

Published in:

Games and Economic Behavior, 36 (1), July 2001, pp. 57-73.

BRANCHING TIME, PERFECT INFORMATION GAMES AND BACKWARD INDUCTION

Giacomo Bonanno^a

Department of Economics,

University of California,

Davis, CA 95616-8578 USA

e-mail: gfbonanno@ucdavis.edu

~~www.econ.ucdavis.edu/~bonanno~~ <http://faculty.econ.ucdavis.edu/faculty/bonanno/>

November 1998, revised November 1999.

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

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. The branching time framework is extended by adding agents and by defining the notion of prediction. A syntactic characterization of backward induction in terms of the property of internal consistency of prediction is given.

^aThis paper was presented at the Third Conference on Logic and the Foundations of Game and Decision Theory (LOFT3), Turin, December 1988. I am grateful to two anonymous referees and to the conference participants for useful and constructive comments.

1. Introduction

The logical foundations of game theory have been the object of a recent and growing literature. 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.¹

Here we turn to a different branch of modal logic, namely temporal logic², and propose to view the solution of a game as a prediction about future play. We focus on extensive games with perfect information, which are modeled in a natural way within the framework of branching time logic. 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 and in Section 2 we discuss what properties one should attribute to predictions in general. In Section 3 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. Section 4 contains the main result, namely a syntactic characterization of backward induction in terms of internal consistency of prediction, in the following sense: if at any node it is predictable (that is, possible according to the prediction) that player i 's payoff will be q then player i cannot induce a position where his payoff is greater than q , or it is predictable that it will be greater than q .

This notion of internal consistency 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 theory of social situations. 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.³

¹Surveys of this literature (and an extensive list of references) can be found in Battigalli and Bonanno (1999) and Dekel and Gul (1997).

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

³The 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

2. Agents and predictions in branching time

Definition 2.1. A branching-time frame with agents (BTA frame for short) is a tuple $\langle T; \dot{A}; N; \{R_i\}_{i \in N} \rangle$ where

² T is a (possibly infinite) set of nodes.⁴

² \dot{A} is a binary relation on T (the precedence relation, representing the ordering of time) satisfying the following properties:

(P.0) antisymmetry: if $t_1 \dot{A} t_2$ then $t_2 \not\dot{S} t_1$:

(P.1) transitivity: if $t_1 \dot{A} t_2$ and $t_2 \dot{A} t_3$ then $t_1 \dot{A} t_3$:

(P.2) backward linearity: if $t_1 \dot{A} t_3$ and $t_2 \dot{A} t_3$ then either $t_1 = t_2$ or $t_1 \dot{A} t_2$ or $t_2 \dot{A} t_1$:

² $N = \{1; \dots; n\}$ is a finite set of agents.

² for every $i \in N$, R_i is a binary relation on T satisfying the following property:

(P.3) R_i subrelation of \dot{A} : if $t_1 R_i t_2$ then $t_1 \dot{A} t_2$.

Properties (P.0)-(P.2) constitute the definition of branching time in temporal logic.⁵ In particular, (P.2) expresses the notion that, while a given node may have different possible futures, its past is unique, that is, (P.2) rules out the possibility that two different past histories lead to the same node.

The interpretation of $t_1 R_i t_2$ is that at node t_1 agent i has available an action 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 , $R_i(t) \stackrel{\text{def}}{=} \{t' \in T : t R_i t'\}$ is empty. In such a case agent i does not have any actions available at node t .⁶

game-theoretic issues. Furthermore, while we make use of standard (Kripkean) temporal 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 point and a branch or history through it; the future operator then refers to points in this branch only and, therefore, the resulting logic is that of linear time. A further operator is then added to capture the notions of historical necessity and contingency.

⁴In the philosophical literature the elements of T are usually called moments or points in time. Since our focus is on games, we prefer to call them nodes.

⁵See, for example, Burgess (1984) and Åhrström and Hasle (1995).

⁶A natural requirement might be that different actions of the same agent be either simultaneous or determining different future histories, in the sense that if $t R_i t^0$ and $t R_i t^0$ then $t^0 \dot{S} t^0$ and $t^0 \dot{S} t^0$. Note that simultaneous actions of different agents are not ruled out, that is, it

Definition 2.2. Given a BTA frame, a prediction is a binary relation \hat{A}_p on T satisfying the following properties:

- (P.4) \hat{A}_p subrelation of \hat{A} : if $t_1 \hat{A}_p t_2$ then $t_1 \hat{A} t_2$:
- (P.5) transitivity: if $t_1 \hat{A}_p t_2$ and $t_2 \hat{A}_p t_3$ then $t_1 \hat{A}_p t_3$:
- (P.6) \hat{A}_p is serial when \hat{A} is:⁷ if $t \hat{A} t_1$ for some t_1 , then $t \hat{A}_p t_2$ for some t_2 :
- (P.7) time consistency: if $t_1 \hat{A} t_2$; $t_2 \hat{A} t_3$ and $t_1 \hat{A}_p t_3$ then $t_1 \hat{A}_p t_2$ and $t_2 \hat{A}_p t_3$:

(P.4) expresses the notion that predicting the future consists in selecting a subset of the conceivable future nodes (those that are believed to be most plausible). Note that it is not assumed that the predictable future of a given node be a unique history following that node (that is, we do not require that if $t \hat{A}_p t^0$ and $t \hat{A}_p t^{00}$ then either $t^0 = t^{00}$ or $t^0 \hat{A} t^{00}$ or $t^{00} \hat{A} t^0$). Furthermore, there is no requirement that the predictable future of a given node be a proper subset of its conceivable future, that is, vague predictions are allowed. For example, suppose that $T = f(t_1; t_2; t_3; t_4)g$ and $\hat{A} = f(t_1; t_2); (t_1; t_3); (t_1; t_4)g$. Suppose also that t_2 is a state where it is sunny, t_3 is a state where it rains and t_4 is a state where it snows. Then $\hat{A}_p = \hat{A}$ corresponds to the trivial prediction "tomorrow either it will be sunny or it will rain or it will snow", while $\hat{A}_p = f(t_1; t_2); (t_1; t_3)g$ corresponds to the somewhat vague prediction "tomorrow either it will be sunny or it will rain, but it will not snow" and $\hat{A}_p = f(t_1; t_2)g$ corresponds to the sharp prediction "tomorrow it will be sunny".

The interpretation of \hat{A}_p in terms of prediction (i.e. belief about the future) makes (P.5) (transitivity of \hat{A}_p) 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 Fraassen, 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 t has a \hat{A} -successor) then there must be a predictable future of t (that is, t must have a \hat{A}_p -successor). This is not really a restriction, since the trivial prediction that every conceivable future is plausible (that is, $t_1 \hat{A}_p t_2$ i[®] $t_1 \hat{A} t_2$) is not ruled out.

is possible that, for some t and some i and j with $i \notin j$, both $R_i(t)$ 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.

⁷In the modal logic literature seriality is usually defined globally. We define it as a local property, since in finite games there are decision nodes, which have successors, as well as terminal nodes, that have no successors. Note that, therefore, the modal operators do not satisfy the consistency axiom.

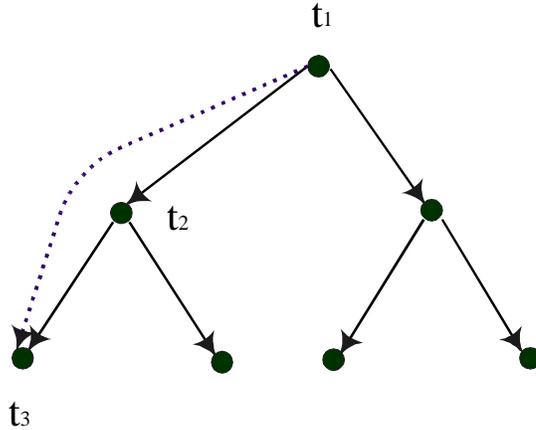


Figure 1

Property (P.7) says the following. Suppose that at node t_1 a conceivable future development is represented by the path $t_1 t_2 t_3$ (that is, $t_1 \hat{A} t_2$ and $t_2 \hat{A} t_3$): this is shown in Figure 1, where a continuous arrow from t to t^0 denotes that $t \hat{A} t^0$ (and the arrows due to transitivity are omitted). Suppose also that t_3 lies in the predictable future of t_1 (that is, $t_1 \hat{A}_p t_3$): this is shown in Figure 1 by a dotted arrow from t_1 to t_3 . Then (P.7) imposes the following requirements:

- (a) since reaching t_3 from t_1 requires going through t_2 , t_2 should lie in the predictable future of t_1 (that is, $t_1 \hat{A}_p t_2$), and
- (b) since reaching t_2 from t_1 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 predictable future of t_2 ($t_2 \hat{A}_p t_3$).

In view of the branching structure of time (there is at most a unique path between any two nodes) (P7) seems a very natural consistency requirement.

Example 2.3. The following is a BTA frame: $T = \{t_1, t_2, \dots, t_7, t_8\}$, $N = \{f, g\}$, $\hat{A} = \{f(t_1; t_2); (t_1; t_3); (t_1; t_4); (t_1; t_5); (t_1; t_6); (t_1; t_7); (t_1; t_8); (t_2; t_4); (t_2; t_5); (t_3; t_6); (t_3; t_7); (t_3; t_8)\}$, $R_1 = \{f(t_1; t_2); (t_1; t_3)\}$, $R_2 = \{f(t_2; t_4); (t_2; t_5); (t_3; t_6); (t_3; t_7); (t_3; t_8)\}$. This frame is shown in Figure 2 where, as before, an arrow from t to t^0 indicates that $t \hat{A} t^0$ and all the arrows due to transitivity are omitted (thus the continuous arrows represent the Hasse diagram of $\langle T; \hat{A} \rangle$); furthermore the label i is assigned to the arrow from t to t^0 if and only if $(t; t^0) \in R_i$. The following is a prediction

according to Definition 2.2: $\hat{A}_p = f(t_1; t_3); (t_1; t_6); (t_1; t_7); (t_3; t_6); (t_3; t_7); (t_2; t_5)g$. This is represented in Figure 2 by a dotted line next to an arrow that belongs to both \hat{A} and \hat{A}_p , omitting dotted lines that can be obtained by transitivity (thus the dotted lines alone represent the Hasse diagram of $\langle T; \hat{A}_p \rangle$).

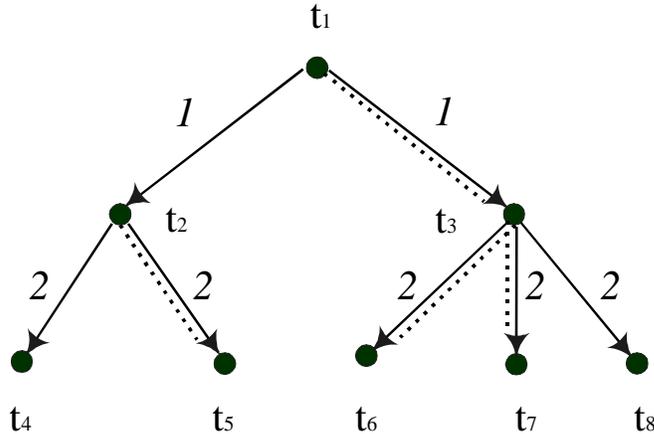


Figure 2

Definition 2.4. An augmented BTA frame is a BTA frame together with a prediction.

Every $t \in T$ should be thought of as a complete description of the world, and sets of nodes represent propositions. In order to establish this interpretation we need to introduce a formal language and the notion of a model based on a frame. We consider a propositional language with the following modal operators:

² Tense and prediction operators⁸: G , G_p , H and H_p . The intended interpretation is as follows:

- $G\hat{A}$: "\it is going to be the case in every conceivable future that \hat{A} "
- $G_p\hat{A}$: "\it is going to be the case in every predictable future that \hat{A} "
- $H\hat{A}$: "\it has always been the case that \hat{A} "
- $H_p\hat{A}$: "\it has always been the case at every past node at which the current node was predicted that \hat{A} "

⁸The names of the operators are the standard ones in temporal logic. G stands for "going to be" and H for "has been".

² Action operators: α_i (for every $i \in \mathbb{N}$), whose intended interpretation is:

$\alpha_i \hat{A}$: "no matter what action agent i takes, it will be the case that \hat{A} ".

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 $_$ (from which the other connectives \wedge , \vee and $\$$ are defined as usual) and the above modal operators.⁹

Given an augmented BTA frame one obtains a model M based on it by adding a function $V : S \rightarrow 2^T$ (where 2^T denotes the set of subsets of T) that associates with every sentence letter p the set of nodes at which p is true. For non-modal formulae truth at a node in a model is defined as usual.¹⁰ Validation for modal formulae is as follows:

$M; t \models G\hat{A} \text{ i}^\circ M; t^0 \models \hat{A}$ for all t^0 such that $t \hat{A} t^0$:

$M; t \models H\hat{A} \text{ i}^\circ M; t^{00} \models \hat{A}$ for all t^{00} such that $t^{00} \hat{A} t$:

$M; t \models G_p \hat{A} \text{ i}^\circ M; t^0 \models \hat{A}$ for all t^0 such that $t \hat{A}_p t^0$:

$M; t \models H_p \hat{A} \text{ i}^\circ M; t^{00} \models \hat{A}$ for all t^{00} such that $t^{00} \hat{A}_p t$:

$M; t \models \alpha_i \hat{A} \text{ i}^\circ M; t^0 \models \hat{A}$ for all t^0 such that $t R_i t^0$.

Thus $G\hat{A}$ ($G_p \hat{A}$) is true at node t if \hat{A} is true at every \hat{A} -successor (\hat{A}_p -successor) of t . Similarly for H and H_p : Let $F_p \hat{A} \stackrel{\text{def}}{=} _ G_p _ \hat{A}$. Then its intended interpretation is:

$F_p \hat{A}$: "at some predictable future node \hat{A} ".¹¹

A formula \hat{A} is valid in model M if $M; t \models \hat{A}$ for all $t \in T$; it is valid on a frame if it is valid in every model based on it.¹²

⁹The set \circledast 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 α_i : (i) for every $p \in S$, $(p) \in \circledast$, (ii) if $\hat{A}, \hat{B} \in \circledast$ then all of the following belong to \circledast : $(\neg \hat{A})$, $(\hat{A} _ \hat{B})$, $G\hat{A}$, $H\hat{A}$, $G_p \hat{A}$, $H_p \hat{A}$ and $\alpha_i \hat{A}$:

¹⁰ $M; t \models \hat{A}$ denotes that \hat{A} is true at node t in model M and $M; t \not\models \hat{A}$ denotes that \hat{A} is false at t . For a sentence letter p , $M; t \models p \text{ i}^\circ t \in V(p)$; furthermore, $M; t \models _ \hat{A} \text{ i}^\circ M; t \not\models \hat{A}$ and $M; t \models (\hat{A} _ \hat{B}) \text{ i}^\circ$ either $M; t \models \hat{A}$ or $M; t \models \hat{B}$. It follows that $M; t \models (\hat{A} \wedge \hat{B}) \text{ i}^\circ M; t \models \hat{A}$ and $M; t \models \hat{A}$, and $M; t \models (\hat{A} \vee \hat{B}) \text{ i}^\circ M; t \models \hat{A}$ implies $M; t \models \hat{A}$:

¹¹Thus $M; t \models F_p \hat{A} \text{ i}^\circ M; t^0 \models \hat{A}$ for some t^0 with $t \hat{A}_p t^0$: F , the dual of G , P , the dual of H , and P_p , the dual of H_p , are defined and interpreted similarly.

¹²A sound and complete axiomatization of augmented BTA frames is given in Bonanno (1998).

3. 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 4 we provide a syntactic characterization of backward induction.

A rooted tree is a pair $\langle T; \preceq \rangle$ where T is a set of nodes and \preceq is a binary relation on T (if $t \preceq t^0$ we say that t immediately precedes t^0 or that t^0 immediately succeeds t) satisfying the following properties:

1. there is a unique node t_0 (the root) with no immediate predecessors;
2. for every node $t \in T \setminus \{t_0\}$ there is a unique path from t_0 to t , that is, there is a unique sequence $\langle x_1; \dots; x_m \rangle$ in T with $x_1 = t_0$, $x_m = t$, and, for every $j = 1; \dots; m - 1$, $x_j \preceq x_{j+1}$;

Given a rooted tree $\langle T; \preceq \rangle$, 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 3.1. A finite extensive form with perfect information is a tuple $\langle T; \preceq; N; \sigma \rangle$ where $\langle T; \preceq \rangle$ is a finite rooted tree, $N = \{1; \dots; n\}$ is a set of players and $\sigma : T \setminus Z \rightarrow N$ is a function that associates with every non-terminal or decision node the player who moves at that node. If $i = \sigma(t)$ and $t \preceq t^0$ we say that the pair $(t; t^0)$ is a choice of player i at node t . Given an extensive form, one obtains a perfect information game by adding, for every player $i \in N$, a payoff or utility function $u_i : Z \rightarrow \mathbb{Q}$ (where Z is the set of terminal nodes and \mathbb{Q} is the set of rational numbers).

Figure 3a shows a perfect information game with three players. There is an arrow from t to t^0 if and only if $t \preceq t^0$ and the vector $(x_1; x_2; x_3)$ written next to a terminal node z is the payoff vector $(u_1(z); u_2(z); u_3(z))$. For every decision node t , the corresponding player $\sigma(t)$ is written next to it.

Lemma 3.2. A finite extensive form with perfect information is a special case of a BTA frame (cf. Definition 2.1).

Proof. Let \hat{A} be the transitive closure of $\frac{1}{2}$, that is, $t \hat{A} t^0$ if and only if there is a $\frac{1}{2}$ -path from t to t^0 . It is straightforward to show that \hat{A} satisfies properties (P.0)-(P.3) of Definition 2.1. Furthermore, if t is a decision node let $tR_i t^0$ if and only if $i = \mathcal{I}(t)$ and $t \frac{1}{2} t^0$, while for every $j \notin \mathcal{I}(t)$, $R_j(t) = \emptyset$. If z is a terminal node, then $R_j(z) = \{z\}$ for all $j \in N$. It is obvious that property (P.3) is satisfied. ■

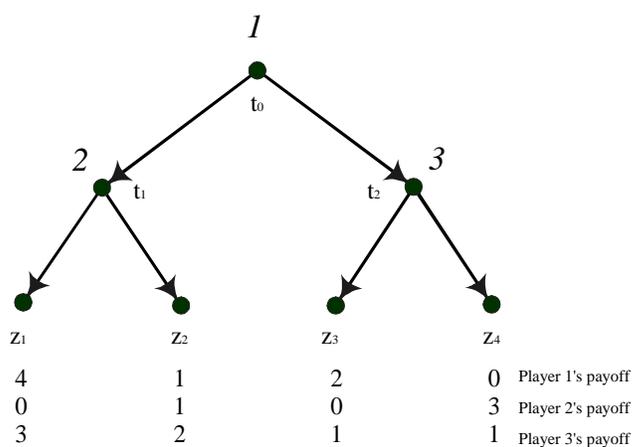


Figure 3a

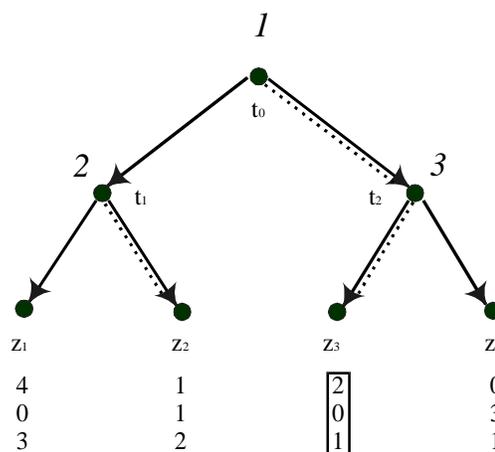


Figure 3b

A well-known procedure for solving a perfect information game is the backward induction algorithm (see, for example, Fudenberg and Tirole, 1991). The algorithm starts at the end of the game and proceeds backwards towards the root:

1. Start at a decision node t whose immediate successors are only terminal nodes (e.g. node t_1 in Figure 3a) and select one choice that maximizes the utility of player $\mathcal{I}(t)$ (in the example of Figure 3a, 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 3b shows one possible outcome of the backward induction algorithm for the game of Figure 3a. The choices selected by the algorithm are shown as

dotted lines next to the corresponding arrows. Note that the backward induction algorithm may yield more than one solution. Multiplicity can arise if there are players who have more than one utility-maximizing choice. For example, in the game of Figure 3a at node t_2 both choices are optimal for Player 3. The selection of choice $(t_2; z_3)$ leads to the solution shown in Figure 3b, while the selection of choice $(t_2; z_4)$ leads to a different solution, namely $f(t_0; t_1); (t_1; z_2); (t_2; z_4)g$.

Definition 3.3. A perfect information game is generic if no player is indifferent between any two terminal nodes, that is, if $\exists i \in N; \exists z, z^0 \in Z$ if $u_i(z) = u_i(z^0)$ then $z = z^0$:

Remark 1. In a generic game the backward induction algorithm yields a unique solution.

We now show that a backward-induction solution is a prediction in the sense of Definition 2.2. To do this we need a more precise definition of backward-induction, which, together with the proof of the following lemma, is given in the Appendix.

Lemma 3.4. Fix a perfect information game. Let \mathcal{I}_{BI} be a backward induction relation for it (cf. Definition A.3 in the Appendix) and \hat{A}_p its transitive closure. Then \hat{A}_p is a prediction in the sense of Definition 2.2.

Definition 3.5. Given a perfect information game, a relation \hat{A}_p on T is called a backward induction prediction if it is the transitive closure of a backward-induction relation for that game.

For every finite perfect information game there is at least one backward-induction prediction, although, as noted above, there may be more than one. However, in generic games (cf. Definition 3.3) there is a unique backward-induction prediction.

Remark 2. It follows from Definitions A.3 (in the Appendix) and 3.5 that, if \hat{A}_p is a backward-induction prediction, then

- (a) if $t \hat{A}_p t^0$ and $t \hat{A}_p t^{00}$ then either $t^0 = t^{00}$ or $t^0 \hat{A}_p t^{00}$ or $t^{00} \hat{A}_p t^0$,
- (b) for every decision node $t \in T \cap Z$, there is a unique $z \in Z$ such that $t \hat{A}_p z$.

4. 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 3.2 showed 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 2) all we need to do is include in the set of sentences (or atomic propositions) sentences of the form $(u_i = q)$ with $i \in \mathbb{N}$ 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 by means 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 2 from a set of sentences S that includes atomic propositions of the form $(u_i = q)$ and $(q_1 \leq q_2)$.

Definition 4.1. Fix a perfect information game and let F be the corresponding BTA frame. A game model is a model based on F (cf. Section 2) obtained in a game language by adding to F a valuation $V : S \rightarrow \{0, 1\}$ satisfying the following properties:

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

$$V(p) = 1 \text{ if } q_1 \leq q_2 \text{ and } V(p) = 0 \text{ otherwise}$$

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

$$V(p) = 1 \text{ if } \exists z \in Z : u_i(z) = q.$$

Thus if M is a game model then, $\forall t \in T, M; t \models (q_1 \leq q_2)$ if q_1 is less than or equal to q_2 and $M; t \models (q_1 \leq q_2)$ otherwise; furthermore, $M; t \models (u_i = q)$ if t is a terminal node with $u_i(t) = q$ and $M; t \models (u_i = q)$ if t is either a decision node or a terminal node with $u_i(t) \neq q$.¹³ The valuation of the other atomic formulae and of the non-atomic formulae is as explained in Section 2.

Consider the following axiom scheme:

$$F_p(u_i = q) \rightarrow \exists_i ((u_i = r) \rightarrow F_p(u_i = r)) \rightarrow (r \leq q) \quad (IC_0)$$

¹³Thus if t is a node whose successors include non-terminal nodes, then the formula $G_p(u_i = q)$ is necessarily false at t , for every player i and for every number q .

(IC_o) says that if it is predictable (i.e. possible according to the prediction) that player *i*'s payoff will be *q* then, no matter what action he takes, if his payoff is *r*, or it is predictable that it will be *r*, then *r* is not greater than *q*. Thus (IC_o) can be viewed as expressing a notion of internal consistency¹⁴ of prediction or recommendation, 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.

The following proposition shows that axiom (IC_o) characterizes the notion of backward induction in generic games.

Proposition 4.2. Let *G* be a generic perfect information game, *F* the associated BTA frame and \hat{A}_p a prediction for *F* (cf. Definition 2.2). Let *M* be any game model based on $\langle F; \hat{A}_p \rangle$ (cf. Definition 4.1). Then the following are equivalent:

- (a) axiom (IC_o) is valid in *M*;
- (b) \hat{A}_p is the backward induction prediction (cf. Definition 3.5).¹⁵

Proof. First we show that if \hat{A}_p is the backward induction prediction then every instance of (IC_o) is true at every $t \in T$. If *t* is a terminal node, then $\forall t' \in T : t \hat{A}_p t' \Rightarrow g = ;$ and therefore $M; t \models F_p(u_i = q)$ for all $i \in N$ and $q \in Q$. Thus (IC_o) is true at *t*. If *t* be a decision node and $i \notin \mathcal{I}(t)$ then $R_i(t) = ;$ and therefore $M; t \models \neg \hat{A}_i$ for every formula \hat{A}_i ; hence (IC_o) is true at *t*. Thus we only need to consider the case where *t* is a decision node and $i \in \mathcal{I}(t)$. Suppose that (IC_o) is false at *t*. Then there are numbers $q, r \in Q$ such that

$$M; t \models F_p(u_i = q) \quad (4.1)$$

and $M; t \not\models \neg ((u_i = r) \wedge F_p(u_i = r)) \vee (r > q)$, that is,

$$\exists t' \in T : t R_i t' \text{ and } M; t' \models ((u_i = r) \wedge F_p(u_i = r)) \wedge (r > q): \quad (4.2)$$

¹⁴Hence the name IC. The subscript 'o' stands for 'optimistic' as will be explained later. As noted 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.

¹⁵Recall that in generic games there is a unique backward induction prediction. Note that the statements "(IC_o) is valid in a game model based on $\langle F; \hat{A}_p \rangle$ " and "(IC_o) is valid in every game model based on $\langle F; \hat{A}_p \rangle$ " are equivalent, since (IC_o) is made up only of atomic propositions of the form $(u_i = q)$ and $(r > q)$ and the valuations of different models coincide on this class of atomic propositions.

By Remark 2 there is a unique $z \in Z$ such that $t \hat{A}_p z$. By (4.1) $u_i(z) = q$: Let t^{00} be the unique immediate successor of t on the \hat{A}_p -path from t to z . By definition of R_i (cf. Lemma 3.2), the t^0 of (4.2) is also an immediate successor of t . Let z^0 be the unique terminal node such that $t^0 \hat{A}_p z^0$ (if t^0 is a terminal node, let $z^0 = t^0$). Then, by (4.2), $u_i(z^0) = r$ and $r > q$. Thus

$$u_i(z^0) > u_i(z): \quad (4.3)$$

By Definition A.2 (in the Appendix), $u_i^{(t^{00})}(t^{00}) = u_i(z)$, $u_i^{(t^0)}(t^0) = u_i(z^0)$ and $u_i^{(t^{00})}(t^{00}) \succ u_i^{(t^0)}(t^0)$, contradicting (4.3).

Next we show that if (IC_0) is valid in M then \hat{A}_p is the backward induction prediction. First of all, by property (P.4) of Definition 2.2 (\hat{A}_p subrelation of \hat{A}), all predictions coincide when restricted to the set of level 0 (or terminal) nodes (they are equal to the empty set). Thus, in particular, \hat{A}_p restricted to T_0 coincides with the backward-induction prediction restricted to T_0 : Now we show that \hat{A}_p restricted to T_1 (the set of level 1 nodes: cf. Definition A.1 in the Appendix) coincides with the restriction of the backward-induction prediction to T_1 . Let $\hat{t} \in T_1$ and let $\hat{Z} = \{z \in Z : \hat{t} \hat{A}_p z\}$. By Properties (P.4) and (P.6) of Definition 2.2 (\hat{A}_p subrelation of \hat{A} , and \hat{A}_p serial if \hat{A} is serial), $\hat{Z} \setminus \{\hat{t}\} \subseteq T : \hat{t} \hat{A}_p t \in \hat{Z}$. Fix an arbitrary $\hat{z} \in \hat{Z} \setminus \{\hat{t}\} \subseteq T : \hat{t} \hat{A}_p \hat{z}$. Then, letting $i = \#(\hat{t})$ and $q = u_i(\hat{z})$,

$$M; \hat{t} j = F_p(u_i = q): \quad (4.4)$$

Furthermore, it must be the case that

$$q \prec u_i(z); \quad \exists z \in \hat{Z}: \quad (4.5)$$

In fact, suppose that, for some $z^0 \in \hat{Z}$, $u_i(z^0) = r > q$: Then $M; z^0 j = (u_i = r) \wedge : (r \cdot q)$. Since $\hat{t} R_i z^0$, $M; \hat{t} j = : \alpha_i ((u_i = r) _ F_p(u_i = r) _ (r \cdot q))$. Thus, by (4.4) (IC_0) would be false at \hat{t} , contrary to the hypothesis that (IC_0) is valid in M . Since the game is generic, if $z \in \hat{Z}$ is such that $z \notin \hat{z}$ then, by (4.5), $u_i(z) < q$; it follows that $\hat{t} \hat{A}_p t \in \hat{z}$. Thus, restricted to T_1 , \hat{A}_p coincides with the backward induction prediction. Next we show that if \hat{A}_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 2.2, $\exists t^{00} \in T$ such that $\hat{t} \hat{A}_p t^{00}$. If t^{00} is not a terminal node, let t^0 be the unique immediate successor of \hat{t} on the \hat{A} -path from \hat{t} to t^{00} . Then, by Property (P.7) of Definition 2.2, $\hat{t} \hat{A}_p t^0$. Clearly, $\#(t^0) = k$; hence, by our supposition that

\hat{A}_p coincides with the backward-induction prediction when restricted to $\bigcup_{j=0}^k T_k$, there is a unique $z^0 \in Z$ such that $t^0 \hat{A}_p z^0$. Let $i = \mathfrak{I}(\hat{t})$ and $q = u_i(z^0)$: Then

$$M; \hat{t} \vDash F_p(u_i = q); \quad (4.6)$$

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

$$u_i(z_{t^0}) \geq u_i(z_t); \quad \forall t \in T : \hat{t} \not\leq t \quad (4.7)$$

Suppose not. Then there exists a $t \in T$ such that $\hat{t} \not\leq t$ and $u_i(z_t) = r > q = u_i(z_{t^0})$. Two cases are possible: (1) $t \in Z$, or (2) $t \notin Z$. In case (1), $M; \hat{t} \vDash (u_i = r) \wedge : (r \cdot q)$; while in case (2) $M; \hat{t} \vDash F_p(u_i = r) \wedge : (r \cdot q)$. Thus in either case $M; \hat{t} \vDash : \mathfrak{A}_i(((u_i = r) \wedge F_p(u_i = r)) \rightarrow (r \cdot q))$. Hence, by (4.6), (IC_0) is false at \hat{t} , contradicting the hypothesis that (IC_0) is valid in M . Since the game is generic, it follows from (4.7) that $\exists z \in Z : \hat{t} \hat{A}_p z \wedge fz_{t^0}g = fz_tg$ and, therefore, if t is an immediate successor of \hat{t} and $\hat{t} \hat{A}_p t$ then $t = t^0$. Thus the restriction of \hat{A}_p to T_{k+1} coincides with the restriction to T_{k+1} of the backward induction prediction. ■

In non-generic games it is still true that if \hat{A}_p is a backward induction prediction then (IC_0) is valid, since a backward induction prediction is such that the predictable future of any node t is always a unique path (cf. Remark 2). However, in non-generic games it is possible to satisfy (IC_0) with a relation that includes more than one path out of some nodes. This is illustrated in Figure 4, 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_0) in every model based on this game.

When there is a multiplicity of predictable paths, the issue arises of how to compare sets of predictable payoffs. Greenberg (1990) proposes two notions of internal consistency (or stability). According to the notion of optimistic internal consistency, a player will reject a recommendation x if she can induce a position where, among the recommendations made, there is one which she prefers to x . That is, the player looks at the best possible outcome among those recommended at the position to which she is contemplating a deviation. On the other hand, according to the notion of conservative internal consistency, a player will reject a recommendation x if she can induce a position where every recommendation

is better than x . That is, the player looks at the worst possible outcome among those recommended at the position to which she is contemplating a deviation.

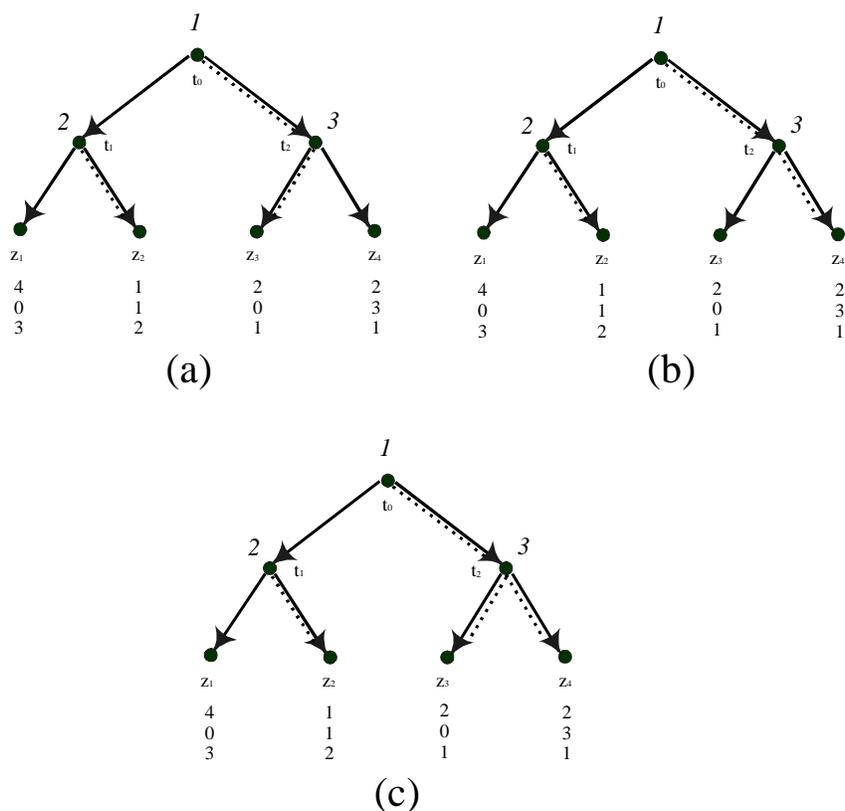


Figure 4

The next proposition shows that axiom (IC_o) captures the notion of optimistic internal consistency. In order to characterize the alternative notion of conservative internal consistency, we first introduce a short-hand notation:

$$G_p^s(u_i = q) \stackrel{\text{def}}{=} F_p(u_i = q) \wedge (F_p(u_i = s) \rightarrow (q \geq s))$$

Thus $G_p^s(u_i = q)$ says that, according to the prediction, player i 's payoff can be exactly q and will be at least q , that is, his minimum payoff will be q . Thus, for every player i and node t , $G_p^s(u_i = q)$ is true at t if and only if $q = \min\{u_i(z) : z \in Z \text{ and } t \in \hat{A}_p(z)\}$. Consider now the following axiom:

$$G_{\hat{p}}(u_i = q) \leq \alpha_i(((u_i = r) \rightarrow G_{\hat{p}}(u_i = r)) \leq (r \cdot q)) \quad (IC_c)$$

(IC_c) says that if, according to the prediction, player i's payo[®] will be at least q, then, no matter what action player i takes, if her payo[®] is r, or is predicted to be at least r, then r is not greater than q. The following proposition shows that (IC_c) captures the notion of conservative internal consistency. Furthermore, in generic games both (IC_o) and (IC_c) characterize backward induction.

Proposition 4.3. Let G be an arbitrary perfect information game, F the associated BTA frame and \hat{A}_p a prediction for F. Let M be any game model based on $\langle F; \hat{A}_p \rangle$. Then:

(a) (IC_o) is valid in M i[®], for every node t and every immediate successor t⁰ of t, $\max_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$ and $t^0 \in \hat{A}_p$ \cdot $\min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$ if t⁰ is a decision node and $u_{\eta(t)}(t^0) \cdot \min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$ if t⁰ is a terminal node;

(b) (IC_c) is valid in M i[®], for every node t and every immediate successor t⁰ of t, $\min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$ \cdot $\min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$ if t⁰ is a decision node and $u_{\eta(t)}(t^0) \cdot \min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$ if t⁰ is a terminal node;

(c) if the game is generic then the following are equivalent: (c.1) (IC_o) is valid in M, (c.2) (IC_c) is valid in M and (c.3) \hat{A}_p is the backward-induction prediction.

Proof. We only sketch the proof, since it follows directly from the definitions and the arguments used in the proof of Proposition 4.2. We shall concentrate on the case where t⁰ is a decision node. Preliminaries: let $q = \min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$, $m = \min_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$, and $M = \max_{u_{\eta(t)}(z) : z \in Z} u_{\eta(t)}(z) > \hat{A}_p(z)$. Thus $M; t \models F_p(u_{\eta(t)} = q)$, $M; t^0 \models F_p(u_{\eta(t)} = m) \wedge F_p(u_{\eta(t)} = M)$, $M; t \models G_{\hat{p}}(u_{\eta(t)} = q)$ and $M; t^0 \models G_{\hat{p}}(u_{\eta(t)} = m)$.

For part (a), if $M > q$ then, since $(t; t^0) \in R_{\eta(t)}$, (IC_o) is violated at t. For the converse, first note that, by the definition of prediction, the sets $\{z \in Z : t \in \hat{A}_p(z)\}$ and $\{z \in Z : t^0 \in \hat{A}_p(z)\}$ are non-empty; thus the numbers q and M are well-defined. Furthermore, (IC_o) can be violated at t only if, for some immediate successor t⁰ of t, and for some $v > q$, $M; t^0 \models F_p(u_{\eta(t)} = v)$. But, since $v > M$, this would require $M > q$.

For part (b), if $m > q$ then (IC_c) is satisfied at t. Conversely, if (IC_c) is satisfied at t, given that the sets $\{z \in Z : t \in \hat{A}_p(z)\}$ and $\{z \in Z : t^0 \in \hat{A}_p(z)\}$ are non-empty and thus the numbers q and m are well defined, it must be that $m > q$.

The equivalence of (c.1) and (c.3) was established in Proposition 4.2. The equivalence of (c.2) and (c.3) follows from an argument similar to the one used in the proof of Proposition 4.2. ■

To see the difference between (IC_o) and (IC_c) , consider the game of Figure 4(c) modified so that $u_1(z_4) = 4$. Then the prediction shown by the dotted lines validates (IC_c) but not (IC_o) . In fact, by (a) of Proposition 4.3, since t_2 is an immediate successor of player 1's node t_0 , for (IC_o) to be satisfied at t_0 it must be that $\max\{u_1(z) : z \in Z \text{ and } t_2 \dot{A}_p z\} = \min\{u_1(z) : z \in Z \text{ and } t_0 \dot{A}_p z\}$; that is, given the prediction shown by the dotted arrows, it must be that $u_1(z_3) = u_1(z_4)$.

The following proposition characterizes the predictions that validate (IC_o) in arbitrary (that is, possibly non-generic) perfect information games. Part (a) of the proposition states that validity of (IC_o) implies that, for every decision node t , if there are multiple predictable paths out of t , they all lead to the same payoff for the player moving at t . The second part states that \dot{A}_p is obtained by extending a backward-induction relation subject to the constraint that, whenever an arrow from a node t to one of its immediate successors is added, the player who moves at t and all the players who move at predecessors t^{00} of t that satisfy the condition $t^{00} \dot{A}_p t$ are indifferent between the terminal nodes previously reachable from t and any other terminal node that becomes reachable due to the addition.

Proposition 4.4. Let G be a perfect information game, F the corresponding BTA frame and \dot{A}_p a prediction for F . Let M be a game model based on $\langle F; \dot{A}_p \rangle$ where axiom (IC_o) is valid. Then,

- (a) $\forall t \in T; \forall q_1, q_2 \in Q$, if $M; t \models F_p(u_{\eta(t)} = q_1) \wedge F_p(u_{\eta(t)} = q_2)$ then $q_1 = q_2$;
- (b) \dot{A}_p is the transitive closure of a subrelation \dot{A}_p of \dot{A} satisfying the following properties: (b.1) there is a backward relation \dot{A}_{BI} which is contained in \dot{A}_p , and (b.2) if $(t; t^0) \in \dot{A}_p$ and $(t; t^0) \in \dot{A}_{BI}$ then, for every t^{00} such that either $t^{00} = t$ or $t^{00} \dot{A}_p t$, $u_{\eta(t^{00})}(z) = u_{\eta(t^{00})}(z^0)$, where z is the unique terminal node \dot{A}_{BI} -reachable from t and z^0 the unique terminal node \dot{A}_{BI} -reachable from t^0 (if t^0 is a terminal node, then $z^0 = t^0$).

Proof. Part (a) is a corollary of (a) of Proposition 4.2. We only sketch the proof of part (b). Starting from a backward-induction relation (which, by Proposition 4.2, validates (IC_o)), by part (a) one can extend it without violating (IC_o) only by adding paths that leave all the players involved indifferent between the terminal nodes that become reachable due to the addition and the terminal nodes previously reachable. ■

5. 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 into 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 provided a syntactic characterization of backward induction in terms of the property of internal consistency of prediction and characterized the two notions of optimistic and conservative internal consistency.

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.

A. APPENDIX

Definition A.1. Given a finite perfect information game, for $k \geq 0$ define the set T_k of level k nodes 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) $t \notin T_{k-1}$, (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 $k-1$.

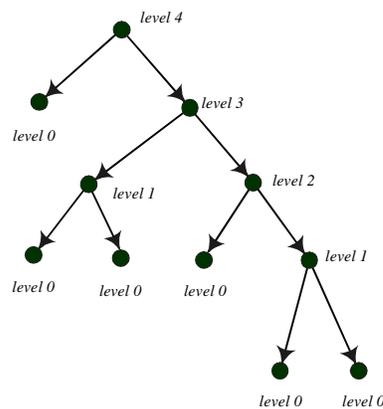


Figure 5

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 if and only if k is the length of the maximal path from t to a terminal node, as illustrated in Figure 5.

Definition A.2. Given a finite perfect information game, define, for $k \geq 1$ and $i \in N$, binary relations $\succsim_{B_i}^k$ on T and functions $u_i^k : T_k \rightarrow \mathbb{Q}$ recursively as follows:

$\succsim_{B_i}^1$ definition of $\succsim_{B_i}^1$:

(1) if $t \succsim_{B_i}^1 t^0$ then (a) $t \in T_1$ (that is, t is a level-1 node) and $t \succ t^0$, (b) $u_{\eta(t)}(t^0) \geq u_{\eta(t)}(t^{00})$ for all t^{00} such that $t \succ t^{00}$, (c) if $t \succ_{B_i}^1 t^0$ and $t \succ_{B_i}^1 t^{00}$ then $t^0 = t^{00}$ and

(2) $t \succ_{B_i}^1 t^0$ for some t^0 ;¹⁶

u_i^1 definition of $u_i^1 : T_1 \rightarrow \mathbb{Q}$: $u_i^1(t) = u_i^0(t^0)$ where $u_i^0 = u_i$ and t^0 is the unique node such that $t \succ_{B_i}^1 t^0$;¹⁷

$\succsim_{B_i}^k$ definition of $\succsim_{B_i}^k$ for $k > 1$:

(1) if $t \succ_{B_i}^k t^0$ then (a) $t \in T_k$ (that is, t is a level- k node) and $t \succ t^0$, (b) $u_{\eta(t)}^{\ell(t)}(t^0) \geq u_{\eta(t)}^{\ell(t)}(t^{00})$ for all t^{00} such that $t \succ t^{00}$, (c) if $t \succ_{B_i}^k t^0$ and $t \succ_{B_i}^k t^{00}$ then $t^0 = t^{00}$ and

(2) $t \succ_{B_i}^k t^0$ for some t^0 ;

u_i^k definition of $u_i^k : T_k \rightarrow \mathbb{Q}$ for $k > 1$: $u_i^k(t) = u_i^{\ell(t)}(t^0)$ where t^0 is the unique node such that $t \succ_{B_i}^k t^0$.

For the example of Figure 3b above, the following satisfy Definition A.1: $\succsim_{B_1}^1 = f(t_1; z_2); (t_2; z_3)g$; $\succsim_{B_1}^2 = f(t_0; t_2)g$, $(u_1^1(t_2); u_2^1(t_2); u_3^1(t_2)) = (2; 0; 1)$, $(u_1^1(t_1); u_2^1(t_1); u_3^1(t_1)) = (1; 1; 2)$, $(u_1^2(t_0); u_2^2(t_0); u_3^2(t_0)) = (2; 0; 1)$. Note that there may be several relations $\succsim_{B_i}^k$ and functions u_i^k that satisfy Definition A.1.

¹⁶Thus $\succsim_{B_i}^1$ mimics the first step of the backward induction algorithm: for every "last decision node" t , $\succsim_{B_i}^1$ associates with t a unique immediate successor t^0 which maximizes the payoff of the player assigned to node t :

¹⁷Thus, 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^0 selected by $\succsim_{B_i}^1$. 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 .

Definition A.3. Given a finite perfect information game $\langle T; \mathcal{I}_2; N; \mathcal{I}; f_{u_i g_{i2N}} \rangle$ a binary relation \mathcal{I}_{2BI} on T is called a backward induction relation if

$$\mathcal{I}_{2BI} = \bigcup_{k=1}^{\infty} \mathcal{I}_{2BI}^k$$

where the \mathcal{I}_{2BI}^k are relations obtained according to Definition A.2.¹⁸

Thus, for the example of Figure 3b, the following relation satisfies Definition A.3: $\mathcal{I}_{2BI} = \{ (t_0; t_2); (t_1; z_2); (t_2; z_3) \}$. Note that a given perfect information game might have more than one backward-induction relation. For example, for the game of Figure 3a, one backward induction relation is the one just described, which is illustrated in Figure 3b, and a different one is $\{ (t_0; t_1); (t_1; z_2); (t_2; z_4) \}$.

Proof of Lemma 3.4. We need to show that \hat{A}_p satisfies properties (P.4)-(P.7) of Definition 2.2. First of all, it is clear from Definition A.2 that \hat{A}_p is a subrelation of \hat{A} (the transitive closure of \mathcal{I}_2 : see Lemma 3.2). By construction, \hat{A}_p is transitive. It is easy to see from Definition A.2 that t is such that there is no t^0 with $t \hat{A}_p t^0$ only if and only if t is a terminal node; thus property (P.6) is satisfied. Finally, if $t_1 \hat{A}_p t_3$ and $t_1 \hat{A} t_2$ and $t_2 \hat{A} t_3$ then: (1) by definition of \hat{A} , there is a \mathcal{I}_2 -path from t_1 to t_3 through t_2 , (2) by definition of \hat{A}_p , there is a \mathcal{I}_{2BI} -path from t_1 to t_3 , which, since \mathcal{I}_{2BI} is a subrelation of \mathcal{I}_2 , is also a \mathcal{I}_2 -path from t_1 to t_3 . By definition of tree, the \mathcal{I}_2 -path from t_1 to t_3 is unique; hence the \mathcal{I}_{2BI} -path from t_1 to t_3 goes through t_2 . Thus, by definition of \hat{A}_p , we have that $t_1 \hat{A}_p t_2$ and $t_2 \hat{A}_p t_3$, that is, property (P.7) is satisfied. ■

References

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

¹⁸In game-theoretic terms, \mathcal{I}_{2BI} corresponds to the pure-strategy profile associated with a backward induction solution.

- [4] Bonanno, Giacomo (1998), Branching time logic, perfect information games and backward induction, Working Paper #98-13, University of California Davis.
- [5] Burgess, John, Basic tense logic, in: D. Gabbay and F. Guentner (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, 87-172.
- [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] Horty, John (1996), Agency and obligation, *Synthese*, 108: 269-307.
- [13] Horty, John and Nuel Belnap (1995), The deliberative stit: a study of action, omission, ability and obligation, *Journal of Philosophical Logic*, 24: 583-644.
- [14] von Neumann, John and Oscar Morgenstern (1947), *Theory of games and economic behavior*, Princeton University Press.
- [15] Åhrström, Peter and Per F. V. Hasle, *Temporal logic*, Kluwer, 1995.