, $\rightarrowtail_{BI}$ corresponds to the strategy profile associated with a backward induction solution.

(a) if $t \prec_p t'$ and $t \prec_p t''$ and both t' and t'' are immediate successors of t then $t' = t''$,

(b) for every decision node $t \in T \backslash Z$, there is a unique $z \in Z$ such that $t \prec_p z$.

5. A characterization of backward induction

The relationship between an extensive *form* with perfect information and a perfect information *game* is similar to the relationship between a frame and a model. Lemma 4.2 shows that an extensive form with perfect information is a special case of a BTA frame. To view a perfect information *game* as a model (as defined in Section 3) all we need to do is include in the set of sentences (or atomic propositions) sentences of the form $(u_i = q)$ with $i \in N$ and $q \in \mathbb{Q}$, whose intended interpretation is "player i's utility (or payoff) is q". We also need to add the standard ordering of the rational numbers in the form of sentences of the form $(q_1 \leq q_2)$ whose intended interpretation is "the rational number q_1 is less than or equal to the rational number q_2". A *game language* is a language obtained as explained in Section 3 from a set of sentences S that includes atomic propositions of the form $(u_i = q)$ and $(q_1 \leq q_2)$.

Definition 5.1. *Let $\mathcal{G}$ be a perfect information game and $\mathcal{F}$ be the corresponding BTA frame (cf. Lemma 4.2). A game model is a model based on $\mathcal{F}$ (cf. Section 3) obtained in a game language by adding to $\mathcal{F}$ a valuation $V : S \rightarrow 2^T$ satisfying the following properties:*

- if $p \in S$ is of the form $(q_1 \leq q_2)$ with $q_1, q_2 \in \mathbb{Q}$ then

$$V(p) = T \text{ if } q_1 \leq q_2 \text{ and } V(p) = \emptyset \text{ otherwise}$$

- if $p \in S$ is of the form $(u_i = q)$ then

$$V(p) = \{z \in Z : u_i(z) = q\}.$$

Thus if $\mathcal{M}$ is a game model then, $\forall t \in T$, $\mathcal{M}, t \models (q_1 \leq q_2)$ if q_1 is less than or equal to q_2 and $\mathcal{M}, t \models \neg(q_1 \leq q_2)$ otherwise; furthermore, $\mathcal{M}, t \models (u_i = q)$ if t is a terminal node with $u_i(t) = q$ and $\mathcal{M}, t \models \neg(u_i = q)$ if t is either a decision node or a terminal node with $u_i(t) \neq q$. The valuation of the other atomic formulae and of the non-atomic formulae is as explained in Section 3.

Consider the following axiom scheme:

$$F_p(u_i = q) \to \Box_i \left(((u_i = r) \vee F_p(u_i = r)) \to (r \leq q)\right) \tag{IC}$$

(IC) says that if at some predicted future time Player i's payoff is q then, no matter what action Player i takes, it will be the case that if Player i's payoff is, or is predicted to be, r then r is not greater than q. It we think of the prediction as a "recommendation" to the players, then (IC) says that if the recommendation is that (the game be played in such a way that) Player i get a payoff of q then it is not possible for Player i to take an action after which his payoff is greater than q or the recommendation is that (the game be played in such a way that) Player i get a payoff greater than q. Thus (IC) can be viewed as expressing a notion of *internal consistency* of prediction or recommendation (hence the name IC), in the sense that no player can increase his payoff by deviating from the recommendation, using the recommendation itself to predict his future payoff after the deviation.[20]

The following propositions show that axiom (IC) characterizes the notion of backward induction.

Proposition 5.2. *Let $\mathcal{G}$ be a perfect information game and $\prec_p$ a backward induction prediction for $\mathcal{G}$ (cf. Definition 4.9). Then axiom (IC) is valid in every game model based on the augmented frame $\langle \mathcal{F}, \prec_p \rangle$, where $\mathcal{F}$ is the BTA frame associated with $\mathcal{G}$ (cf. Lemma 4.2).*

Proof. Fix an arbitrary game model $\mathcal{M}$ based on $\langle \mathcal{F}, \prec_p \rangle$. We have to show that every instance of (IC) is true at every $t \in T$. If t is a terminal node, then $\{t' \in T : t \prec_p t'\} = \emptyset$ and therefore $\mathcal{M}, t \models \neg F_p(u_i = q)$ for all $i \in N$ and $q \in \mathbb{Q}$. Thus (IC) is true at t. If t be a decision node and $i \neq \iota(t)$ then $R_i(t) = \emptyset$ and therefore $\mathcal{M}, t \models \Box_i \phi$ for every formula ϕ; hence (IC) is true at t. Thus we only need to consider the case where t is a decision node and $i = \iota(t)$. Suppose that (IC) is false at t. Then there are numbers $q, r \in \mathbb{Q}$ such that

$$\mathcal{M}, t \models F_p(u_i = q) \tag{5.1}$$

and $\mathcal{M}, t \not\models \Box_i \left(((u_i = r) \vee F_p(u_i = r)) \to (r \leq q)\right)$, that is,

[20] As remarked in the introduction, the notion of internal consistency is due to von Neumann and Morgenstern (1947) and is central to Joseph Greenberg's (1990) theory of social situations.

$$\exists t' \in T : tR_i t' \text{ and } \mathcal{M}, t' \models ((u_i = r) \vee F_p(u_i = r)) \wedge \neg(r \leq q)). \tag{5.2}$$

By Remark 4 (Section 3) there is a unique $z \in Z$ such that $t \prec_p z$. By (5.1) $u_i(z) = q$. Let t'' be the unique immediate successor of t on the $\prec_p$-path from t to z. By definition of R_i (cf. Lemma 4.2), the t' of (5.2) is also an immediate successor of t. Let z' be the unique terminal node such that $t' \prec_p z'$. Then, by (5.2), $u_i(z') = r$ and $r > q$. Thus

$$u_i(z') > u_i(z). \tag{5.3}$$

By Definition 4.6, $u_i^{\ell(t'')}(t'') = u_i(z)$, $u_i^{\ell(t')}(t') = u_i(z')$ and $u_i^{\ell(t'')}(t'') \geq u_i^{\ell(t')}(t')$, contradicting (5.3). ■

The next proposition gives a converse to Proposition 5.2 for generic games (cf. Definition 4.4).

Proposition 5.3. *Let $\mathcal{G}$ be a generic perfect information game, $\mathcal{F}$ the associated BTA frame and $\prec_p$ a prediction for $\mathcal{F}$. Let $\mathcal{M}$ be any game model based on $\langle \mathcal{F}, \prec_p \rangle$ (cf. Definition 5.1). If axiom (IC) is valid in $\mathcal{M}$ then $\prec_p$ is the backward induction prediction.*[21]

Proof. First of all, by property (P.4) of Definition 2.3 ($\prec_p$ subrelation of $\prec$), all predictions coincide when restricted to the set of level 0 (or terminal) nodes (they are equal to the empty set). Thus, in particular, $\prec_p$ restricted to T_0 coincides with the backward-induction prediction restricted to T_0. Now we show that $\prec_p$ restricted to T_1 (the set of level 1 nodes: cf. Definition 4.5) coincides with the restriction of the backward-induction prediction to T_1. Let $\hat{t} \in T_1$ and let $\hat{Z} = \{z \in Z : \hat{t} \rightarrowtail z\}$. By Properties (P.4) and (P.6) of Definition 2.3 ($\prec_p$ subrelation of $\prec$, and $\prec_p$ serial if $\prec$ is serial), $\hat{Z} \cap \{t \in T : \hat{t} \prec_p t\} \neq \emptyset$. Fix an arbitrary $\hat{z} \in \hat{Z} \cap \{t \in T : \hat{t} \prec_p t\}$. Then, letting $i = \iota(\hat{t})$ and $q = u_i(\hat{z})$,

$$\mathcal{M}, \hat{t} \models F_p(u_i = q). \tag{5.4}$$

[21] Recall that in generic games there is a unique backward induction prediction. Note also that the statements "(IC) is valid in *a* game model based on $\langle \mathcal{F}, \prec_p \rangle$" and "(IC) is valid in *every* game model based on $\langle \mathcal{F}, \prec_p \rangle$" are equivalent, since (IC) is made up only of atomic propositions of the form $(u_i = q)$ and $(r \geq q)$ and the valuations of different models coincide on this class of atomic propositions.

Furthermore, it must be the case that

$$q \geq u_i(z), \quad \forall z \in \hat{Z}. \tag{5.5}$$

In fact, suppose that, for some $z' \in \hat{Z}$, $u_i(z') = r > q$. Then $\mathcal{M}, z' \models (u_i = r) \wedge \neg(r \leq q)$. Since $\hat{t}R_i z'$, $\mathcal{M}, \hat{t} \models \neg\Box_i\left((u_i = r) \rightarrow (r \leq q)\right)$. Thus, by (5.4) (IC) would be false at $\hat{t}$, contrary to the hypothesis that (IC) is valid in $\mathcal{M}$. Since the game is generic, if $z \in \hat{Z}$ is such that $z \neq \hat{z}$ then, by (5.5), $u_i(z) < q$; it follows that $\{t \in T : \hat{t} \prec_p t\} = \{\hat{z}\}$. Thus, restricted to T_1, $\prec_p$ coincides with the backward induction prediction. Next we show that if $\prec_p$ and the backward-induction prediction coincide when restricted to $\bigcup_{j=0}^{k} T_k$ for $k \geq 1$, then they coincide when restricted to T_{k+1}. Fix an arbitrary $\hat{t} \in T_{k+1}$. By Property (P.6) of Definition 2.3, $\exists t'' \in T$ such that $\hat{t} \prec_p t''$. If t'' is not a terminal node, let t' be the unique immediate successor of $\hat{t}$ on the $\prec$-path from $\hat{t}$ to t''. Then, by Property (P.7) of Definition 2.3, $\hat{t} \prec_p t'$. Clearly, $\ell(t') \leq k$; hence, by our supposition that $\prec_p$ coincides with the backward-induction prediction when restricted to $\bigcup_{j=0}^{k} T_k$, there is a unique $z' \in Z$ such that $t' \prec_p z'$. Let $i = \iota(\hat{t})$and $q = u_i(z')$.Then

$$\mathcal{M}, \hat{t} \models F_p(u_i = q). \tag{5.6}$$

For every $t \in T$ such that $\hat{t} \rightarrowtail t$, if t is not a terminal node let z_t be the unique terminal node such that $t \prec_p z_t$ (once again, uniqueness is guaranteed by our supposition; if t is a terminal node, let $z_t = t$). We want to show that

$$u_i(z_{t'}) \geq u_i(z_t), \quad \forall t \in T : \hat{t} \rightarrowtail t \tag{5.7}$$

Suppose not. Then there exists a $\tilde{t} \in T$ such that $\hat{t} \rightarrowtail \tilde{t}$ and $u_i(z_{\tilde{t}}) = r > q = u_i(z_{t'})$. Two cases are possible: (1) $\tilde{t} \in Z$, or (2) $\tilde{t} \notin Z$. In case (1), $\mathcal{M}, \tilde{t} \models (u_i = r) \wedge \neg(r \leq q)$,while in case (2) $\mathcal{M}, \tilde{t} \models F_p(u_i = r) \wedge \neg(r \leq q)$. Thus in either case $\mathcal{M}, \hat{t} \models \neg\Box_i\left(((u_i = r) \vee F_p(u_i = r)) \rightarrow (r \leq q)\right)$. Hence, by (5.6), (IC) is false at $\hat{t}$, contradicting the hypothesis that (IC) is valid in $\mathcal{M}$. Since the game is generic, it follows from (5.7) that $\{z \in Z : \hat{t} \prec_p z\} = \{z_{t'}\}$ and, therefore, if t is an immediate successor of $\hat{t}$ and $\hat{t} \prec_p t$ then $t = t'$. Thus the restriction of $\prec_p$ to T_{k+1} coincides with the restriction to T_{k+1} of the backward induction prediction. ■

The reason why Proposition 5.3 is not true, as stated, for non-generic games is that (cf. Remark 4), while a backward induction prediction is such that the predicted future of any node t is always a unique path, in non-generic games it is possible to satisfy (IC) with a relation that includes more than one path out of some nodes. This is illustrated in Figure 8 below, where (a) and (b) are the only backward induction relations, while the relation illustrated in (c) is not a backward-induction relation; however, it is easy to see that all three validate (IC) in every model based on this game.

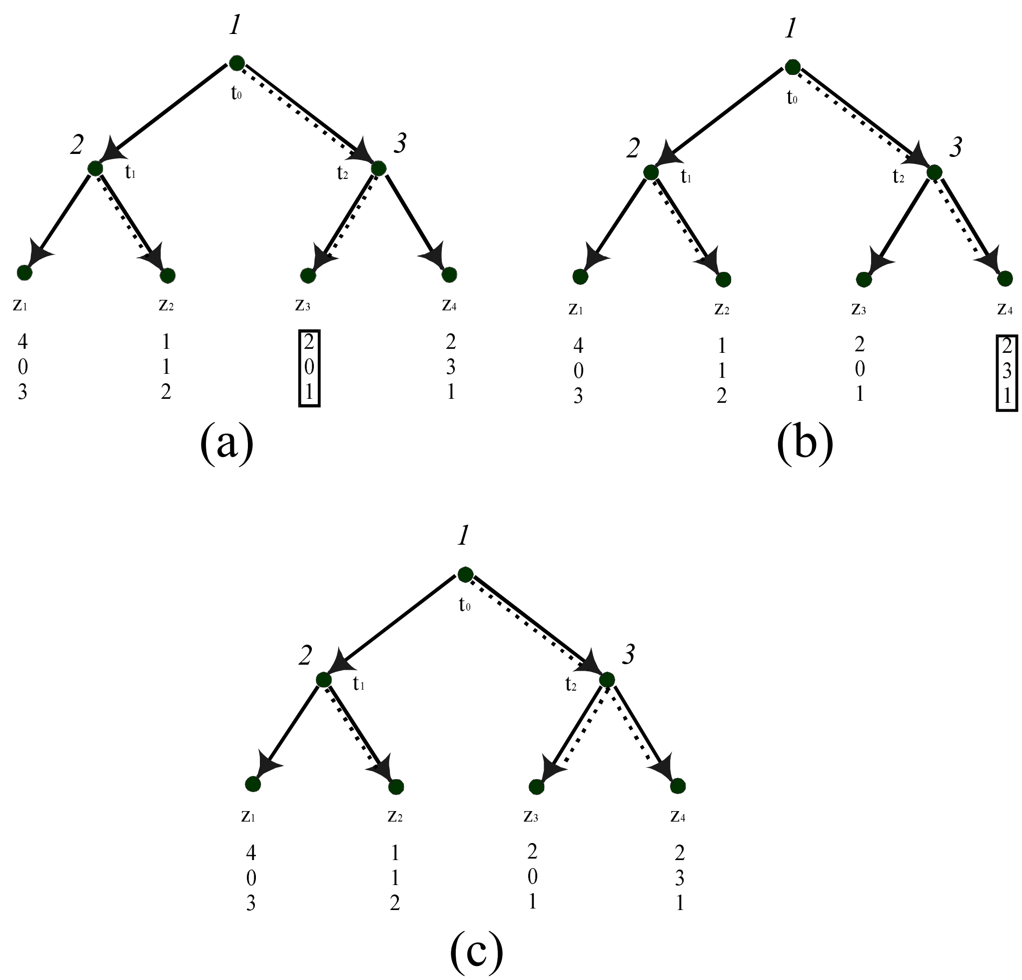

Figure 8

In order to generalize Proposition 5.3 to non-generic games we need the following lemma.

Lemma 5.4. *Let $\mathcal{G}$ be a perfect information game, $\mathcal{F}$ the corresponding BTA frame and $\prec_p$ a prediction for $\mathcal{F}$. Let $\mathcal{M}$ be a game model based on $\langle \mathcal{F}, \prec_p \rangle$ where axiom (IC) is valid. Then, $\forall t \in T, \forall q_1, q_2 \in \mathbb{Q}$,*

$$\textit{if} \quad \mathcal{M}, t \models F_p(u_{\iota(t)} = q_1) \wedge F_p(u_{\iota(t)} = q_2) \quad \textit{then} \quad q_1 = q_2.$$

Proof. Fix an arbitrary $t \in T$ and let $i = \iota(t)$. Suppose that

$$\mathcal{M}, t \models F_p(u_i = q_1) \wedge F_p(u_i = q_2). \tag{5.8}$$

Since (IC) is true at t, for every $r \in \mathbb{Q}$,

$$\mathcal{M}, t \models \Box_i \left(((u_i = r) \vee F_p(u_i = r)) \rightarrow (r \leq q_1) \right) \tag{5.9}$$

and

$$\mathcal{M}, t \models \Box_i \left(((u_i = r) \vee F_p(u_i = r)) \rightarrow (r \leq q_2) \right) \tag{5.10}$$

Furthermore, by (5.8), there exist $z_1, z_2 \in Z$ such that $t \prec_p z_1$, $t \prec_p z_2$, $u_i(z_1) = q_1$ and $u_i(z_2) = q_2$. For $j = 1, 2$ let t_j be the immediate successor of t on the $\prec$-path from t to z_j. Then

$$tR_it_1 \text{ and } tR_it_2. \tag{5.11}$$

Furthermore, by Property (P.7) of Definition 4.9, either $t_1 = z_1$ or $t_1 \prec_p z_1$ and either $t_2 = z_2$ or $t_2 \prec_p z_2$. Hence

$$\mathcal{M}, t_1 \models (u_i = q_1) \vee F_p(u_i = q_1) \tag{5.12}$$

and

$$\mathcal{M}, t_2 \models (u_i = q_2) \vee F_p(u_i = q_2). \tag{5.13}$$

It follows from (5.9), (5.11) and (5.12) that, for all $r \in \mathbb{Q}$, $\mathcal{M}, t_1 \models (r \leq q_1)$; in particular, $\mathcal{M}, t_1 \models (q_2 \leq q_1)$. Similarly, it follows from (5.10), (5.11) and (5.13) that $\mathcal{M}, t_2 \models (q_1 \leq q_2)$. Hence, by Definition 5.1, $q_1 = q_2$. ■

Definition 5.5. *Given a perfect information game* $\langle T, \rightarrowtail, N, \iota, \{u_i\}_{i \in N} \rangle$*, let* $\rightarrowtail_{eq}$ *be a subrelation of* $\rightarrowtail$ *and let* $\rightarrowtail_{BI}$ *be a backward-induction relation (cf. Definition 4.7). We say that* $\rightarrowtail_{eq}$ *is equivalent to* $\rightarrowtail_{BI}$ *if*

(1) $\rightarrowtail_{eq}$ *contains* $\rightarrowtail_{BI}$ *and*

(2) if $(t, t') \in \rightarrowtail_{eq}$ *and* $(t, t') \notin \rightarrowtail_{BI}$ *then, letting* z *be the unique terminal node* $\rightarrowtail_{BI}$*-reachable from* t *and* z' *be the unique terminal node* $\rightarrowtail_{BI}$*-reachable from* t'*,* $u_{\iota(t)}(z) = u_{\iota(t)}(z')$*.*

Thus a super-relation of a backward-induction relation is equivalent to it if, whenever an arrow from a node t to one of its immediate successors is added to the backward-induction relation, the player who moves at t is indifferent between the terminal node reachable from t by the backward-induction relation and any other terminal node that becomes reachable due to the addition.

Definition 5.6. *A prediction for a perfect information game is equivalent to a backward-induction prediction if it is the transitive closure of a subrelation of* $\rightarrowtail$ *which is equivalent to a backward-induction relation.*

The following proposition generalizes Proposition 5.3 to perfect information games that are not necessarily generic.

Proposition 5.7. *Let* $\mathcal{G}$ *be a perfect information game,* $\mathcal{F}$ *the associated BTA frame and* $\prec_p$ *a prediction for* $\mathcal{F}$*. Let* $\mathcal{M}$ *be any game model based on* $\langle \mathcal{F}, \prec_p \rangle$*. If axiom (IC) is valid in* $\mathcal{M}$ *then* $\prec_p$ *is equivalent to a backward induction prediction.*

We omit the proof of Proposition 5.7 since it follows directly from Lemma 5.4 with an argument similar to the one used in the proof of Proposition 5.3.

6. Conclusion

The logical foundations of game-theoretic solution concepts have so far been developed within the confines of epistemic logic. The purpose of this paper was to show that a different branch of modal logic, namely temporal logic, can offer new insights on the logic of solution concepts. We proposed to view the solution of a game as a complete prediction about future play. After having extended the branching time framework by adding agents and by defining the notion of prediction, we showed that perfect information games are a special case of extended branching time frames and that the backward-induction solution can be viewed as a prediction. We concluded by providing a characterization of backward induction in terms of the property of internal consistency of prediction.

The analysis in this paper was confined to perfect information games. In future work we hope to extend this approach to general games in extensive form.

References

[1] Battigalli, Pierpaolo and Giacomo Bonanno (1998), Recent results on belief, knowledge and the epistemic foundations of game theory, Working Paper, University of California Davis [forthcoming in *Research in Economics*].

[2] Belnap, Nuel and Michael Perloff (1988), Seeing to it that: a canonical form for agentives, *Theoria*, 54: 175-199.

[3] van Benthem, Johan, *The logic of time*, Kluwer Academic Publishers, 1991.

[4] Bonanno, Giacomo (1998), The logic of prediction, Working Paper, University of California Davis.

[5] Burgess, John, Basic tense logic, in: D. Gabbay and F. Guenthner (eds.), *Handbook of philosophical logic*, Vol. II, 1984, 89-133, D. Reidel Publishing Company.

[6] Chellas, Brian (1992), Time and modality in the logic of agency, *Studia Logica*, 51: 485-517.

[7] Dekel, Eddie and Faruk Gul (1997). Rationality and knowledge in game theory. in *Advances in Economics and Econometrics* (D. Kreps and K. Wallis, Eds.). Cambridge UK: Cambridge University Press.

[8] van Fraassen, Bas, Belief and the will, *Journal of Philosophy*, 1984, 81: 235-256.

[9] Fudenberg, Drew and Jean Tirole (1991). *Game Theory.* Cambridge MA: MIT Press.

[10] Goldblatt, Robert, *Logics of time and computation*, CSLI Lecture Notes No. 7, 1992.

[11] Greenberg, Joseph (1990), *The theory of social situations*, Cambridge University Press.

[12] Halpin, John, Indeterminism, indeterminateness, and tense logic, *Journal of Philosophical Logic*, 1988, 17:207-219.

[13] Horty, John (1996), Agency and obligation, *Synthese*, 108: 269-307.

[14] Horty, John and Nuel Belnap (1995), The deliberative stit: a study of action, omission, ability and obligation, *Journal of Philosophical Logic*, 24: 583-644.

[15] von Neumann, John and Oscar Morgenstern (1947), *Theory of games and economic behavior*, Princeton University Press.

[16] Øhrstrøm, Peter and Per F. V. Hasle, *Temporal logic*, Kluwer, 1995.

[17] Osborne, Martin and Ariel Rubinstein (1994): *A Course in Game Theory*. Cambridge MA: MIT Press.

[18] Zermelo, E. (1913), Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels, in: E.W. Hobson and A.E.H. Love, Eds., *Preceedings of the Fifth International Congress of Mathematicians*, Vol. II, Cambridge University Press, pp. 501-504.