Logic for games of perfect information and epistemic conditions for backward induction and for subgame perfectness

Tai-Wei Hu*

Department of Economics, Pennsylvania State University

This version: 2007/03/28

Abstract

We propose a logical system in which a notion of the structure of a game is formally defined and the meaning of sequential rationality is formulated. We provide a set of decision criteria which, given sufficiently high order of mutual belief of the game structure and of every player following these criteria, entails Backward Induction decisions in generic perfect information games. We say that a player is rational if the player follows these criteria in his/her decisions. The set of mutual beliefs is also necessary, in the sense that any mutual belief of lower order can not entail the Backward Induction decisions. These conditions are determined by the length of the game structure, and they are never involved with common belief. Moreover, we give a set of epistemic conditions for subgame perfect equilibria for any perfect information game, which requires every player follow these decision criteria and there be mutual belief of the the equilibrium strategy and of the game structure.

JEL classification: C72, D81

Keywords: Backward Induction, Game Theory, Knowledge, Decision Theory, Subgame Perfect Equilibrium

---

*E-mail: tuh129@psu.edu; Address: 608 Kern Graduate Building, University Park, PA 16802-3306, USA.

1 Introduction

This paper develops a multi-player epistemic logic with additional structures in which we give a precise notion of the structure of an extensive form game and we express the meaning of sequential rationality[1] by means of a set of decision criteria. An agent is said to be rational if this agent follows these criteria in making his/her decisions. We consider the epistemic conditions on mutual belief of players' belief on the decision criteria they follow, and their belief of the structure of the game. These decision criteria, with some adequate epistemic conditions, entail Backward Induction decisions in generic extensive form games with perfect information. Moreover, these criteria refines the equilibrium with subgame perfectness. This logical system is a combination of a epistemic logic in Halpern and Moses[(8)] with a logic of causality similar to the one developed in Giordano and Schwind[(7)], and it is able to express players' beliefs and the structures of extensive form games.

We shall first illustrate the meaning of the structure of a game with the following 2-player simple game playing on the matrix:

×		
○	○	
	×	○

The circle player and the cross player put a circle and a cross in a cell, respectively, in an alternative manner, and the player who first have a row, a column, or a diagonal of circles or crosses wins. We call a matrix with some crosses and some circles in its cells a situation, and we call putting a circle or a cross in a particular cell an action. We say that a situation is an ending situation if the matrix has entries in all its cells. Denote the set of situations and ending situations by N and Z, respectively, and denote the set of actions by A. Notice that not every situation is possible, nor is every action in every situation.

[1]This concept is the 'substantial rationality' in Aumann[(1)]. The term 'sequential rationality' is more often used in the literature, but with a somehow loose meaning attached to it.

The rule of the game can then be defined as a function from $N \times A$ to $N \cup Z$, which indicates how the situations would involve as players take actions. Moreover, we specify the situations succeeding each particular situation by giving a relation on N, and we specify a beginning situation, denoted by v^0, which in this case is the empty matrix. Finally, we specify the preference relation on the ending situations for each player in this game. In the logical system we develop, all these pieces of information about the structure of a game can be formulated explicitly, and hence we could formulate the sentence that says there is certain finite order of mutual belief of the structure of the game, but there may not be common belief. This is the novel part of this paper, since in the literature on epistemic conditions of game theory, almost all papers assume that the structure of the game is common knowledge.[2]

We follow the interpretation of decision criteria in Kaneko[(10)] and assume that players use these criteria to reach their final decisions, together with their predictions from their beliefs and logical derivations. However, since we are dealing with dynamic games here, we need the notion of sequential rationality. Within our logical system, we can formulate the concept of sequential best responses, and so the decision criteria that capture the concept of sequential rationality, in an abstract way, without any reference to a particular game. Since rationality is defined as following the decision criteria, it then is possible that a player i believes that another player j follows the criteria but i may not believe that j knows the structure of the game.

The set of decision criteria we shall investigate in this paper is called DC, which consists of two parts: RC and PC. The first criterion, RC, requires that any decision to take an action in a particular situation implies that the action leads to a weakly preferred consequence to any other consequence lead by some other available action, given the agent's predictions of future decisions. The second criterion, PC, requires that the agent to take an action if it leads to a strictly preferred consequence to all other consequences

[2]The literature of games with incomplete information deals with incomplete information of payoff functions, and it seems not very relevant to talk about the structures of the games for normal form games. Aumann and Brandenburger[(2)] is another paper with models in which there may be no common belief of the payoff functions.

lead by other available actions, given the agent's predictions of future decisions. These criteria are intended to capture the meaning of sequential rationality and thus the meaning of the preferences of players and implicitly assumes consequentialism in players' decisions. It may be remarkable that all our analysis is concerned with *ex ante* decisions or plans, which are differentiated from actual actions or moves in *ex post* plays. We could not find such a distinction in previous papers like Aumann[(1)], Balkenborg and Winter[(3)], or Samet[(15)], since all these papers have taken a model theoretic approach, and have treated rationality as an attribute of a particular action.[3]

Our first finding is that, given every player following these decision criteria, there is a set of sufficient and necessary epistemic conditions such that with these conditions DC entails Backward Induction decisions. In a given generic game[4] with length K, these epistemic conditions are mutual beliefs of players following these criteria and of the structure of the game up to order $K-1$. This result gives a more precise meaning of substantial rationality in Aumann[(1)], which is regarded as a feature of a habitual payoff maximizer. In our framework, this habitual payoff maximizer can be interpreted as an agent who follows DC for his/her decisions. With this interpretation, it might be easier to identify situations in which this assumption may apply. Following this particular decision criteria should be an attribute of a combination of a player and of a situation.

It may not be surprising that these conditions do not demand common belief of the game structure or common belief of following these criteria. Balkenborg and Winter[(3)] has already argued that common belief of rationality is not necessary for Aumann[(1)]'s result.[5] Balkenborg and Winter[(3)] also have reached a necessary condition, which almost coincides with ours regarding rationality but not structure of the game, for a special class of extensive form games with perfect information. Clausing[(6)] also has provided a

[3]Although the conditions in these literature often read as belief in players' rationality, it more or less means players' actions' rationality. In those models, rationality is defined in terms of the consistency between actions and preferences.

[4]We use the term to mean a game with perfect information which has a unique subgame perfect equilibrium.

[5]The authors even have conjectured that common belief of payoff information is not necessary either. However, they did not explicitly prove this conjecture.

theorem which states that even common belief of the game structure is not necessary. However, the structure of the game is formulated as a belief revision system consistent with the game structure, which should be implications of the game structure rather than primitive elements of the structure. Moreover, in his formulation, being rational in a particular game is not independent of knowledge of the game. We adopt this approach instead of the belief revision approach taken there and Board[(5)], among others, to model the structure of the game since we are dealing with *ex-ante* decision-making processes.

Our second result concerns equilibrium refinement. Given an extensive form game, we show that there exists a model for this game in which there is mutual belief of the structure of the game and of all the players following a particular strategy and every player follows that strategy and DC if and only if the strategy profile is a subgame perfect equilibrium. This result has similar flavor to those in Aumann and Brandenburger[(2)]. In their paper there is a preliminary observation which states that rationality and mutual knowledge of players' actions in normal form games entail that the actions constitute a Nash equilibrium. We show that DC refines the equilibrium to have subgame perfectness. However, since we deal with pure strategies here, the results are silent about mixed equilibria.

The rest of the paper is organized as follows: section 2 provides an example to illustrate the need for a new logical system, section 3 formulates such a system and states its completeness and soundness, section 4 gives formulations of game structures and decision criteria, and in section 5 we present our main results regarding Backward Induction and subgame perfect equilibrium. Section 6 gives some conclusion remarks.

2 An Example

In this section we shall discuss an example to illustrate potential defects of classical logical system in formulating game structures. We defer the formal definitions to later sections, but we introduce some sentences and make inferences somehow informally whenever necessary. Consider an extensive form game with the following game tree:

$U \qquad\qquad\qquad u$

$v_0 - - - - v_1 - - - - z_3$

$|\qquad\qquad\quad |$

$| \; D \qquad\qquad | \; d$

$z_1 \qquad\qquad\quad z_2$

Player 1 has to take an action at v_0, while player 2 has to take an action at v_1. To describe the rules of the game, we consider the following sentences: a. $d_1(v_0, U)$, $d_1(v_0, U)$, $d_2(v_1, u)$, $d_2(v_1, d)$; the intended meaning of the sentence $d_1(v_0, U)$ is player 1 taking action U in situation specified by v_0, etc.; b. $o(v_0)$, $o(v_1)$, $o(z_1)$, $o(z_2)$, $o(z_3)$; the intended meaning of $o(v_1)$ is that v_0 occurs, etc. The rules should specify that $d_1(v_0, U)$ would lead to $o(v_1)$, $d_1(v_0, D)$ would lead to $o(z_1)$, etc. Let $\wedge$, $\sim$, and $\Rightarrow$ stand for conjunction, negation, and material implication in classical propositional logic.

One possible axiomatization of the game structure includes the following sentences as axioms: $d_1(v_0, U) \Rightarrow o(v_1)$, $d_1(v_0, D) \Rightarrow o(z_1)$, $d_2(v_1, u) \Rightarrow o(z_3)$, $d_2(v_1, d) \Rightarrow o(z_3)$, and $\wedge\{o(z) \Rightarrow \sim o(z') : z, z' = z_1, z_2, z_3, z \neq z'\}$. Suppose that player 2 decides to take d at v_1 and suppose that player 1 knows this. This could be formulated as $o(v_1) \Rightarrow d_2(v_1, d)$. Assume that z_1 is strictly preferred to z_2, and z_2 is strictly preferred to z_3 for player 1, which could be formulated as $P_1(z_1, z_2) \wedge \sim P_1(z_2, z_1)$ and $P_1(z_2, z_3) \wedge \sim P_1(z_3, z_2)$. Then D is a sequential best response while U is not, given player 2's decision.

A natural way to formulate a being a sequential best response at v_0, $a = D$ or U, in this context is as follows:

$$\wedge\{(d_1(v_0, a) \Rightarrow o(z)) \wedge (d_1(v_0, a') \Rightarrow o(z')) \Rightarrow P_1(z, z') : a' \neq a, z, z' = z_1, z_2, z_3\}.$$

It can be shown that U not being a best response can be derived form the above axioms plus the condition $o(v_1) \Rightarrow d_2(v_1, d)$. However, to show that D is a best response, it requires that $\sim (d_1(v_0, D) \Rightarrow o(z_3))$ be provable from some the axioms, for otherwise the sentence

$$(d_1(v_0, D) \Rightarrow o(z_3)) \wedge (d_1(v_0, U) \Rightarrow o(z_2)) \wedge \sim P_1(z_3, z_2)$$

is consistent with the axioms for the game, and so

$$(d_1(v_0, D) \Rightarrow o(z_3)) \wedge (d_1(v_0, U) \Rightarrow o(z_2)) \Rightarrow P_1(z_3, z_2)$$

is not provable from the axioms for the game. By classical propositional logic, it can be shown that

$$\sim (d_1(v_0, D) \Rightarrow o(z_3)) \Rightarrow d_1(v_0, D)$$

is a tautology. This, by completeness of the classical propositional logic, then requires that $d_1(v_0, D)$ be derivable from the axioms, which is not plausible.

Thus, we need another formulation of the causality in extensive form games than the material implication. In particular, Aumann[(1)] could avoid this problem because in his setup payoffs are implicitly given by the strategies, and the structure is not explicitly formulated.

3 The Epistemic Logic with Causality

We shall develop a system, called ECL suitable for describing the structure of the game, players' beliefs, and their inferences in this section. We call this system an epistemic logic with causality with n agents $\{1, ..., n\}$. In extensive form games, the causal relations of actions to situations are expressed by a tree structure, and our logical system will be able to capture the tree structure. Since we shall discuss finite extensive form games with perfect information, we adopt propositional logic formulation.

First we give a general definition of a language that will be specific to games in next section. There is a countable set of primitive propositions, which is denoted by Φ, with typical elements p and q. There are three logical connectives: $\sim$(negation), $\wedge$(and), and $>$ (causality). $\sim \varphi$ means that φ does not hold, $\varphi \wedge \chi$ means that both φ and χ hold, and $\varphi > \chi$ means that φ causes χ. Epistemic operators are as follows: B_1, ..., B_n. $B_i(p)$ means that i believes p.

Formulas are defined as follows:

We define L^0 first:

a.1 Any p in Φ belongs to L^0.

a.2 If $\zeta, \xi \in L^0$, then $\zeta \wedge \xi \in L^0$, and $\sim \zeta \in L^0$.

a.3 Any element of L^0 is a result of finite applications of a.1 and a.2.

Suppose that we have defined $L^{k-1}, k > 0$. We define L^k inductively as follows:

b.1 Any φ in L^{k-1} belongs to L^k.

b.2 If $\zeta \in L^0$ and $\psi \in L^{k-1}$, then $(\zeta > \psi) \in L^k$.

b.3 If $\varphi, \psi \in L^k$, then $\varphi \wedge \psi \in L^k, \sim \varphi \in L^k$, and $B_i(\varphi) \in L^k, i = 1, ..., n$.

b.4 Any element of L^k is a result of finite applications of b.1, b.2, and b.3.

Define $L = \cup_{k=0}^{\infty} L^k$. The language is similar to the one in standard epistemic logic except for the new connective $>$. We complicate the formation of formulas because we do not allow the precedent of any causation to have epistemic or causation element. For example, $B_i(p) > q$ is not allowed in L, nor is $(p > q) > r$. However, both $B_i(p > B_j(q))$ and $p > B_i(q > r)$ are allowed. We restrict our language in order to simplify the presentation of semantics, and these sentences are the only components we need for the expressions of the extensive form games.[6] As usual, we use $\varphi \vee \chi$ for the abbreviation of $\sim (\sim \varphi \wedge \sim \chi)$ and we use $\varphi \Rightarrow \chi$ for $\sim \varphi \vee \chi$, $\varphi, \chi \in L$. In what follows, we use the following notation: by $\wedge\{\chi_1, ..., \chi_k\}$ we mean $\chi_1 \wedge ... \wedge \chi_k$.[7]

Our logic consists of two parts — semantics and syntax. The syntax gives the formal description of an axiomatic system and proofs of theoremhood in that system, while semantics gives formal models for the system in which the truth values of sentences are given. The two systems are connected by the soundness and completeness theorem, which states that every theorem in the syntax is true in the semantics and that every true sentence in the semantics is provable in the syntax. In the following subsections, we

[6] For more discussions of a suitable language to express causality, please see Giordano and Schwind[(7)].

[7] This is not precise in terms of logic. However, since all propositional tautologies are true and provable in our logic, this does not matter too much.

shall present the syntax and the semantics for our logic, and prove its soundness and completeness.

3.1 Syntax

In this subsection, we consider the syntax for our logic ELC. Let $\varphi, \psi, \chi \in L$, and let $\xi, \zeta, \eta \in L^0$.

Logical Axiom Schemes:

(PC1) $\varphi \Rightarrow (\chi \Rightarrow \varphi)$.

(PC2) $(\varphi \Rightarrow (\psi \Rightarrow \chi)) \Rightarrow ((\varphi \Rightarrow \psi) \Rightarrow (\varphi \Rightarrow \chi))$.

(PC3) $((\sim \varphi) \Rightarrow (\sim \chi)) \Rightarrow (((\sim \varphi) \Rightarrow \chi) \Rightarrow \varphi)$.

Causal Axiom Schemes:

(Tran) $(\zeta > \xi) \wedge (\xi > \eta) \Rightarrow (\zeta > \eta)$.

(Det) $(\zeta >\sim \varphi) \Rightarrow\sim (\zeta > \varphi)$.

Epistemic Axiom Schemes ($i = 1, ..., n$):

(K) $(B_i(\varphi) \wedge B_i(\varphi \Rightarrow \chi)) \Rightarrow B_i(\chi)$.

(D) $\sim B_i(\varphi \wedge \sim \varphi)$.

(4) $B_i(\varphi) \Rightarrow B_i(B_i(\varphi))$.

Inference Rules:

(MP) From φ and $\varphi \Rightarrow \chi$, infer χ.

(RCEA) From $\zeta \Leftrightarrow \xi$, infer $(\zeta > \varphi) \Leftrightarrow (\xi > \varphi)$.

(RCK) From $(\varphi_1 \wedge ... \wedge \varphi_k) \Rightarrow \chi$, infer $((\zeta > \varphi_1) \wedge ... \wedge (\zeta > \varphi_k)) \Rightarrow (\zeta > \chi)$.

(NEC) From φ infer $B_i(\varphi)$.

(CN) From φ infer $(\zeta > \varphi)$.

Axioms (PC1-3) are standard in propositional logic. Notice also that by (PC1-3) and completeness of classical propositional logic, any tautology (including substitution of any formula in L with any primitive proposition in the original tautology) in classical logic is also provable in ELC. Axiom (Tran) captures the transitivity in the causal relations in extensive form games. (Det) captures the assumption that any action has a deterministic result in extensive form games. If effect, it rules out inconsistent consequences from any causes. Axioms (K), (D), and (4) are standard in epistemic logic, and we refer the readers to Kaneko[(11)] for more detailed discussions regarding the plausibility and importance of these axioms. (MP) and (NEC) are standard inference rules in epistemic logics. We regard (RCK), (RCEA), and (CN) as intuitive requirements — they are standard in the literature of conditional logic, and notice that they are inference rules, and so they can be applied only when $(\varphi_1 \wedge ... \wedge \varphi_k) \Rightarrow \chi$ or $\zeta \Leftrightarrow \xi$ or φ are theorems.

A proof in ELC is a finite ordered set of formulas $(\varphi_1, ..., \varphi_k)$ in L such that each φ_j is either an instance of the axiom schemes or a result of applications of the rules to formulas preceding it. We say that $\varphi \in L$ is a theorem of ELC if there is a proof $(\varphi_1, ..., \varphi_j)$ with $\varphi_k = \varphi$, and this is denoted by $\vdash \varphi$. We list several lemmas that will be useful for later results.

Lemma 3.1. $\vdash B_i(\varphi \wedge \chi) \Leftrightarrow B_i(\varphi) \wedge B_i(\chi)$.

Proof. See, for example, Kaneko[(11)]. □

Lemma 3.2. *a.* $\vdash (\eta \Rightarrow \zeta)$ *and* $\vdash \xi \Rightarrow (\zeta \Rightarrow \chi)$ *implies that* $\vdash \eta \wedge \xi \Rightarrow \chi$.

b. $\vdash ((\eta \wedge \zeta) \Rightarrow \xi) \Leftrightarrow (\eta \Rightarrow (\zeta \Rightarrow \xi))$.

Proof. See the appendix. □

3.2 Semantics

In this subsection we present a model theoretical semantics for ELC. The semantics we present here is based on the semantics adopted in Halpern and Moses[(8)] for epistemic logic and we change it to accommodate causality using selection functions along the lines developed by Lewis[(12)] and others. The closest system to the present one may be the causal logic in Giordano and Schwind[(7)]. The primitive constitutes are frames usually adopted in modal logic, and a frame consists of a set of possible worlds W and a binary relation R_i on W for each agent i. The interpretation of these relations is standard. We have two additional elements — a collection Σ of subsets of Ω, which is supposed to include every event that can potentially be a cause, and a function $f : \Sigma \times W \to 2^{\Omega}$. The subset corresponding to a particular event in Σ at $w \in W$ could be interpreted as the set of plausible worlds by amending the world w to accommodate the event. We give a formal definition as follows.

Definition 3.1. Given the language L defined in the beginning of this section, we define a model M as a $(n+4)-$tuple: $< W, \Sigma, f, R_1, ..., R_n, V >$, where W is the set of possible worlds, f is a mapping from $\Sigma \times W \to 2^W$, where Σ is an algebra of subsets of W, and each R_i is a mapping from W to 2^W, $i = 1, ..., n$; V is a mapping from $\Phi \times W$ to $\{\top, \bot\}$ such that for all $p \in \Phi$, $\{w \in W : V(p, w) = \top\} \in \Sigma$. We then extend V to $L \times W$ as follows:

(a) For ζ, $\xi \in L^0$:

(a.1) $V(\sim \zeta, w) = \top$ if and only if $V(\zeta, w) = \bot$.

(a.2) $V(\zeta \wedge \xi, w) = \top$ if and only if $V(\zeta, w) = V(\xi, w) = \top$.

(b) Suppose that V has been defined for all $(\varphi, w) \in L^{k-1} \times W$;

(b.1) $V(\zeta > \varphi, w) = \top$ if and only if for all $w' \in f([\zeta], w)$, $V(\varphi, w') = \top$, where $[\zeta] = \{w \in W : V(\zeta, w) = \top\}$.

(b.2) $V(\sim \varphi, w) = \top$ if and only if $V(\varphi, w) = \bot$.

(b.3) $V(\zeta \wedge \xi, w) = \top$ if and only if $V(\zeta, w) = V(\xi, w) = \top$.

(b.4) $V(B_i(\varphi), w) = \top$ if and only if for all $w' \in R_i(w)$, $V(\varphi, w') = \top$.

Define $[\varphi] = \{w \in W : V(\varphi, w) = \top\}$, for all $\varphi \in L$. We also require f and $R_1, ..., R_n$ satisfy the following conditions:

(s-Tran) for all $w \in W$ and for all X_1, X_2, $X_3 \in \Sigma$, $f(X_1, w) \subseteq X_2$ and $f(X_2, w) \subseteq X_3$ imply that $f(X_1, w) \subseteq X_3$;

(s-Det) for all $w \in W$ and for all $X \in \Sigma$, $f(X, w) \neq \emptyset$;

(s-D) for all $w \in W$, $R_i(w) \neq \emptyset$;

(s-4) for all $w \in W$ and $w' \in R_i(w)$ and $w'' \in R_i(w')$ imply that $w'' \in R_i(w)$.

The set of models satisfying all these restrictions is denoted by $\mathfrak{M}$.

Given a model M, we say that $\varphi \in L$ is true at state w if and only if $V(\varphi, w) = \top$, and this is denoted by $(M, w) \models \varphi$. We say that $\varphi \in L$ is true in the model M if and only if $(M, w) \models \varphi$ for all $w \in W$ and this is denoted by $M \models \varphi$. We say that $\varphi \in L$ is valid (w.r.t. $\mathfrak{M}$) if and only if $M \models \varphi$ for all $M \in \mathfrak{M}$. Validity of φ is denoted by $\models \varphi$.

Lemma 3.3. *For any $\zeta \in L^0$, we have $[\zeta] \in \Sigma$, and so the extension of V to $L \times W$ is well-defined.*

Proof. Clearly $[p] \in \Sigma$ for any $p \in \Phi$. Now, suppose that $[\zeta]$, $[\xi] \in \Sigma$, then $[\sim \zeta] = W - [\zeta] \in \Sigma$ and $[\zeta \wedge \xi] = [\zeta] \cap [\xi] \in \Sigma$. □

Lemma 3.4. *We have the following properties for $\models$:*

a. $(M, w) \models \varphi$ *if and only if* $(M, w) \not\models \sim \varphi$*;*

b. $(M, w) \models \varphi \Rightarrow \chi$ *if and only if* $(M, w) \models \chi$ *or* $(M, w) \not\models \varphi$*;*

c. $(M.w) \models \varphi \vee \chi$ *if and only if* $(M, w) \models \varphi$ *or* $(M, w) \models \chi$*;*

d. $M \models \varphi \Rightarrow \chi$ *if and only if* $[\varphi] \subseteq [\chi]$.

Proof. a. Suppose that $(M, w) \models \varphi$, then $V(\sim \varphi, w) = \bot$ and so $(M, w) \not\models \sim \varphi$. Conversely, if $(M, w) \not\models \sim \varphi$, then $V(\sim \varphi, w) = \bot$ and so $V(\varphi, w) = \top$.

c. Suppose that $(M.w) \models \varphi \vee \chi = \sim (\sim \varphi \wedge \sim \chi)$. Then $(M, w) \not\models (\sim \varphi \wedge \sim \chi)$, and so $(M, w) \not\models \sim \varphi$ or $(M, w) \not\models \sim \chi$. By (a), this is equivalent to $(M, w) \models \varphi$ or $(M, w) \models \chi$.

b. By c. and $\varphi \Rightarrow \chi = \sim \varphi \vee \chi$.

d. Suppose that $M \models \varphi \Rightarrow \chi$. Then for all w, $(M,w) \models \chi$ or $(M,w) \not\models \varphi$. Thus, if $(M,w) \models \varphi$, then $(M,w) \models \chi$, and so we have $[\varphi] \subseteq [\chi]$. Suppose $(M,w) \not\models \varphi \Rightarrow \chi$. Then $(M,w) \not\models \chi$ and $(M,w) \models \varphi$ and so $w \notin [\chi]$ but $w \in [\varphi]$. $\square$

3.3 Completeness

In this subsection, we shall give a theorem which states that a formula in L is a theorem in ELC if and only if it is valid in all models in $\mathfrak{M}$. The "if" part of the theorem is called the completeness of the logic ELC, and the "only if" part of the theorem is called soundness.

We shall say that a formula φ in L is consistent if and only if it is not the case that $\vdash \sim \varphi$. A finite set of formulas $\{\varphi_1, ..., \varphi_K\}$ is consistent if and only if it is not the case that $\vdash \sim \varphi_1 \vee .... \vee \sim \varphi_K$. An arbitrary subset of L is consistent if and only if its every finite subset is consistent. A subset $\Gamma \subseteq L$, is maximal consistent if and only if Γ is consistent and for any $\varphi \notin \Gamma$, $\Gamma \cup \{\sim \varphi\}$ is inconsistent.

We state the completeness and soundness of the logic ELC with respect to $\mathfrak{M}$ in the following theorem. The proof can be found in the appendix.

Theorem 3.1. *For any $\varphi \in L$, $\models \varphi$ if and only if $\vdash \varphi$.*

4 Extensive Form Game Logic

In this section we shall develop a logic for extensive form games based on ELC. This logic enables us to formulate the concepts and decision criteria necessary for Backward Induction as an ex-ante decision. In order to describe the structure of a game explicitly, we take a set of decision nodes, a set of terminal nodes, a set of actions, and all possible succeeding relations of the decision nodes as primitive elements. We take the assignment of players at each decision node as an attribute of that node. The players are assumed to know these primitive elements implicitly, and this is common knowledge. However, they may not know the structure of a particular game, and this knowledge is never common belief in this paper.

This formulation can accommodate games with perfect information and, potentially, multi-stage games, but our focus is the former. Therefore, we shall only present the formulation of games with perfect information, and leave the latter for further research.

4.1 The Language

We shall now define the extensive form game logic. Let $I = \{1, ..., n\}$ be the set of players. Let N_i^0 denote the set of decision nodes for player i, $i = 1, ..., n$, and let Z^0 denote the set of terminal nodes. Define $N^0 = \cup_{i \in I} N_i^0$. Since we shall consider games with perfect information only, $N_i^0 \cap N_j^0$ is assumed to be empty for any pair $(i, j) \in I^2$. Also, for any $v \in N^0$, we use $\tau(v)$ to denote the index i such that $v \in N_i^0$. Both N^0 and Z^0 are assumed to be large but finite. Let A^0 be the set of all possible actions, which is also assumed to be finite. Without loss of generality, let $v^0 \in N_1^0$ be the initial node, and let $z^* \in Z^0$ be a fixed element that will be interpreted as an impossible outcome.

The set Φ of atomic formulas is as follows:

a. For each $v \in N_i^0$ and $a \in A^0$, $d_i(v, a)$ is an atomic formula, and we denote the set of these formulas as Φ_{GR}.

b. For each $v \in N^0 \cup Z^0$, $o(v)$ is an atomic formula, and we denote the set of these formulas as Φ_{NZ}.

c. For each pair v, $v' \in N^0$, $S(v, v')$ is an atomic formula, and we denote the set of these formulas as Φ_S.

c. For each z, $z' \in Z^0$, $P_i(z, z')$ is an atomic formula, and we denote the set of these formulas as Φ_S.

The intended meaning for $d_i(v, a)$ is that the action taken by player i in situation v is a, and the intended meaning for $o(v)$ is that the situation v occurs. We shall also denote the sentence by $(o(v) > d_i(v, a))$ by $D_i(v, a)$, and it is intended to mean that player i

decides to take action a in situation v. We distinguish $D_i(v,a)$ from $d_i(v,a)$ as different concepts — actual behavior and plans are not the same objects — but they should be closely related. Now we shall express a game structure with the language.

First, we define a description as follows:

Definition 4.1. A description consists of three sets of formulas (G_C, G_S, G_P) that satisfy the following properties:

(C1) $G_C = \Gamma \cup \{(o(z) >\sim o(z')) : z,\ z' \in Z^0,\ z' \neq z\}$, for some Γ being a set such that every formula in it has the form $d_i(v,a) > o(v')$ for some $v \in N_i^0$, $v' \in N^0 \cup Z^0$, and some $a \in A^0$, and if $(d_i(v,a) > o(z))$ and $(d_i(v',a') > o(z))$ belong to G_C for some $(v,a) \neq (v',a')$, then $z = z^*$;

(C2) for each pair $(v,a) \in N_i^0 \times A^0$, if $(d_i(v,a) > o(v')) \in G_C$ for some $v' \in N^0 \cup Z^0$, then for any other $v'' \neq v'$, $(d_i(v,a) > o(v'')) \notin G_C$, and for all $a' \in A^0$, $(d_i(v,a') > o(v'')) \in G_C$ for some $v'' \in N^0 \cup Z^0$, and for at least one $a'' \in A^0$, such v'' is not z^*;

(P1) $\{P_i(z,z^*), \sim P_i(z^*,z), P_i(z^*,z^*), P_i(z,z) : z \in Z^0 - \{z^*\}, i \in I\} \subseteq G_P$, and every formula in G_P has the form $P_i(z,z')$ or $\sim P_i(z,z')$ for some $i \in I$, $z, z' \in Z^0$;

(P2) for every pair $z, z' \in Z^0$, if there exist $(v,a),(v',a') \in N_i^0 \times A^0$ such that $d_i(v,a) > o(z), d_i(v',a') > o(z') \in G_C$, then, for each $i' \in I$, $P_{i'}(z,z')$ or $P_{i'}(z',z)$ or both belong to G_P and either $P_{i'}(z,z')$ or $\sim P_{i'}(z,z')$ belongs to G_P;

(S1) for every pair $(v,v') \in (N^0)^2$, either $S(v,v')$ or $\sim S(v,v')$ belongs to G_S and any formula in G_S has this form, and $\sim S(v,v^0) \in G_S$ for all $v \in N^0$, and if $d_{\tau(v)}(v,a) > o(v') \in G_C$, $v \in N^0$, $v' \in N^0 \cup Z^0$, then $S(v,v') \in G_S$ or $S(v'',v) \in G_S$ for some $v'' \in N^0$ or $v = v^0$;

(S2) if $S(v,v') \in G_S$, then there is a unique $a \in A^0$ such that $(d_{\tau(v)}(v,a) > o(v')) \in G_C$ and if $S(v,v') \in G_S$ but there is no $v'' \in N^0$ such that $S(v',v'') \in G_S$, then there is some $(a,z) \in A^0 \times Z^0$ such that $(d_{\tau(v')}(v',a) > o(z)) \in G_C$.

The set G_C gives us the rules of the game — it specifies how situations involve as players take actions in turns. The sentences with the form $(o(z) >\sim o(z'))$ ensure that every consequence is final, and we identify the consequences with the terminal nodes. We require that, in (C2), every action in every situation leads to some other situation in order to make our description unique. This requirement, together with the conditions we give later, asserts that any infeasible action leads a least preferred consequence z^*, which is an interpretation of infeasibility consistent with Harsanyi[(9)]. G_P gives the information about the personal evaluation of the results for every player, and here we impose only completeness. G_S provides the information necessary for the player to know what are the possible decision nodes of all the players in the future, and so the players could make predictions regarding future plays. This set is required, as in (S2), to be consistent with the rules of the game specified in G_C.

Consider an extensive form game:

$$g = [I, (N, Z, \prec), \tau, A, \{t_v\}_{v \in N}, \{\succeq_i\}_{i \in I}],$$

where

$I = \{1, ..., n\}$ is the set of players;

$(N, Z, \prec)$ is a tree: a. $\prec$ is a binary relation on $N \cup Z$ such that for all v, not $v \prec v$; b. there is a unique element $v_0 \in N$ such that for all $v \in N$, not $v \prec v_0$, and for all other $v \in N \cup Z$, $v' \prec v$ for a unique $v' \in N$; c. for all $v \in Z$, and for all $v' \in N \cup Z$, not $v \prec v'$.

τ is a function from N to I (player assignment);

$\succeq_i$ is a preference relation on Z, $i = 1, ..., n$;

A is a finite set of actions;

t_v is a one-to-one mapping from $Succ(v)$ to A, with $Succ(v) = \{v' \in N \cup Z : v \prec v'\}$; We denote the range of t_v by $A(v)$.

For each $v \in N$, define $Sub(v) = \{v' \in N$: for some $v_1, ..., v_k$, $v \prec v_1 \prec ... \prec v_k = v'\}$ and $Ter(v) = \{v' \in Z$: for some $v_1, ..., v_k, v \prec v_1 \prec ... \prec v_k = v'\}$. Define $N_i = \{v \in N : \tau(v) = i\}$.

A strategy s in g is a mapping from N to A such that $s(v) \in A(v)$. Given a node

$v \in N$, we define the sub-game $g(v)$ to be

$$[I, ((\{v\} \cup Sub(v)) \cap N, Ter(v) \cap Z, \prec), \tau, A, \{t_{v'}\}_{v' \in \{v\} \cup Sub(v)}, \{\succeq_i\}_{i \in N}].$$

For any strategy s, we define $s|_v$ to be the same strategy but restrict its domain to be $Sub(v)$ or $\{v\} \cup Sub(v)$, and it shall be clear which case it is from the context. Moreover, we define $Path(v, z) = \{v' \in N : z \in Ter(v') \text{ and } v' \in Sub(v) \cup \{v\}\}$ if $z \in Ter(v)$. We also use $Path(z)$ to denote $Path(v^0, z)$. For a given strategy s and a node $v \in N$, we define $Z(v; s) = \{z \in Z : \text{for some } v = v_1, ..., v_k = z,\ t_{v_{l-1}}(v_l) = s(v_{l-1}),\ l = 2, ..., k\}$.

We assume that $N_i \subseteq N_i^0$, $Z \subseteq Z^0 - \{z^*\}$, and $A = A^0$. Also, we assume that $v^0 \in N_1^0$ is identified with the initial node in g, i.e., we assume that the game always begins with player 1.

Given an extensive form game with perfect information, the following description is a description for g:

a. $G_C = \{(d(v, a) > o(v') : v, v' \in N \cup Z, t_v^{-1}(a) = v'\} \cup \{d(v, a) > o(z^*) : v \in N, a \notin A(v)\} \cup \{(o(z) >\sim o(z')) : z,\ z' \in Z^0,\ z' \neq z\}$;

b. $G_P = \{P_i(z, z') : (z, z') \in \succeq_i, z, z' \in Z\} \cup \{\sim P_i(z, z') : (z, z') \notin \succeq_i, z, z' \in Z\} \cup \{P_i(z, z^*), \sim P_i(z^*, z) : z \in Z^0 - \{z^*\},\ i = 1, ..., n\}$;

c. $G_S = \{S(v, v') : v, v' \in N, v \prec v'\} \cup \{\sim S(v, v') : v \in N, v' \in N, v \not\prec v'$, or $(v, v') \in (N^0)^2 - N^2\}$;

d. $G = \wedge(G_C \cup G_P \cup G_S)$ is then defined to be the axiom for g, and (G_C, G_S, G_P) is called the description for g.

An arbitrary description may (G_C, G_S, G_P) may not be a description of any game. The following lemma characterizes the conditions for a description to be a description of a game in extensive form with perfect information. It also shows that if such a game exists, it is unique.

Lemma 4.1. *Given a description (G_C, G_S, G_P), the formula $G = \wedge(G_C \cup G_P \cup G_S)$ is an axiom for some extensive form game with perfect information if and only if:*

a. G is consistent;

b. $(\wedge G_P) \wedge (\wedge\{P_i(z,z') \wedge P_i(z',z'') \Rightarrow P_i(z,z'') : z,z',z'' \in Z^0\})$ *is consistent.*

c. $(\wedge G_S) \wedge (\wedge\{S(v,v') \Rightarrow \sim S(v'',v') : v,v',v'' \in N^0\})$ *is consistent;*

d. $(\wedge G_S) \wedge (\wedge\{S(v,v') \Rightarrow \sim (S(v',v_1) \wedge ... \wedge S(v_n,v)) : \{v_1,...,v_n\} \subseteq N^0\})$ *is consistent;*

e. $(\wedge G_S) \wedge (\wedge\{S(v,v') \Rightarrow (\vee\{S(v'',v) : v'' \in N^0\}) : v \neq v^0, v,v' \in N^0\})$ *is consistent.*

Proof. Suppose that $G = \wedge(G_C \cup G_P \cup G_S)$ is an axiom for some extensive form game, then it is easy to check that these properties hold. We delay the proof for the fact that G is consistent to theorem 5.1.

Conversely, let (G_C, G_P, G_S) be a given description such that these requirements are satisfied. Define the game $g = [I, (N, Z, \prec), \tau, A, \{t_v\}_{v\in N}, \{\succeq_i\}_{i\in I}]$ as follows: $N = \{v^0\} \cup \{v \in N^0 : S(v^0, v_1), ..., S(v_{n-1}, v_n) \in G_S$ for some $v_1, ..., v_{n-1}, v_n = v \in N^0\}$, $Z = \{z \in Z^0 : d_{\tau(v)}(v,a) > o(z) \in G_C$ for some $v \in N^0, a \in A^0$ and for any other $(v',a') \neq (v,a), d_{\tau(v')}(v',a') > o(z) \notin G_C\}$, $v \prec v'$ if and only if $S(v,v') \in G_S$ for any $v,v' \in N$, $v \prec z$ if and only if $d_{\tau(v)}(v,a) > o(z) \in G_C$ for some $a \in A^0$ for any $v \in N$ and $z \in Z$, $t_v(v') = a$ if and only if $d_{\tau(v)}(v,a) > o(v') \in G_C$, and $z \succeq_i z'$ if and only if $P_i(z,z') \in G_P$ for any $z,z' \in Z$. Finally, we define $t_v(v')$ to be the action $a \in A^0$ such that $d_{\tau(v)}(v,a) > o(v') \in G_C$ for $v \in N, v' \in N \cup Z$.

First we show that $(N, Z, \prec)$ is a tree. For each $v \in N - \{v^0\}$, there exists a node v' in N such that $S(v',v) \in G_C$ by definition of N. Moreover, if there were 2 nodes $v',v'' \in N$ such that $S(v',v), S(v'',v) \in G_C$, then we have

$$\vdash (\wedge G_C) \Rightarrow S(v',v) \wedge S(v'',v),$$

which then implies that

$$\vdash\sim ((\wedge G_C) \wedge \sim (S(v',v) \vee S(v'',v)),$$

a contradiction to property c. Thus, such a node is unique. Now, for any $v \neq v^0$, $v \in N$, there is a sequence $v_1, ..., v_n = v$ such that $S(v^0, v_1), ..., S(v_{n-1}, v) \in G_S$, and this implies that $v_{n-1} \in N$ or $S(v^0, v) \in G_S$. Therefore, for any $v \in N$ other than v^0, there exists some $v' \in N$ such that $v' \prec v$. Moreover, there is no $v \in N$ such that $v \prec v^0$ by (S1).

Also, by definition of Z, there is a unique node $v \in N$ such that $v \prec z$ for any $z \in Z$. Thus, $(N, Z, \prec)$ is a tree. Define $Prec(z) = \{(v, a) \in N^0 \times A^0 : d_{\tau(v)}(v, a) > o(z) \in G_C\}$. If $Prec(z)$ is not a singleton sets, $z = z^*$ by (C1), for any $z \in Z^0$. This also implies that if $z \notin Z \cup \{z^*\}$, then $Prec(z) = \emptyset$ and for all $z \in Z$, $|Prec(z)| = 1$. Moreover, for any $v \in N$, by (C2), there are some $a \in A^0$ and $v' \neq z^*$ such that $d_{\tau(v)}(v, a) > o(v') \in G_C$, and so $v \prec v'$.

We show that for the constructed g, its description is (G_C, G_S, G_P). Clearly, for $v, v' \in N$, $v \prec v'$ if and only if $S(v, v') \in G_S$. Now, let $v \in N^0 - N, v' \in N$. Since, as we have proven, $S(v'', v') \in G_S$ for some $v'' \in N$, and since v'' cannot be the same as v, it follows that $\sim S(v, v') \in G_S$ by (S1). On the other hand, if $S(v', v) \in G_S$, then $v \in N$, a contradiction, and thus $\sim S(v', v) \in G_S$ by (S1). Consider now $v, v' \in N^0 - N$ and suppose that $S(v, v') \in G_S$. Since $v \neq v^0$, there is another node $v_1 \in N^0$ such that $S(v_1, v) \in G_S$ by e. and (S1). To see this, if $S(v'', v) \notin G_S$ for all $v'' \in N^0$, then $\sim S(v'', v) \in G_S$ for all $v'' \in N^0$, i.e.

$$\vdash G_S \Rightarrow (\wedge\{\sim S(v'', v) : v'' \in N^0\}),$$

which implies

$$\sim (G_S \wedge (\vee\{S(v'', v) : v'' \in N^0\})),$$

a contradiction to e.

If $v_1 \in N$, then $v \in N$, impossible. Suppose that we have found $v_n \notin N$ such that $S(v_n, v_{n-1}) \in G_S$, then there exists $v_{n+1} \notin N$ such that $S(v_{n+1}, v_n) \in G_S$ by the same argument. Since N^0 is finite, let v_n, v_{n+k} be the first pair such that $k > 0$, $v_n = v_{n+k}$. Then we have

$$\vdash (\wedge G_S) \Rightarrow S(v_{n+k}, v_{n+k-1}) \wedge .... \wedge S(v_{n+1}, v_n),$$

which implies that

$$\vdash (\wedge G_S) \Rightarrow S(v_n, v_{n+k-1}) \wedge (S(v_n, v_{n+k-1}) \wedge .... \wedge S(v_{n+1}, vn)),$$

and so

$$\vdash \sim (\wedge G_S) \vee (S(v_n, v_{n+k-1}) \wedge S(v_n, v_{n+k-1}) \wedge .... \wedge S(v_{n+1}, v_n)),$$

holds, which is a contradiction to d. Therefore, G_S represents the relation $\prec$ in g.

Consider some $v \in N$. We show that $d_{\tau(v)}(v,a) > o(v') \in G_C$ if and only if $(v,v') \in N \times (N \cup Z)$ and $t_v(v') = a$, or $v' = z^*$ and $a \notin A(v)$. Let $d_{\tau(v)}(v,a) > o(v') \in G_C$. If $v' \in N^0$, then $S(v,v') \in G_S$. By the argument above, we have $v' \in N$ and by (S2) $t_v(v')$ is well defined and it has value a. Now suppose that $v' \in Z^0$ and suppose that $t_v(v')$ is not defined. Then, $Prec(z)$ is not a singleton, and so, as we have seen, $v' = z^*$ and $a \notin A(v)$. On the other hand, if $t_v(v')$ is defined, then $v' \in Z$ and so $t_v(v') = a$.

Conversely, let $(v,v') \in N \times (N \cup Z)$ and $t_v(v') = a$. By definition of t_v, we have $d_{\tau(v)}(v, t_v(v')) > o(v') \in G_C$. On the other hand, let $a \notin A(v)$. By (C2) and (S2), $d_{\tau(v)}(v, t_v(z)) > o(z) \in G_C$ for some $z \in Z^0$. If $|Prec(z)| = 1$, then $z \in Z$ and $t_v(z) = a$, i.e., $a \in A(v)$, a contradiction, and thus, $|Prec(z)| > 1$ and so $z = z^*$.

Suppose that $v \in N^0 - N$. Then, if there is some $(a,v') \in A^0 \times N^0$ such that $d_{\tau(v)}(v,a) > o(v') \in G_C$, then by (S1), $S(v,v') \in G_S$ and so $v \in N$, a contradiction. If there is some $(a,z) \in A^0 \times Z^0$ such that $d_{\tau(v)}(v,a) > o(z) \in G_C$, then either $S(v'',v) \in G_S$ for some $v'' \in N^0$ or $v = v^0$, and both cases imply that $v \in N$, a contradiction. Thus, G_C is derived from g.

It is easy to check that, by definition, G_P is derived from $\{\succeq_i\}_{i \in I}$. We shall show that each $\succeq_i$ is reflexive, transitive, and complete. It is reflexive and complete by (P1) and (P2). Suppose that $z \succeq_i z'$ and $z' \succeq_i z''$, $z, z', z'' \in Z$. Then $P_i(z,z')$ and $P_i(z',z'')$ are in G_P. If $\sim P_i(z,z'') \in G_P$, it is to see that this contradicts with b. Then, by (P2), $P_i(z,z'') \in G_P$, and so $z \succeq_i z''$. □

We end this section with some notations and several lemmas for future reference. Let N' be a subset of N. We define $Sub(v,v') = \vee\{S(v_1,v_2) \wedge ... \wedge S(v_{k-1},v_k) : v = v_1, v_k = v', v_1, ..., v_k \in N^0\}$, and define $Sub(v,N') = (\wedge\{Sub(v,v') : v' \in N'\}) \wedge (\wedge\{\sim Sub(v,v') : v' \notin N'\})$.

Lemma 4.2. *Let G be the axiom of a game g.*

(a) $\vdash G \Rightarrow Sub(v,v')$ if $v,v' \in N$, and $v' \in Sub(v)$.

(b) $\vdash G \Rightarrow \sim Sub(v, v')$ *if* $v, v' \in N$ *and* $v' \notin Sub(v)$*, or if* $v' \in N^0 - N$.

(c) $\vdash G \Rightarrow Sub(v, N')$ *if* $N' = Sub(v)$ *and* $\vdash G \Rightarrow \sim Sub(v, N')$ *otherwise.*

Lemma 4.3. *Let* G *be the axiom of a game* g*. For all* $v \in N$ *and* $a \in A(v)$ *such that* $t_v^{-1}(a) \in Z$*,* $z' \neq t_v^{-1}(a)$ *implies that* $\vdash G \Rightarrow (d_{\tau(v)}(v, a) >\sim o(z'))$.

Proof. Clearly, we have $\vdash G_C \Rightarrow (d_{\tau(v)}(v, a) > o(t_v^{-1}(a))$ and $\vdash G_D \Rightarrow (o(t_v^{-1}(a)) >\sim o(z'))$. Thus, we have $\vdash G \Rightarrow (d_{\tau(v)}(v, a) > o(t_v^{-1}(a)) \wedge (o(t_v^{-1}(a)) >\sim o(z'))$. Now, by (Tran), we have $\vdash (d_{\tau(v)}(v, a) > o(t_v^{-1}(a)) \wedge (o(t_v^{-1}(a)) >\sim o(z')) \Rightarrow (d_{\tau(v)}(v, a) >\sim o(z'))$. Thus, we have $\vdash G \Rightarrow (d_{\tau(v)}(v, a) >\sim o(z'))$. □

Lemma 4.4. *Let* G *be the axiom of a game* g*. Suppose that* $((v_1, a_1), ..., (v_k, a_k), z)$ *is a path from* v_1 *to* $z \in Z$*, then we have*

$\vdash G \wedge (\wedge\{o(v_l) > d_{\tau(v_l)}(v_l, a_l) : l = 2, ..., k\}) \Rightarrow (d_{\tau(v_1)}(v_1, a_1) > o(z)) \wedge (\wedge\{d_{\tau(v_1)}(v_1, a_1) >\sim o(z') : z' \neq z\})$.

Proof. We prove by induction on the size of k. For $k = 1$, it follows directly from the lemma 4.3.

Suppose that it holds for all $l < k$, $k > 1$. Then we have

$$\begin{aligned} \vdash \quad & G \wedge (\wedge\{o(v_l) > d_{\tau(v_l)}(v_l, a_l) : l = 3, ..., k\}) \\ \Rightarrow \quad & (d_{\tau(v_2)}(v_2, a_2) > o(z)) \wedge (\wedge\{d_{\tau(v_2)}(v_2, a_2) > o(z') : z' \neq z\}). \end{aligned}$$

Now, by (Tran), we have

$\vdash (o(v_2) > d_{\tau(v_2)}(v_2, a_2)) \wedge (d_{\tau(v_1)}(v_1, a_1) > o(v_2)) \Rightarrow (d_{\tau(v_1)}(v_1, a_1) > d_{\tau(v_2)}(v_2, a_2))$.

Therefore,

$$\begin{aligned} \vdash \quad & G \wedge (\wedge\{o(v_l) > d_{\tau(v_l)}(v_l, a_l) : l = 2, ..., k\}) \\ \Rightarrow \quad & (d_{\tau(v_1)}(v_1, a_1) > d_{\tau(v_2)}(v_2, a_2)) \wedge (d_{\tau(v_2)}(v_2, a_2) > o(z)) \\ \wedge \quad & (\wedge\{d_{\tau(v_2)}(v_2, a_2) >\sim o(z') : z' \neq z\}). \end{aligned}$$

By (Tran), we have

$$\vdash (d_{\tau(v_1)}(v_1, a_1) > d_{\tau(v_2)}(v_2, a_2)) \wedge (d_{\tau(v_2)}(v_2, a_2) > o(z))$$
$$\wedge(\wedge\{d_{\tau(v_2)}(v_2, a_2) >\sim o(z') : z' \neq z\})$$
$$\Rightarrow (d_{\tau(v_1)}(v_1, a_1) > o(z)) \wedge (\wedge\{d_{\tau(v_1)}(v_1, a_1) >\sim o(z') : z' \neq z\}).$$

Thus, we have

$$\vdash \quad G \wedge (\wedge\{o(v_l) > d_{\tau(v_l)}(v_l, a_l) : l = 2, ..., k\})$$
$$\Rightarrow \quad (d_{\tau(v_1)}(v_1, a_1) > o(z)) \wedge (\wedge\{d_{\tau(v_1)}(v_1, a_1) >\sim o(z') : z' \neq z\}).$$

□

By lemma 4.4, a player i who has firm predictions of future decisions in the sense that i believes that the decisions predicted would be carried out if the decision node is reached could determine the consequence of each action i could take in a particular node.

4.2 Game Theoretical Concepts

We shall formulate the game theoretic concepts with the language, and then formulate the decision criteria for Backward Induction in this subsection. We formulate the concept of sequential best response at a particular node $v \in N^0$ for a given set of prediction of decisions at nodes in a set $N' \subseteq N^0$ in such a way that the formula itself does not depend on the formulation of the game, and the player may be able to learn whether a particular action at v is a sequential best response or not after the player learns the structure of the game and some of other players' decisions.

We shall say that an action a at a particular decision node (weakly) dominates another a' if the consequence a leads to is better the consequence lead by a'. This is formally defined as:

$$Dom_{\tau(v),v}(a, a') =$$
$$\wedge\{(d_{\tau(v)}(v, a) > o(z)) \wedge (d_{\tau(v)}(v, a') > o(z')) \Rightarrow P_{\tau(v)}(z, z') : z, z' \in Z^0\}. \quad (1)$$

We also define strict dominance:

$$SDom_{\tau(v),v}(a,a') = \wedge\{(d_{\tau(v)}(v,a) > o(z)) \wedge (d_{\tau(v)}(v,a') > o(z')) \Rightarrow P_{\tau(v)}(z,z') \wedge \sim P_{\tau(v)}(z',z) : z,z' \in Z^0\}. \quad (2)$$

We can then formulate sequential best responses:

1. Suppose that $N' = \emptyset$. Then

$$Best_{\tau(v),N'}(v,a) = \wedge\{Dom_{\tau(v),v}(a,a') : a' \neq a \in A\}. \quad (3)$$

2. Otherwise, given N' and $s \in A^{N'}$,

$$Best_{\tau(v),N',s}(v,a) = (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v' \in N'\}) \Rightarrow (\wedge\{Dom_{\tau(v),v}(a,a') : a' \neq a \in A\}). \quad (4)$$

We also define strictly best responses as follows:

1. Suppose that $N' = \emptyset$. Then

$$SBest_{\tau(v),N'}(v,a) = \wedge\{SDom_{\tau(v),v}(a,a') : a' \neq a \in A\}. \quad (5)$$

2. Otherwise, given N' and $s \in A^{N'}$,

$$SBest_{\tau(v),N',s}(v,a) = (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v' \in N'\}) \Rightarrow (\wedge\{SDom_{\tau(v),v}(a,a') : a' \neq a \in A\}). \quad (6)$$

The intended meaning of $Best_{\tau(v),N',s}(v,a)$ is that a is a sequential best response at v, given that $s \in A^{N'}$ will be played at all nodes in N', and N' should be the set of nodes succeeding v. Before learning the structure of the game, the player is unaware of what N' will be, and our approach is able to explicitly give the arguments for this thought process. Moreover, the player will not be able to infer which action is a best response at v unless the player has predictions of others' decisions at all nodes in N'. We remark that this formula could be adequate for expressing the idea of sequential best responses in multi-stage games, with some modification. Here we list some intuitive properties of $SBest$ and $Best$:

Lemma 4.5. *(a) For any $a, a' \in A$, we have*

$$\vdash SDom(a, a') \Rightarrow Dom(a, a').$$

(b) For any $v \in N^0$, $a \in A$, and for any $N' \subseteq N^0$ and $s \in A^{N'}$, we have

$$\vdash SBest_{\tau(v),N',s}(v, a) \Rightarrow Best_{\tau(v),N',s}(v, a).$$

We now turn to the decision criteria that lead to Backward Induction decisions. The first criterion requires that, at each node v, if the decision maker i has predictions of all the decisions at all nodes succeeding v, then i should decide to take action a if a is a strictly sequential best responses with respect to i's predictions. The second principle demands any decision to take an action a to be a sequential best response with respect to i's prediction. These two criteria correspond to PC_v and RC_v, respectively. We shall denote the predictions $B_i(D_{\tau(v)}(v, a) \wedge (\wedge\{\sim D_{\tau(v)}(v, a') : a' \neq a, a' \in A\}))$ of decisions at v as $PD_i(v, a)$, and we shall denote $\wedge\{PD_i(v, s(v)) : v \in N'\}$ as $PD_i(N', s)$, for any $s \in A^{N'}$.

We give the formal formulation as follows:

a. Rationality Criteria: $(v \in N^0)$

a.1 Suppose that $N' = \emptyset$. Then

$$RC_{v,N'} = \wedge\{D_{\tau(v)}(v, a) \wedge Sub(v, N') \Rightarrow \quad Best_{\tau(v),N'}(v, a) : a \in A\}; \tag{7}$$

a.2 Suppose that $N' \neq \emptyset$. Then

$$\begin{aligned} RC_{v,N'} = \wedge\{&D_{\tau(v)}(v, a) \wedge Sub(v, N') \wedge PD_{\tau(v)}(N', s) \\ &\Rightarrow B_{\tau(v)}(Best_{\tau(v),N',s}(v, a)) : a \in A,\ s \in A^{N'}\}; \end{aligned} \tag{8}$$

a.3 Define

$$RC_v = \wedge\{RC_{v,N'} : N' \subseteq N\}. \tag{9}$$

This criterion requires players' decisions be consistent with their preference. If in situation v, every action will lead to final results, then any plan should be a best response.

If the result of an action is contingent on future actions, and if the player has beliefs on these future actions, the plan of the player is required to be a consciously best response given these future actions.

b. Preference Criteria: $(v \in N^0)$

b.1 Suppose that $N' = \emptyset$. Then

$$PC_{v,N'} = \wedge\{SBest_{\tau(v),N'}(v,a) \wedge Sub(v,N') \Rightarrow \quad D_{\tau(v)}(v,a) : a \in A\}; \tag{10}$$

b.2 Suppose that $N' \neq \emptyset$. Then

$$\begin{aligned} PC_{v,N'} = \wedge\{Sub(v,N') \wedge PD_{\tau(v)}(N',s) \wedge B_{\tau(v)}(SBest_{\tau(v),N',s}(v,a)) \\ \Rightarrow D_{\tau(v)}(v,a) : a \in A,\ s \in A^{N'}\}; \end{aligned} \tag{11}$$

b.3 Define

$$PC_v = \wedge\{PC_{v,N'} : N' \subseteq N\}. \tag{12}$$

This criterion requires players' decision to be determined by their preferences given their beliefs, if the preferences are conclusive. If in situation v, every action will lead to final results, then the player should decide to take the action that leads to the most preferred result. If the player is aware of future actions relevant to current actions and is also aware of a most preferred action given these future actions, then the player should take the most preferred action.

c. The compound decision criterion at $v \in N^0$:

$$DC_v = (Sub(v^0,v) \Rightarrow RC_v \wedge PC_v), \tag{13}$$

and

$$DC_{v^0} = (RC_{v^0} \wedge PC_{v^0}),$$

and the compound decision criterion for the game:

$$DC = (\wedge_{v \in N^0} DC_v). \tag{14}$$

These criteria are formulated without any reference a particular game. To apply these criteria in a particular situation v, DC_v require the players to follow these decision criteria only if v is a situation where decision is necessary. The sentence $Sub(v^0, v)$ is equivalent to say that v is a decision node in the game for player $\tau(v)$. To say that a player is rational in the traditional sense can be replaced by the assumption that the player follows DC. With this definition, being rational can be an attribute of the combination of a player and a situation. It is not necessary for a player to follow the same decision criterion in all situations, and sometimes it is commonly known when players like to adopt criteria like DC in certain situations. Thus, this may help to identify the scope of the application of Backward Induction and many other Game-theoretic solution concepts.

5 Backward Induction and Subgame Perfectness

In this section we present the main results for Backward Induction decisions and subgame perfectness. Before we analyze the consequences of these criteria and beliefs of them, we first argue that they are consistent so that our inference will not be trivial. We shall define, for any formula φ in L,

$$B^1(\varphi) = \wedge\{B_i(\varphi) : i \in I\},$$

$$B^1_{-i}(\varphi) = \wedge\{B_j(\varphi) : j \in I, j \neq i\},$$

and define

$$B^k_i(\varphi) = B_i(B^{k-1}_{-i}(\varphi)),$$

$$B^k_{-i}(\varphi) = \wedge\{B^k_j(\varphi) : j \in I, j \neq i\},$$

and

$$B^k(\varphi) = \wedge\{B^k_i(\varphi) : i \in I\}$$

for $k > 1$. We shall also use $B^0\varphi$ to denote φ.

We shall also define, for any strategy profile s of g, the formula that all decisions follow s as follows:

$$D_g(s) = (\wedge\{D_{\tau(v)}(v, s(v)) : v \in N\}) \wedge (\wedge\{\sim D_{\tau(v)}(v, a) : v \in N, a \neq s(v)\}).$$

Theorem 5.1. *Let G be the axiom for an extensive for game with perfect information g. The formula $\wedge\{B^k(DC \wedge G \wedge D_g(s)) : k = 0, ..., K\}$ is consistent for any extensive game g and any subgame perfect equilibrium s for g, for any $K \in \mathbb{N}$.*

Proof. See the appendix. □

By this theorem, these decision criteria and common belief of them are consistent. Moreover, it states that for any subgame perfect equilibrium s, these decision criteria are consistent with common belief (by this we mean arbitrary high order of mutual belief) of decisions following s. We shall show that these criteria lead to Backward Induction decisions if there is a unique subgame perfect equilibrium. Unlike many other papers on Backward Induction, a strategy is not a primitive element in our model. Players may not have any decision at some node v, and this can be expressed as $\wedge\{\sim D_{\tau(v)}(v, a) : a \in A\}$. All their decisions, in our model, are derived from their decision criteria and their predictions on other players' decisions.

We define $dep(v) = \max\{k : \text{for some } v_1, ..., v_k,\ v \prec v_1 \prec ... \prec v_k \in Z\}$ to be the length of the game beginning from v, which turns out to be the exact measurement of complexity of the mutual belief sufficient and necessary for Backward Induction decisions.

Theorem 5.2. *Given a game g and its axiom G, Suppose that $dep(v^0) = K$ and suppose that g has a unique subgame perfect equilibrium s. Then*

$$\vdash (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-1\}) \Rightarrow D_g(s).$$

Proof. We claim that for any $v \in N$ and the sub-game $g(v)$,

$$\vdash (\wedge\{B^k(DC \wedge G) : k = 0, ..., dep(v)-1\}) \Rightarrow D_{g(v)}(s|_v).$$

We prove by induction on the depth of v. Define $z(v)$ be such that $Z(v, s) = \{z(v)\}$ for any $v \in N$.

Suppose that $dep(v) = 1$. Then for any $a \in A$,

$$\vdash (PC_v \wedge RC_v \wedge Sub(v, \emptyset))$$
$$\Rightarrow (D_{\tau(v)}(v, a) \Rightarrow Best_{\tau(v),\emptyset}(v, a)) \wedge (SBest_{\tau(v),\emptyset}(v, a) \Rightarrow D_{\tau(v)}(v, a)),$$

and

$$\vdash G \Rightarrow Sub(v, \emptyset) \wedge \{\sim Sub(v, N') : N' \subseteq N^0,\ N' \neq \emptyset\}.$$

Thus, for any $a \in A$,

$$\vdash G \wedge PC_v \wedge RC_v \Rightarrow$$
$$(D_{\tau(v)}(v, a) \Rightarrow Best_{\tau(v),\emptyset}(v, a)) \wedge (SBest_{\tau(v),\emptyset}(v, a) \Rightarrow D_{\tau(v)}(v, a)). \tag{15}$$

By lemma 4.3 and (Tran), for all $a \in A$,

$$\vdash \wedge G_C \Rightarrow$$
$$(d_{\tau(v)}(v, a) > o(t_v^{-1}(a))) \wedge (\wedge\{\sim (d_{\tau(v)}(v, a) > o(z)) : z \neq t_v^{-1}(a)), z \in Z^0\}). \tag{16}$$

Since s is the unique subgame perfect equilibrium, it follows that $z(v) \succeq_{\tau(v)} t_v^{-1}(a)$ and $t_v^{-1}(a) \not\succeq_{\tau(v)} z(v)$ for any $a \in A(v)$, $a \neq s(v)$. Thus, we have

$$\vdash \wedge G_P \Rightarrow (\wedge\{P_{\tau(v)}(z(v), t_v^{-1}(a)) \wedge \sim P_{\tau(v)}(t_v^{-1}(a), z(v)) : a \neq s(v), a \in A\}). \tag{17}$$

(We take $t_v^{-1}(a) = z^*$ for any $a \notin A(v)$.)

Combining (16) and (17), we have

$$\vdash G \Rightarrow (\wedge\{SDom(s(v), a) \wedge \sim Dom(a, s(v)) : a \neq s(v), a \in A\}).$$

Thus, we have

$$\vdash G \Rightarrow SBest_{\tau(v),\emptyset}(v, s(v)) \wedge (\wedge\{\sim Best_{\tau(v),\emptyset}(v, a) : a \neq s(v), a \in A\}). \tag{18}$$

Moreover, since $v \in N$, $\vdash G \Rightarrow Sub(v^0, v)$ or $v = v^0$, and so $\vdash G \wedge DC_v \Rightarrow PC_v \wedge RC_v$.

Combining (15) and (18), we have

$$\vdash G \wedge DC \Rightarrow$$
$$D_{\tau(v)}(v, s(v)) \wedge (\wedge\{\sim D_{\tau(v)}(v, a) : a \neq s(v), a \in A\}). \tag{19}$$

Suppose that $dep(v) = K$ and suppose that for any $v' \in Sub(v)$, $(K \geq 2)$

$$\vdash (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\}) \Rightarrow D_{g(v')}(s|_{v'}). \tag{20}$$

Then, assuming that $\tau(v) = i$, by (NEC) and (K),

$$\vdash B_i(\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\})$$
$$\Rightarrow (\wedge\{PD_i(v', s(v')) : v' \in Sub(v)\}).$$

Therefore, by axiom (4), we have

$$\vdash (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-1\})$$
$$\Rightarrow (\wedge\{PD_i(v', s(v')) : v' \in Sub(v)\}). \quad (21)$$

Moreover, if we let $z(a)$ be the result of playing a at v, $a \in A$ ($z(a) = z^*$ if $a \notin A(v)$), given everyone follows s in every $v' \in Sub(v)$, i.e., $z(a) \in Z(t_v^{-1}(a); s)$, then we have, by lemma 4.4,

$$\vdash G \wedge (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v' \in Sub(v)\})$$
$$\Rightarrow (d_{\tau(v)}(v, a) > o(z(a))) \wedge (\wedge\{\sim (d_{\tau(v)}(v, a) > o(z')) : z' \neq z(a), z' \in Z^0\}). \quad (22)$$

Since s is the unique subgame perfect equilibrium, we have that $z(s(v)) \succeq_{\tau(v)} z(a')$ and $z(a') \not\succeq_{\tau(v)} z(s(v))$ for any $a' \neq s(v)$, $a' \in A(v)$. Thus, we have (notice that $P_i(z, z^*)$, $\sim P_i(z^*, z) \in G_P$ for any $z \in Z$)

$$\vdash \wedge G_P \Rightarrow (\wedge\{P_i(z(s(v)), z(a)) \wedge \sim P_i(z(a), z(s(v))) : a \neq s(v), a \in A\}). \quad (23)$$

We have for any $a \in A$,

$$\vdash G \wedge (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v' \in Sub(v)\})$$
$$\Rightarrow (SBest_{i,Sub(v),s|_v}(v, a) \Leftrightarrow (\wedge\{SDom_{i,v}(a, a') : a' \neq a, a' \in A\}))$$
$$\wedge(Best_{i,Sub(v),s|_v}(v, a) \Leftrightarrow (\wedge\{Dom_{i,v}(a, a') : a' \neq a, a' \in A\})). \quad (24)$$

Combining (22) and (23),

$$\vdash G \wedge \{o(v') > d_{\tau(v')}(v', s(v')) : v' \in Sub(v)\}$$
$$\Rightarrow (\wedge\{SDom_{i,v}(s(v), a) : a \neq s(v), a \in A\})$$
$$\wedge(\wedge\{\sim Dom_{i,v}(a, s(v)) : a \neq s(v), a \in A\}). \quad (25)$$

Combining (24) and (25),

$$\vdash G \wedge (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v' \in Sub(v)\})$$
$$\Rightarrow (SBest_{i,Sub(v),s|_v}(v, s(v)) \wedge (\wedge\{\sim Best_{i,Sub(v),s|_v}(v, a) : a \neq s(v), a \in A\})).$$

By (NEC), (K), (D), and lemma 3.1,

$$\vdash B_i(G) \wedge (\wedge\{B_i(o(v') > d_{\tau(v')}(v', s(v'))) : v' \in Sub(v)\})$$
$$\Rightarrow (B_i(SBest_{i,Sub(v),s|_v}(v, s(v)))$$
$$\wedge(\wedge\{\sim B_i(Best_{i,Sub(v),s|_v}(v, a)) : a \neq s(v), a \in A\})). \tag{26}$$

Now, we shall apply the decision criterion DC_v. Since $v \in N$, we have

$$\vdash G \Rightarrow Sub(v^0, v)$$

or $v = v^0$, and so by lemma 3.2,

$$\vdash G \Rightarrow (DC_v \Rightarrow PC_v \wedge RC_v).$$

By lemma 3.2,

$$\vdash \quad PC_v \wedge RC_v \wedge B_i(SBest_{i,Sub(v),s|_{Sub(v)}}(v, s(v)))$$
$$\wedge(\wedge\{\sim B_i(Best_{i,Sub(v),s|_v}(v, a)) : a \neq s(v), a \in A\})$$
$$\wedge Sub(v, Sub(v)) \wedge PD_i(Sub(v), s|_{Sub(v)})$$
$$\Rightarrow D_i(v, s(v)) \wedge (\wedge\{\sim D_i(v, a) : a \neq s(v), s \in A\}). \tag{27}$$

Clearly we have

$$\vdash G \Rightarrow Sub(v, Sub(v)).$$

Thus, combining these results with (26) and (27), it follows that

$$\vdash G \wedge DC_v \wedge PD_i(Sub(v), s|_v) \wedge B_i(G)$$
$$\wedge(\wedge\{B_i(o(v') > d_{\tau(v')}(v', s(v'))) : v' \in Sub(v)\})$$
$$\Rightarrow D_i(v, s(v)) \wedge (\wedge\{\sim D_i(v, a) : a \neq s(v), s \in A\}). \tag{28}$$

Now, by definition:

$$\vdash PD_i(Sub(v), s|_v)$$
$$\Rightarrow (\wedge\{B_i((o(v') > d_{\tau(v')}(v', s(v'))) : v' \in Sub(v)\}). \tag{29}$$

Combining (21), (28), and (29),

$$\vdash DC_v \wedge B_i(G)$$
$$\wedge(\wedge\{B^k(DC \wedge G) : k = 0, ..., K-1\})$$
$$\Rightarrow D_i(v, s(v)) \wedge (\wedge\{\sim D_i(v, a) : a \neq s(v), s \in A\}). \tag{30}$$

Notice that, by definition and the the tautology $\varphi \wedge \chi \Rightarrow \varphi$,

$$\vdash (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-1\}) \Rightarrow DC_v \wedge B_i(G),$$

it follows that, combining with (20) and (30),

$$\vdash (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-1\}) \Rightarrow D_{g(v)}(s|_v).$$

□

A corollary of this theorem is that common belief of rationality is not necessary for Backward Induction decisions, and similar results have been reported in Clausing[(6)] and Balkenborg and Winter[(3)]. Furthermore, as another corollary, common belief of the game structure is not necessary either. This gives us a set of sufficient epistemic conditions and decision criteria for Backward Induction decisions. The following theorem states that these conditions are also essentially necessary for general extensive form games with perfect information. This result, to the best of our knowledge, is novel to the literature.

Theorem 5.3. *Suppose that $dep(v_0) = K \geq 2$. Then*

$$(\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\}) \wedge (\wedge\{\sim D_{\tau(v_0)}(v_0, a) : a \in A\})$$

is a consistent formula.

Proof. See the appendix. □

Corollary 5.1. *Suppose that $dep(v^0) = K \geq 2$ and suppose that s is a subgame perfect equilibrium for g. Then*

$$\nvDash (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\}) \Rightarrow D_g(s).$$

The corollary of the theorem states that any lower order of mutual belief of the game and of every player following DC than the length of the game minus 1 would not entail any decision for the player in the initial node even if the player follows DC. This states that without sufficient epistemic conditions, the decision criteria DC are not able to entail a particular decision. However, this does not exclude the possibility of making decisions when the player use DC together with others. For example, if the player decide to use a strategy if it is dominant in the normal from representation of the game, then it is easy to imagine a situation where this rule entails a definite decision given that the player knows the game but knows nothing more. The story behind theorem 5.2 and theorem 5.3 are not about equilibrium strategies; the epistemic conditions may be satisfied because of learning or judgement, but no belief of equilibrium play is assumed in deriving these results.

One may also wonder if we impose stronger assumptions on the content of the belief, whether we could lower the complexity of mutual belief. The following theorem states that if there is mutual knowledge of following a particular strategy profile and the game structure, and if every player follows DC, then the strategy profile must be a subgame perfect equilibrium. We shall give a lemma before we present the theorem.

Lemma 5.1. *Let (G_C, G_S, G_P) be a description that describes g. If there is model M and a world w in it such that*

$$(M, w) \models G \wedge Best_{\tau(v),Sub(v),s}(v, a) \wedge (\wedge\{D_{\tau(v')}(v', s(v')) : v' \in Sub(v)\}),$$

then a is a sequential best response at v w.r.t. s.

Proof. By the definition of $Best_{\tau(v),Sub(v),s}(v, a)$, we also have that

$$(M, w) \models G \wedge Dom_{\tau(v),v}(v, a).$$

Since $(M, w) \models G \wedge (\wedge\{D_{\tau(v')}(v', s(v')) : v' \in Sub(v)\})$, by lemma 4.4, we have $(M, w) \models d_{\tau(v)}(v, a') > o(z(a'))$ for each $a' \in A^0$, where $z(a') = Z(a'; s')$, s' is any extension of s to g. Let a' be any action in A^0 other than a. Then $(M, w) \models (d_{\tau(v)}(v, a') > o(z(a'))) \wedge (d_{\tau(v)}(v, a) > o(z(a)))$. Since $(M, w) \models Dom_{\tau(v),v}(v, a)$, it follows that $(M, w) \models P_{\tau(v)}(z(a), z(a'))$. Since $(M, w) \models G_P$, $(M, w) \models P_{\tau(v)}(z(a), z(a'))$ if and only if $z(a) \succeq_{\tau(v)} z(a')$ or $a' \notin A(v)$. Thus, a is a sequential best response at v w.r.t. s in g. □

Theorem 5.4. *Given a game description (G_C, G_S, G_P), the formula*

$$B^1(G \wedge D_g(s)) \wedge (DC \wedge G \wedge D_g(s))$$

is consistent if and only if s is a subgame perfect equilibrium in the game g described by (G_C, G_S, G_P).

Proof. The 'if' part is proved in theorem 5.1. Suppose that $B^1(DC \wedge G \wedge D_g(s)) \wedge (DC \wedge G \wedge D_g(s))$ is a consistent formula. Then by theorem 3.1, there is a model M and a world w in it such that $(M, w) \models B^1(D_g(s)) \wedge (DC \wedge G \wedge D_g(s))$. Let g be the game described by (G_C, G_S, G_P).

We shall show that $s(v)$ is a best move w.r.t. $s|_v$ at every node $v \in N$. Let $v \in N$ be a node such that $Sub(v) = \emptyset$. Then $(M, w) \models DC \wedge G \wedge D_g(s)$. This implies that $(M, w) \models G \wedge D_{\tau(v)}(v, s(v)) \wedge RC_{v,\emptyset}$, which in turn implies that $(M, w) \models G \wedge Best_{\tau(v),\emptyset}(v, s(v))$. By lemma 5.1, $s(v)$ is a sequential best response at v.

Consider any $v \in N$ such that $Sub(v) \neq \emptyset$. Then

$$(M, w) \models PD_{\tau(v)}(Sub(v), s|v) \wedge Sub(v, Sub(v)) \wedge D_{\tau(v)}(v, s(v)) \wedge RC_{v,Sub(v)}.$$

Thus, we have

$$(M, w) \models B_{\tau(v)}(Best_{\tau(v),Sub(v),s|v}(v, s(v))).$$

Let $w_1 \in R_{\tau(v)}(w)$, then

$$(M, w_1) \models Best_{\tau(v),Sub(v),s|v}(v, s(v)) \wedge G \wedge (\wedge\{D_{\tau(v')}(v', s(v')) : v' \in Sub(v)\}).$$

By lemma 5.1, $s(v)$ is a sequential best response at v w.r.t. s. □

Corollary 5.2. *Let s be a strategy profile for a perfect information game g that is not a subgame perfect equilibrium. Let G be its axiom. Then we have*

$$\vdash \sim (B^1(G \wedge D_g(s))) \vee \sim (DC \wedge G \wedge D_g(s)).$$

Proof. By theorem 5.4,

$$B^1(G \wedge D_g(s)) \wedge (DC \wedge G \wedge D_g(s))$$

is not consistent. Thus, its negation is provable. □

We interpret an equilibrium strategy as a strategy that is expected to be followed by all the players. With this interpretation, the decision criteria DC refines the equilibrium to satisfy subgame perfectness. Theorem 5.4 shows any equilibrium strategy in an environment with mutual knowledge of the structure of the game and every player following DC implies that the equilibrium satisfies subgame perfectness. The condition here is similar to the epistemic conditions for Nash equilibrium in Aumann and Brandenburger[(2)], especially the preliminary observation there, but we do not consider mixed strategies here. This result is also tight, since by Theorem 5.3, for any game g with $dep(v^0) > 1$, $DC \wedge G$ is consistent with any decision at v^0.

6 Conclusion

In this paper we develop an epistemic logic with causality and prove its completeness and soundness. We use this logic to formalize the notion of the structure of a game, and give a set of decision criteria that entail Backward Induction decisions, given sufficient epistemic conditions. We also prove that the set of epistemic conditions provided are also necessary. Also, we provide some epistemic conditions for subgame perfect equilibrium. We give a set of precise conditions, which may be tested empirically in many real applications. However, one more possibility, that people do not possess sufficient inference ability, is not investigated here. We conjecture that it is possible to estimate the complexity of the inferences for Backward Induction decisions.

Another line of extension may be the investigation of multi-stage games. We conjecture that a generalization of Theorem 5.4 should hold for general multi-stage games. Also, we have not touched the problem of belief revision in extensive form games, which is extensively discussed in the literature. Our result suggests that if an agent who has planned to play the Backward Induction actions observe an off-equilibrium action can infer that at least one condition we give in Theorem 5.2 is wrong. However, there seems no obvious way to say which one is it. Our approach, nonetheless, provides a framework in which one could model the situation where the fault has to do with belief in the structure of the game.

7 Appendix

We provide proofs of several theorems in the appendix.

Before the proof of theorem 3.1, we give a lemma first, and its proof can be found in standard textbook for Modal logic.

Lemma 7.1. *Let* Γ *be a maximal consistent set.*

a. For each $\varphi \in L$*, either* $\varphi \in \Gamma$ *or* $\sim \varphi \in \Gamma$*, but not both.*

b. $\varphi \wedge \chi \in \Gamma$ *if and only if* $\varphi \in \Gamma$ *and* $\chi \in \Gamma$.

c. $\vdash \varphi$ *implies* $\varphi \in \Gamma$.

d. $\varphi \Rightarrow \chi \in \Gamma$ *implies that if* $\varphi \in \Gamma$, $\chi \in \Gamma$.

e. If for all maximal consistent Γ, $\varphi \in \Gamma$ *implies* $\chi \in \Gamma$*, then* $\vdash \varphi \Rightarrow \chi$.

Proof of theorem 3.1:

Proof. ($\Leftarrow$) Suppose that $\{\varphi_1, ..., \varphi_k\}$ is a proof for φ. We show by induction that for any $M \in \mathfrak{M}$ and any $w \in W$, $(M, w) \models \varphi_j$, $j = 1, .., k$.

a. $j = 1$. Then φ_1 is an instance of the axiom schemes.

a.1 (PC1-3) φ_1 is an instance of PC tautology. Then $M \models \varphi_1$ by lemma 3.4.

a.2 (Tran) $\varphi_1 = (\zeta > \xi) \wedge (\xi > \eta) \Rightarrow (\zeta > \eta)$. Suppose that $(M, w) \models (\zeta > \xi)$ and $(M, w) \models (\xi > \eta)$. Then $f([\zeta], w) \subseteq [\xi]$ and $f([\xi], w) \subseteq [\eta]$. By (s-Tran), we have $f([\zeta], w) \subseteq [\eta]$ and so $(M, w) \models (\zeta > \eta)$.

a.3 (K) $\varphi_1 = (B_i(\varphi) \wedge B_i(\varphi \Rightarrow \chi)) \Rightarrow B_i(\chi)$. Suppose that $(M, w) \models B_i(\varphi)$ and $(M, w) \models B_i(\varphi \Rightarrow \chi)$, i.e., $R_i(w) \subseteq [\varphi] \cap [\varphi \Rightarrow \chi]$. Now, $(M, w') \models \varphi$ and $(M, w') \models \varphi \Rightarrow \chi$ imply that $(M, w') \models \chi$. Thus, $[\varphi] \cap [\varphi \Rightarrow \chi] \subseteq [\chi]$ and hence $(M, w) \models B_i(\chi)$.

a.4 (D) $\varphi_1 =\sim B_i(\varphi \wedge \sim \varphi)$. There exists some $w' \in R_i(w)$ by (s-D). Then $w' \notin [\varphi] \cap [\sim \varphi] = \emptyset$.

a.5 (4) $\varphi_1 = B_i(\varphi) \Rightarrow B_i(B_i(\varphi))$. Suppose that $(M, w) \models B_i(\varphi)$. Then $R_i(w) \subseteq [\varphi]$. Let $w' \in R_i(w)$, then $w'' \in R_i(w')$ implies that $w'' \in R_i(w)$. Therefore, $w'' \in [\varphi]$. Thus, $R_i(w') \subseteq [\varphi]$, and so $R_i(w) \subseteq [B_i(\varphi)]$.

a.6 (Det) $\varphi_1 = (\zeta >\sim \varphi) \Rightarrow\sim (\zeta > \varphi)$. Suppose that $(M, w) \models (\zeta >\sim \varphi)$. Then, there exists some $w' \in f([\zeta], w) \subseteq W - [\varphi]$ since $f([\zeta], w) \neq \emptyset$ by (s-Det). Thus, $f([\zeta], w) \nsubseteq [\varphi]$ and so $(M, w) \not\models (\zeta > \varphi)$.

b. Suppose that $(M, w) \models \varphi_j$ for all $j = 1, .., k-1$. Then:

b.1 φ_k is an instance of the axiom schemes. This case is covered by a.

b.2 (MP) φ_k is a result of MP from $\varphi_l, \varphi_{l'} = \varphi_l \Rightarrow \varphi_k$, $l, l' < k$. By induction hypothesis, $\models \varphi_l$ and $\models \varphi_l \Rightarrow \varphi_k$. Thus, we have $\models \varphi_k$.

b.3 (NEC) φ_k is a result of NEC from φ_l and $\varphi_k = B_i(\varphi_l)$. By the induction hypothesis, $\models \varphi_l$ and so $[\varphi_l] = W$ for any model M. Thus, $R_i(w) \subseteq [\varphi_l]$.

b.4 (RCEA) φ_k is a result of RCEA from $\varphi_l = \zeta \Leftrightarrow \xi$ and $\varphi_k = (\zeta > \varphi) \Leftrightarrow (\xi > \varphi)$. By the induction hypothesis, $\models \varphi_l$ and so $[\zeta] = [\xi]$. Thus, $f([\zeta], w) = f([\xi], w)$, and so $[(\zeta > \varphi)] = [(\xi > \varphi)]$.

b.5 (RCK) φ_k is a result of RCEA from $\varphi_l = \varphi_1 \wedge ... \wedge \varphi_m \Rightarrow \chi$ and $\varphi_k = ((\zeta > \varphi_1) \wedge ... \wedge (\zeta > \varphi_m)) \Rightarrow (\zeta > \chi)$. By the induction hypothesis, $\models_K \varphi_l$ and so $[\varphi_1] \cap ... \cap [\varphi_m] \subseteq [\chi]$.

Now, if $w \in [(\zeta > \varphi_1) \wedge ... \wedge (\zeta > \varphi_k)]$, then $f([\zeta], w) \subseteq [\varphi_1] \cap ... \cap [\varphi_m] \subseteq [\chi]$ and so $w \in [(\zeta > \chi)]$.

b.6 (CN) φ_k is a result of CN from φ_l and $\varphi_k = (\zeta > \varphi_l)$. By the induction hypothesis, $\models \varphi_l$ and so $[\varphi_l] = W$ for any model M. Thus, $f([\zeta], w) \subseteq [\varphi_l]$.

Thus, by mathematical induction, any provable formula φ is L is also valid.

($\Rightarrow$) It suffices to show that, for every consistent formula φ, there is a model M and a state w such that $(M, w) \models \varphi$. This proves that $\nvdash\sim \varphi$ implies $\not\models\sim \varphi$. For any $\chi \in L$, we have $\models \chi$ if and only if $\models\sim\sim \chi$, and $\vdash \chi$ if and only if $\vdash\sim\sim \chi$. Thus, if we have $\models \chi$, it follows that $\models\sim (\sim \chi)$, and so we have $\vdash\sim (\sim \chi)$, which implies $\vdash \chi$.

We shall construct a canonical model M such that if φ is consistent, there is a state w such that $(M, w) \models \varphi$. First we shall enumerate formulas in L, and then we shall show that, given a consistent formula, the set of maximal consistent formulas containing it exists. Since every atomic formula is consistent, it follows that the collection of maximal consistent sets is not empty.

Since the set L is denumerable, let $\varphi_1, ..., \varphi_k, ...$ be a enumeration. We claim that if $p \in \Phi$, then there is a maximal consistent set $\mathfrak{F}$ that contains p. This fact is easily verifiable using the usual techniques. Notice that this works for any consistent formula $\varphi \in L$ instead of p, and hence there is always a maximal consistent set that contains a particular consistent formula φ.

Now, if we define $W = \{w \subseteq L : w \text{ is a maximal consistent subset of } L\}$ and define Σ be the smallest algebra generated by $\{|p| \subseteq W : |p| = \{w \in W : p \in w\}, p \in \Phi\}$, then we have that for any $A \in \Sigma$, there is a formula $\zeta \in L^0$ such that $A = |\zeta| = \{w \in W : \zeta \in w\}$. To see this, any $A \in \Sigma$ can be expressed as $\cup_{k=1}^{m} \cap_{l=1}^{m_k} A_{kl}$, where A_{kl} is of the form $|p_{kl}|$ or $|\sim p_{kl}|$ for some $p_{kl} \in \Phi$. Let η_{kl} be p_{kl} if $A_{kl} = |p_{kl}|$ and let η_{kl} be $\sim p_{kl}$ otherwise. It is easy to check that $A = | \vee_{k=1}^{m} (\wedge_{l=1}^{m_k} \eta_{kl})|$. Clearly, for any $\zeta \in L^0$, $|\zeta| \in \Sigma$.

Define $M =< W, f, R_1, ..., R_n, V >$ as follows:

a. $W = \{w \subseteq L : w \text{ is a maximal consistent subset of } L.\}$.

b. Σ is the smallest algebra generated by $\{|p| \subseteq W : |p| = \{w \in W : p \in w\}, p \in \Phi\}$.

c. for each $w \in W$, define $w/\zeta = \{\varphi \in L : (\zeta > \varphi) \in w\}$ and then define $f(|\zeta|, w) = \{w' \in W : w/\zeta \subseteq w'\}$;

(Since every set in Σ can be represented as $|\eta|$ for some $\eta \in L^0$, we identify a set in Σ with its representation. Notice that by (RCEA), $\vdash \zeta \Leftrightarrow \xi$ implies that $\vdash (\zeta > \varphi) \Leftrightarrow (\xi > \varphi)$ and so if $\vdash \zeta \Leftrightarrow \xi$, $w/\zeta = w/\xi$. Since $\vdash \zeta \Leftrightarrow \xi$ implies that $|\zeta| = |\xi|$, f is well-defined.)

d. for each $w \in W$, define $w/B_i = \{\varphi \in L : B_i(\varphi) \in w\}$, and then define $R_i(w) = \{w' \in W : w/B_i \subseteq w'\}, i = 1, .., n$;

e. for each $p \in \Phi$ and $w \in W$, $V(p, w) = \top$ if and only if $p \in w$.

We define an operator $[\,] : L \rightarrow 2^W$ as before such that for all $\varphi \in L$, $V(\varphi, w) = \top$ if and only if $w \in [\varphi]$. Define $|\varphi| = \{w \in W : \varphi \in w\}$, for all $\varphi \in L$.

First we show that for all $\varphi \in L$, $|\varphi| = [\varphi]$. We will show by induction on the structure of φ.

a. Let $p \in \Phi$. Then, $|p| = [p]$ by definition.

b. We then show that $|\zeta| = [\zeta]$, for all $\zeta \in L^0$ by induction.

b.1 By lemma 3.4, $w \in [\varphi \wedge \chi] = [\varphi] \cap [\chi] = |\varphi| \cap |\chi|$ if and only if both φ and χ belong to w if and only if $\varphi \wedge \chi \in w$ if and only if $w \in |\varphi \wedge \chi|$, for all φ, $\chi \in L^0$.

b.2 By lemma 3.4, $w \in [\sim \varphi] = W - [\varphi] = W - |\varphi|$ if and only if $\varphi \notin w$ if and only if $\sim \varphi \in w$, for all $\varphi \in L^0$.

c. Suppose that $|\varphi| = [\varphi]$, $|\chi| = [\chi]$, and $|\zeta| = [\zeta]$, for all φ, $\chi \in L^{k-1}$, and $\zeta \in L^0$.

c.1 Let $\varphi \in L^{k-1}$. We show that $|B_i(\varphi)| = [B_i(\varphi)]$ and $|(\zeta > \varphi)| = [(\zeta > \varphi)]$, $\zeta \in L^0$.

c.1.1 Suppose that $w \in [B_i(\varphi)]$. If w/B_i is consistent, then $R_i(w) \subseteq [\varphi] = |\varphi|$ and so $w/B_i \cup \{\sim \varphi\}$ is inconsistent. (for otherwise, there exists $w' \in W$ and $w/B_i \cup \{\sim \varphi\} \subseteq w'$, which contradicts to the fact that $R_i(w) \subseteq |\varphi|$.) Therefore, either w/B_i being consistent or not, there is a subset $\{\chi_1, ..., \chi_k\} \subseteq w/B_i$ such that $\vdash \sim (\chi_1 \wedge ... \wedge \chi_k) \vee \varphi$. (Notice that k may be 0, and $\{\sim \varphi\}$ is inconsistent, i.e., $\vdash \sim\sim \varphi$) Then by (NEC) and (K), we have

$\vdash (B_i(\chi_1) \wedge ... \wedge B_i(\chi_k)) \Rightarrow B_i(\varphi)$. Thus, $B_i(\varphi) \in w$.

c.1.2 Suppose that $w \in |B_i(\varphi)|$. Then $\varphi \in w/B_i$ and so $R_i(w) \subseteq [\varphi]$.

c.1.3 Suppose $w \in [(\zeta > \varphi)]$. If w/ζ is consistent, then, $w/\zeta \cup \{\sim \varphi\}$ is inconsistent, for otherwise $w/\zeta \cup \{\sim \varphi\} \subseteq w'$ for some $w' \in W$. Thus, either w/ζ being consistent or not, for some $\{\chi_1, ..., \chi_k\} \subseteq w/\zeta$ we have $\vdash \sim (\chi_1 \wedge ... \wedge \chi_k) \vee \varphi$, that is, $\vdash_K (\chi_1 \wedge ... \wedge \chi_k) \Rightarrow \varphi$. (Notice that k may be 0, and $\{\sim \varphi\}$ is inconsistent, i.e., $\vdash \sim\sim \varphi$.) By (RCK) or (CN) (in case when $k = 0$), we have $\vdash_K ((\zeta > \chi_1) \wedge ... \wedge (\zeta > \chi_k)) \Rightarrow (\zeta > \varphi)$. Now, $\{\chi_1, ..., \chi_k\} \subseteq w/\zeta$ implies that $(\zeta > \chi_1), ..., (\zeta > \chi_k) \in w$, and so $(\zeta > \varphi) \in w$. Thus, if $w \in [(\zeta > \varphi)]$, then $w \in |(\zeta > \varphi)|$.

c.1.4 Suppose $w \in |(\zeta > \varphi)|$, then $\varphi \in w/\zeta$, and so $w/\zeta \subseteq w'$ implies that $\varphi \in w'$. Thus, $f([\zeta], w) = f(|\zeta|, w) \subseteq \varphi$, and hence $w \in [(\zeta > \varphi)]$.

c.2 Suppose that $\varphi, \chi \in L^k$, and suppose that $[\varphi] = |\varphi|$ and $[\chi] = |\chi|$.

c.2.1 By lemma 3.4, $w \in [\varphi \wedge \chi] = [\varphi] \cap [\chi] = |\varphi| \cap |\chi$ if and only if both φ and χ belong to w if and only if $\varphi \wedge \chi \in w$ if and only if $w \in |\varphi \wedge \chi|$.

c.2.2 By lemma 3.4, $w \in [\sim \varphi] = W - [\varphi] = W - |\varphi$ if and only if $\varphi \notin w$ if and only if $\sim \varphi \in w$.

Thus, we have shown that $(M, w) \models \varphi$ if and only if $\varphi \in w$. Since any consistent formula belongs to an element in W, it follows that for any consistent formula φ, there is a state w such that $(M, w) \models \varphi$. It remains to show that $M \in \mathfrak{M}$ and we shall show that M satisfies (s-Tran), (s-Det),(s-D), and (s-4).

a. (s-Tran) Suppose that $f(A, w) \subseteq B$ and $f(B, w) \subseteq C$, $A, B, C \in \Sigma$. Let $A = [\zeta]$, $B = [\xi]$ and $C = [\eta]$. Then we have $\zeta > \xi \in w$ and $\xi > \eta \in w$. By (Tran), we have $\zeta > \eta \in w$. Thus, for all $w' \in f(A, w)$ $\eta \in w'$, i.e., $f(A, w) \subseteq C$.

b. (s-Det) Let $A = [\zeta]$. Then if w/ζ is inconsistent, there are formulas $\chi_1, ..., \chi_k \in w/\zeta$ such that $\vdash_K \sim (\chi_1 \wedge ... \wedge \chi_k)$. By (CN), we have then $\zeta > \sim (\chi_1 \wedge ... \wedge \chi_k) \in w$. Now $\vdash_K ((\zeta > \chi_1) \wedge ... \wedge (\zeta > \chi_k)) \Rightarrow (\zeta > (\chi_1 \wedge ... \wedge \chi_k))$ holds by (RCK) and so $(\zeta > (\chi_1 \wedge ... \wedge \chi_k)) \in w$. But by (Det), $\zeta > \sim (\chi_1 \wedge ... \wedge \chi_k)$ is not consistent with

$(\zeta > (\chi_1 \wedge ... \wedge \chi_k))$. Thus, w/ζ is consistent and so $f(A, w)$ cannot be empty.

c. (s-D) Let $w \in W$. Suppose that w/B_i is inconsistent. Then we have $\vdash\sim (\chi_1 \wedge ... \wedge \chi_k)$ for some $\chi_1, ..., \chi_k \in w/B_i$. Thus, we have $\vdash B_i(\sim (\chi_1 \wedge ... \wedge \chi_k))$ by (NEC) and so $B_i(\sim (\chi_1 \wedge ... \wedge \chi_k)) \in w$. Now, since $B_i(\chi_1), ..., B_i(\chi_k) \in w$, $B_i(\chi_1 \wedge ... \wedge \chi_k) \in w$. Therefore, $B_i(\sim (\chi_1 \wedge ... \wedge \chi_k)) \wedge B_i(\chi_1 \wedge ... \wedge \chi_k) \in w$, which contradicts to the consistency of w.

d. (s-4) Let $w' \in R_i(w)$ and $w'' \in R_i(w')$. Then $w/B_i \subseteq w'$ and $w'/B_i \subseteq w''$. Let $\varphi \in w/B_i$, then $B_i(\varphi) \in w/B_i \subseteq w'$ and so $\varphi \in w'/B_i \subseteq w''$. Therefore, $w/B_i \subseteq w''$. □

Proof of theorem 5.1:

Proof. Let $Z = \{z_1, ..., z_S\}$ and suppose that, without lost of generality, z_1 is the path derived from the subgame perfect equilibrium s.

Define a model $M = (W, f, R_1, ..., R_n, V)$ as follows:

a. $W = Z \cup \{(v, a) : v \in N,\ a \neq s(v), \text{ but } a \in A(v),\ t_v^{-1}(a) \notin Z\} \cup \{z^*\}$.

b. We define V as follows:

b.1. For any $z \in Z$, let $v(z) \in N$ satisfy (1) there exist $v(z) = v_1, ..., v_k = z$ such that $t_{v_{l-1}}(v_l) = s(v_{l-1})$, $l = 2, ..., k$; (2) $t_{v'}(v(z)) \neq s(v')$, where $v' \prec v(z)$.

b.1.1. If $v(z)$ exists, define $V(d_{\tau(v)}(v, a), z) = \top$ if and only if $v \in Path(v(z), z)$ and $a = s(v)$, $v \in N$, $z \in Z$, $a \in A$ and define $V(o(v), z) = \top$ if and only if $v \in Path(v(z), z)$, $v \in N \cup Z$.

b.1.2. If $v(z)$ does not exist, then define $V(o(v), z) = \bot = V(d_{\tau(v)}(v, a), z)$ for all $v \in N$ and $a \in A$, and define $V(o(z'), z) = \top$ if and only if $z' = z$ for all $z \in Z$.

b.1.3. For all $z \in Z$, $V(o(z'), z) = \top$ if and only if $z' = z$.

b.1.4. Define $V(o(v), z^*) = \bot$ for all $v \in N$, and define $V(d_{\tau(v)}(v, a), z^*) = \top$ if and only if $a \notin A(v)$ for all $v \in N$.

b.2. For any $(v, a) \in W$, define $V(d_{\tau(v)}(v', a'), (v, a)) = \top$ if and only if $(v', a') = (v, a)$

and define $V(o(v'), (v, a)) = \bot$ for all $v' \in N$.

b.3. For each $w \in W$, $V(P_i(z, z'), w) = \top$ if and only if $(z, z') \in \succeq_i$, $i = 1, .., n$, for any $z, z' \in Z$; $V(P_i(z, z^*), w) = \top$ and $V(P_i(z^*, z), w) = \bot$ for all $z \in Z$; $V(P_i(z', z), w) = \bot$ for all $z, z' \in Z^0 - Z$.

b.5. $V(S(v, v'), z_1) = \top$ if and only if $v \prec v'$ for all $v, v' \in N$ and $V(S(v, v'), z_1) = \bot$ for all $v \in N$, and $v' \in N^0 - N$.

b.6. For nay $w \in W$, $V(d_{\tau(v)}(v, a), w) = \bot = V(o(v), w)$ for any $v \in (N^0 \cup Z^0) - (N \cup Z \cup \{z^*\})$ and any $a \in A$.

c. Let $\Sigma = 2^W$. Define $f(\{z\}, z_1) = \{z\}$ for all $z \in Z \cup \{z^*\}$, $f(\{(v, a)\}, z_1) = [o(t_v^{-1}(a))]$ for all $(v, a) \in W$, and define $f(X, z_1) = X$ for all other $X \in \Sigma$, $X \neq \emptyset$. $f(\emptyset, z_1) = \{z_1\}$.

d. For each $i = 1, ..., n$, we define $R_i(z_1) = \{z_1\}$.

(s-Det) is true. Notice that $V(o(v), z) = \top$ if and only if $v \in Path(v(z), z)$ by (b.1.1.) and since s gives a unique path from each $v' \in N$, $[o(v)]$ is a singleton and is a subset of Z. To check (s-Tran), it suffices to notice that $f(X, z_1) \subset Y$ implies that $f(X, z_1) \subset f(Y, z_1) = Y$ by (c.) for all $X, Y \in \Sigma$.

Then we claim that $(M, z_1) \models DC \wedge G$. Since, by (b.5.), $(M, z_1) \models S(v, v')$ if and only if $v \prec v$, and $(M, z_1) \models\sim S(v, v')$ for all $v \in N$, $v' \in N^0 - N$, it follows that $(M, z_1) \models \wedge G_S$. It is easy to check that $(M, z_1) \models \wedge G_P$ by (b.3). We shall show that $(M, z_1) \models \wedge G_C$ also. Suppose that $a = t_v(v')$.

Case (1): $a = s(v)$. Then, by (b.1.) and (c.), $[d_{\tau(v)}(v, a)] = Z(v; s) =_{def} \{z\}$, and so $o(v') \in Path(v(z), z)$. Thus, $f([d_{\tau(v)}(v, a)], z_1) = \{z\} = [o(v')]$. Therefore, $(M, z_1) \models d_{\tau(v)}(v, a) > o(v')$.

Case (2): $a \neq s(v)$ and $a \in A(v)$.

(2.1) $v' \in N$. Then, by (b.1.), $[d_{\tau(v)}(v, a)] = \{(v, a)\}$ and $f([d_{\tau(v)}(v, a)], z_1) = [o(t_v^{-1}(a))]$. Thus, we have $(M, z_1) \models d_{\tau(v)}(v, a) > o(v')$.

(2.2) $v' \in Z$. Then, by (b.1.), $[d_{\tau(v)}(v,a)] = \{v'\}$ and so, by (c.), $f([d_{\tau(v)}(v,a)], z_1) = \{v'\} = [o(v')]$ Thus, we have $(M, z_1) \models d_{\tau(v)}(v,a) > o(v')$.

Case (3): $a \notin A(v)$. Then, by (b.1.), $[d_{\tau(v)}(v,a)] = \{z^*\}$ and so by (c.), $f([d_{\tau(v)}(v,a)], z_1) = f(\{z^*\}, z_1) = \{z^*\} = [o(z^*)]$. Thus, $(M, z_1) \models d_{\tau(v)}(v,a) > o(z^*)$.

Moreover, for any $z \neq z' \in Z \cup \{z^*\}$, $f([o(z)], z_1) = \{z\} \subseteq W - \{z'\} = [\sim o(z')]$ and so we have $(M, z_1) \models o(z) >\sim o(z')$. Thus, we have $(M, z_1) \models G$.

Notice that $R_i(z_1) = \{z_1\}$ for all i, and so

$$(M, z_1) \models \varphi \Leftrightarrow B_i(\varphi), \tag{31}$$

for any $\varphi \in L$.

For any $v \in N$, $f([o(v)], z_1) = Z(v; s) = [d_{\tau(v)}(v, s(v))]$. Moreover, $Z(v; s) \subseteq W - (\{(v, a')\} \cup \{z^*\}) \subseteq W - [d_{\tau(v)}(v, a')]$. Thus,

$$(M, z_1) \models \wedge\{(o(v) > d_{\tau(v)}(v, s(v))) \wedge \sim (o(v) > d_{\tau(v)}(v, a)) : v \in N, a \neq s(v), a \in A\}. \tag{32}$$

Moreover, since $R_i(z_1) = \{z_1\}$ and so $(M, z_1) \models \wedge\{PD_i(N, s) : i \in I\}$.

Therefore, by (7)-(12),

$$(M, z_1) \models \wedge\{(RC_v \wedge PC_v) \Leftrightarrow$$
$$Best_{\tau(v),Sub(v),s|_v}(v, s(v)) \wedge (\wedge\{\sim SBest_{\tau(v),Sub(v),s|_v}(v, a) : a \neq s(v)\}) : v \in N\}. \tag{33}$$

Combining (32), (33), (31), (5), and (6), it follows that for all $v \in N$,

$$(M, z_1) \models$$
$$(\wedge\{Dom_{\tau(v),v}(s(v), a) : a \neq s(v)\})$$
$$\wedge(\wedge\{\sim SDom_{\tau(v),v}(a, s(v)) : a \neq s(v)\}) \Rightarrow \wedge(RC_v \wedge PC_v).$$

By lemma 4.4 and (32), we have $(M, z_1) \models d_{\tau(v)}(v, s(v)) > o(z(v))$ ($z(v)$ is such that $Z(v; s) = \{z(v)\}$) and $(M, z_1) \models d_{\tau(v)}(v, a) > o(z(t_v^{-1}(a)))$ for any $v \in N$. Moreover, by (Det) and theorem 3.1, it follows that $(M, z_1) \models \sim (d_{\tau(v)}(v, s(v)) > o(z'))$ for any $z' \neq z$.

Since s is a subgame perfect, it follows that for any $v \in N$, $z(v) \succeq_{\tau(v)} z(t_v^{-1}(a))$ for any $a \in A(v)$. Thus, for any $a' \neq s(v)$,

$$(M, z_1) \models (d_{\tau(v)}(v, s(v)) > o(z(v))) \wedge (d_{\tau(v)}(v, a) > o(z(t_v^{-1}(a)))) \wedge (P_{\tau(v)}(z(v), z(t_v^{-1}(a))),$$

and so

$$(M, z_1) \models \sim SDom_{\tau(v),v}(a, s(v)).$$

For any $a \neq s(v)$, $(M, z_1) \models (d_{\tau(v)}(v, s(v)) > o(z)) \wedge (d_{\tau(v)}(v, a) > o(z'))$ if and only if $z = z(v)$, and $z' = z(t_v^{-1}(a))$. Since $(M, z_1) \models P_{\tau(v)}(z(v), z(t_v^{-1}(a)))$ for any $a \in A(v)$ and since $(M, z_1) \models P_{\tau(v)}(z(v), z^*)$, it follows that

$$(M, z_1) \models \wedge\{Dom_{\tau(v),v}(s(v), a) : a \neq s(v)\}.$$

Therefore, we have $(M, z_1) \models \wedge\{RC_v \wedge PC_v : v \in N\}$.

Combining the results, we have $(M, z_1) \models DC \wedge G$. By (32), we also have $(M, z_1) \models D_g(s)$. These give us that $(M, z_1) \models DC \wedge G \wedge D_g(s)$. From (31), it follows that

$$(M, z_1) \models \wedge\{B^k(DC \wedge G \wedge D(s)) : k = 0, ..., K\}.$$

□

We give a general proof for the theorem 5.3.

Theorem 7.1. *Suppose that $dep(v_0) = K \geq 2$. Then*

$$(\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\}) \wedge (\wedge\{\sim D_{\tau(v_0)}(v_0, a) : a \in A\})$$

is a consistent formula.

Proof. Suppose that $\tau(v^0) = 1$. Let z_1 be the terminal node such that there are $v^0 = v_0$, v_1,...,$v_{K+1} = z_1$ and $v_k \prec v_{k+1}$, $k = 0, ..., K$. Consider a model M as follows:

$M = (W, f, R_1, ..., R_n, V)$, where

a. $W = Z \cup \{(v, a) : v \in N, a \neq s(v), a \in A(v),\ t_v^{-1}(a) \notin Z\} \cup \{w_1, ..., w_{K-1}\} \cup \{w^*\} \cup \{z^*\}$, $w_1, ..., w_{K-1}$, and w^* are assumed to be different from any $z \in Z^0$ and any

$(v, a) \in N^0 \times A$ and from each other. We also enumerate Z as $\{z_1, ..., z_K, ..., z_{K^*}\}$, where $K^* = |Z| \geq K$. W.l.o.g., we assume that z_2 is lead by s.

b. V is defined as follows:

b.1. ($w \in Z \cup \{z^*\}$, Φ_{GR}, Φ_{NZ}) For any $z \in Z$, let $v(z) \in N$ satisfy, if any, (1) there exist $v(z) = v_1, ..., v_k = z$ such that $t_{v_{l-1}}(v_l) = s(v_{l-1})$, $l = 2, ..., k$; (2) $t_{v'}(v) \neq s(v')$, where $v' \prec v$.

b.1.1. Case 1: $v(z)$ exists. Assign $V(d_{\tau(v)}(v, a), z) = \top = V(o(v), z)$ if and only if $v \in Path(v(z), z)$ and $a = s(v)$ for all $v \in N \cup Z$ and $a \in A$. For all other $p \in \Phi_{GR} \cup \Phi_{NZ}$, assign $V(p, z) = \bot$.

b.1.2. Case 2: $v(z)$ does not exist. Assign $V(d_{\tau(v)}(v, a), z) = \top = V(o(z'), z)$ if and only if $v \prec z$, $t_v(z) = a$, and $z' = z$, for all $v \in N$ and $z' \in Z$. For all other $p \in \Phi_{GR} \cup \Phi_{NZ}$, assign $V(p, z) = \bot$.

b.1.3. Case 3: $z = z^*$. Assign $V(d_{\tau(v)}(v, a), z^*) = \top = V(o(z^*), z^*)$ if and only if $a \in A$, $a \notin A(v)$, for all $v \in N$. For all other $p \in \Phi_{GR} \cup \Phi_{NZ}$, assign $V(p, z^*) = \bot$.

b.2. ($w = (v, a) \in W$, Φ_{GR}, Φ_{NZ}) Assign $V(d_{\tau(v')}(v', a'), (v, a)) = \top$ if and only if $(v', a') = (v, a)$ for all $(v, a) \in N \times A$. For all other $p \in \Phi_{GR} \cup \Phi_{NZ}$, assign $V(p, (v, a)) = \bot$.

b.3. ($w = w_k$, $k = 1, ..., K-1$, and w^*, Φ_{GR}, Φ_{NZ}) Define $N^k = \{v \in N : dep(v) = k\} \cap Path(z_1)$, $k = 1, ..., K$. Since $dep(v^0) = K$, $N^k \neq \emptyset$ for all $k = 1, ..., K$, and clearly $|N^k| = 1$. Let $N^k = \{v_k\}$. Assign $V(o(v_{k'}), w_k) = \top$ if and only if $k' > k$ for $k' = 1, ..., K$. For all other $p \in \Phi_{GR} \cup \Phi_{NZ}$, assign $V(p, w_k) = \bot$. For all $p \in \Phi_{GR} \cup \Phi_{NZ}$, assign $V(p, w^*) = \bot$.

b.4. (Φ_P and Φ_S) For each $w \in W$, $z, z', z'' \in Z$, and for each $i \in I$, assign $V(P_i(z', z''), w) = \top$ if and only if $z' \succeq_i z''$, and assign $V(P_i(z', z^*), w) = \top$, $V(P_i(z^*, z'), w) = \bot$. Assign $V(S(v, v'), w) = \top$ if and only if $v, v' \in N$ and $v \prec v'$. For all other $p \in \Phi_S \cup \Phi_P$, assign $V(p, w) = \bot$.

c. Let $\Sigma = 2^W$. We consider the construction of f for $z_2, ..., z_K$ and w^* only. For any $v \neq v_k$, $k = 1, ..., K$, $[o(v)] = Z(v; s) =_{def} \{z(v; s)\}$. $[o(v_k)] = \{z(v_k; s), w_1, .., w_{k-1}\}$.

Moreover, $[d_{\tau(v)}(v, s(v))] = \{z(t_v^{-1}(a); s)\}$, $[d_{\tau(v)}(v, a)] = \{(v, a)\}$ if $a \neq s(v), a \in A(v)$ and $Sub(v) \neq \emptyset$, $[d_{\tau(v)}(v, a)] = \{z^*\}$ if $a \notin A(v)$, and $[d_{\tau(v)}(v, a)] = \{t_v^{-1}(a)\}$ if $Sub(v) = \emptyset$ and $a \in A(v)$.

c.1. We define $f(X, w^*) = \{w^*\}$ for all $X \in \Sigma$.

c.2. For $z' = z_2, ..., z_K$, define $f(\{z\}, z') = \{z\}$ for all $z \in Z$, and $f(\{(v, a)\}, z') = [o(t_v^{-1}(a))]$ for all $(v, a) \in W$.

c.3. Consider $z_k \in W$, $k = 2, ..., K$. We define $f(\{z(v_{k'}; s), w_1, ..., w_{k'-1}\}, z_k) = \{z(v_{k'}; s)\}$ if $k' < k$ and $f(\{z(v_{k'}; s), w_1, ..., w_{k'-1}\}, z_k) = \{z(v_{k'}; s), w^*\}$ if $k' \geq k$. For any subset $Y \subseteq \{w_1, ..., w_{K-1}\}$, let $f(Y, z_k) = \{z_1\}$. For all other nonempty $X \in \Sigma$, we define $f(X, z_k) = X$. Let $f(\emptyset, z_k) = \{z_1\}$.

d. Let $i_k = \tau(v_k)$, $k = 2, ..., K$. We define R_1, ..., R_n to be the partitions of $\{z_2, ..., z_K, w^*\}$ as follows: (d.1) $R_{i_2}(w^*) = \{z_2, w^*\}$, and $R_i(w^*) = \{w^*\}$ for all other $i \neq i_2$. (d.2) $R_{i_3}(z_2) = \{z_3, z_2\}$, $R_i(z_2) = \{z_2\}$ for all i other than i_2 and i_3. (d.3) Suppose that $R_i(z_{k-1})$ has been defined for all i, $k > 2$. Then define $R_{i_{k+1}}(z_k) = \{z_k, z_{k+1}\}$ and define $R_i(z_k) = \{z_k\}$ for all i other than i_{k+1} and i_k.

First we check that M is a model that satisfy (s-Tran) and (s-Det). (s-Det) is clearly true. Now, (s-Tran) holds at w^* since $f(X, w^*) = f(Y, w^*)$ for any $X, Y \in \Sigma$. Fix some $k \in \{2, ..., K\}$, and consider z_k. Suppose that $f(X, z_k) \subseteq Y$ and $f(Y, z_k) \subseteq Z$. If $Y \subseteq f(Y, z_k)$, then $f(X, z_k) \subseteq Z$. Suppose that $Y \nsubseteq f(Y, z_k)$. Then, by (c.), Y has the form $\{z(v_{k'}; s), w_1, ..., w_{k'-1}\}$ or is a subset of $\{w_1, ..., w_{K-1}\}$ or is $\{z\}$ for some $z \in Z$.

Case (1): Y has the form $\{z(v_{k'}; s), w_1, ..., w_{k'-1}\}$ for some k'. Since $f(X, z_k) \subseteq Y$, by (c.), $X = \{z(v_{k''}; s), w_1, ..., w_{k''-1}\}$ for some $k'' < k$ and $z(v_{k''}; s) = z(v_{k'}; s)$. Then $f(X, z_k) = \{z(v_{k'}; s)\} \subseteq f(Y, z_k) \subseteq Z$.

Case (2): $Y \subseteq \{w_1, ..., w_{K-1}\}$. Then there is no X such that $f(X, z_k) \subseteq Y$.

Case (3): $Y = \{z\}$. Then $f(X, z_k) = Y \subseteq Z$.

Thus, we have that M satisfy (s-Tran) and (s-Det). We shall show that $(M, z) \models G$, for each $z = z_2, ..., z_K$. Fix such a z_k. By (b.4.), it is easy to check that $(M, z_k) \models \wedge(G_P \cup G_S)$.

By (b.), we have $[o(z)] = \{z\}$ for any $z \in Z \cup \{z^*\}$, and $[o(z)] = \emptyset$ for all $z \in Z^0 - (Z \cup \{z^*\})$. Thus, it is to see from (c.) that $(M, z_k) \models \wedge\{o(z) >\sim o(z') : z, z' \in Z^0\}$. We shall show that $(M, z_k) \models \wedge G_C$ also. Suppose that $a = t_v(v')$.

Case (1): $a = s(v)$. Then, $[d_{\tau(v)}(v, a)] = Z(v; s) =_{def} \{z\}$, and so $o(v') \in Path(v(z), z)$. Thus, $f([d_{\tau(v)}(v, a)], z_k) = \{z\} = [o(v')]$. Therefore, $(M, z_k) \models d_{\tau(v)}(v, a) > o(v')$.

Case (2): $a \neq s(v)$ and $a \in A(v)$.

(2.1) $v' \in N$. Then, by (b.2.), $[d_{\tau(v)}(v, a)] = \{(v, a)\}$ and $f([d_{\tau(v)}(v, a)], z_k) = [o(t_v^{-1}(a))]$. Thus, we have $(M, z_k) \models d_{\tau(v)}(v, a) > o(v')$.

(2.2) $v' \in Z$. Then, by (b.1.2.), $[d_{\tau(v)}(v, a)] = \{v'\}$ and so, by (c.), $f([d_{\tau(v)}(v, a)], z) = \{v'\} = [o(v')]$ Thus, we have $(M, z_k) \models d_{\tau(v)}(v, a) > o(v')$.

Case (3): $a \notin A(v)$. Then, by (b.1.3.), $[d_{\tau(v)}(v, a)] = \{z^*\}$ and so by (c.), $f([d_{\tau(v)}(v, a)], z_k) = f(\{z^*\}, z_1) = \{z^*\} = [o(z^*)]$. Thus, $(M, z_k) \models d_{\tau(v)}(v, a) > o(z^*)$.

Moreover, we claim that $V(D_{\tau(v)}(v, a), z_k) = \perp$ for all $v = v_k, ..., v_K$, $a \in A$, and $V(D_{\tau(v)}(v, a), z_k) = \top$ if and only if $a = s(v)$ for all other $v \in N$, $a \in A$.

(1) Consider any $k' = k, ..., K$. Then, $[o(v_{k'})] = \{z(v_{k'}; s), w_1, ..., w_{k'-1}\}$ and $f([o(v_{k'})], z_k) = \{z(v_{k'}; s), w^*\}$. Since $(M, w^*) \models\sim d_{\tau(v)}(v_{k'}, a)$ for any $a \in A$, $(M, z_k) \models\sim (o(v_{k'}) > d_{\tau(v)}(v_{k'}, a))$ for any $a \in A$.

(2) Consider any $k' = 2, ..., k - 1$. Then, $[o(v_{k'})] = \{z(v_{k'}; s), w_1, ..., w_{k'-1}\}$ and $f([o(v_{k'})], z_k) = \{z(v_{k'}; s)\}$. Since $(M, z(v_{k'}; s)) \models d_{\tau(v)}(v_{k'}, s(v_{k'})) \wedge \sim d_{\tau(v)}(v_{k'}, a)$ for any $a \neq s(v_{k'})$, $a \in A$, it follows that $(M, z_k) \models o(v_{k'}) > d_{\tau(v)}(v_{k'}, a)$ if and only if $a = s(v_{k'})$, $a \in A$.

(3) Consider any $v \neq v_2, ..., v_K$, $v \in N$. Then $[o(v)] = \{z(v; s)\}$ and $f([o(v)], z_k) = \{z(v; s)\}$ and $(M, \{z(v; s)\}) \models d_{\tau(v)}(v, s(v)) \wedge \sim d_{\tau(v)}(v, a)$ for any $a \neq s(v)$, $a \in A$. Therefore, $(M, z_k) \models o(v) > d_{\tau(v)}(v, a)$ if and only if $a = s(v)$, $a \in A$.

Therefore, we have the following, for $k = 2, ..., K$,

$$(M, z_k) \models G \wedge (\wedge\{(o(v) > d_{\tau(v)}(v, s(v)))\wedge \sim (o(v) > d_{\tau(v)}(v, a)) : \\ v \in N, a \in A, a \neq s(v), v \neq v_k, ..., v_K\}). \tag{34}$$

and

$$(M, z_k) \models (\wedge\{\sim (o(v) > d_{\tau(v)}(v, a)) : a \in A, v = v_k, ...v_K\}). \tag{35}$$

But we have

$$(M, w^*) \models\sim G. \tag{36}$$

Now we shall show that for each $k = 2, ..., K$, $(M, z_k) \models DC$. For all $k = 2, ..., K$, we have $(M, z_k) \models\sim Sub(v^0, v)$ for any $v \in N^0 - N$, it follows that for each $k = 2, ..., K$,

$$(M, z_k) \models (\wedge\{DC_v : v \in N\}) \Rightarrow DC.) \tag{37}$$

Since we have

$$(M, z) \models Sub(v, Sub(v)) \wedge (\wedge\{\sim Sub(v, N') : N' \subseteq N,\ N' \neq Sub(v)\}),$$

it follows that

$$(M, z) \models PC_v \wedge RC_v \Leftrightarrow PC_{v,Sub(v)} \wedge RC_{v,Sub(v)}. \tag{38}$$

Since s is a subgame perfect equilibrium, and by (34), it is easy to check that for each $k = 2, ..., K$ and for all $v \in N$, $v \neq v_k, ..., v_K$,

$$(M, z_k) \models Best_{\tau(v),Sub(v),s|_v}(v, s(v)) \\ \wedge(\wedge\{\sim SBest_{\tau(v),Sub(v),s|_v}(v, a) : a \neq s(v), a \in A\}), \tag{39}$$

and for $v = v_k, ..., v_K$, since by (35),

$$(M, z_k) \models\sim (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v' \in Sub(v)\}),$$

it follows that

$$(M, z_k) \models \wedge\{Best_{\tau(v),Sub(v),s|_v}(v, a) \\ \wedge SBest_{\tau(v),Sub(v),s|_v}(v, a) : a \in A\}). \tag{40}$$

Similarly, since, for any $v \in N$,

$$(M, w^*) \models \sim (\wedge\{o(v') > d_{\tau(v')}(v', s(v')) : v \in Sub(v)\})$$
$$\wedge(\wedge\{\sim (d_{\tau(v)}(v, a) > o(z)) : z \in Z^0, a \in A\}),$$

it follows that

$$(M, w^*) \models \wedge\{Best_{\tau(v),Sub(v),s|_v}(v, a)$$
$$\wedge SBest_{\tau(v),Sub(v),s|_v}(v, a) : a \in A\}). \tag{41}$$

Thus, for each $k = 2, ..., K$ and for each $v \in N$, $v \neq v_k, ..., v_K$,

$$(M, z_k) \models (\wedge\{\sim B_i(SBest_{\tau(v),Sub(v),s|_v}(v, a)) : i \in I, a \neq s(v), a \in A\}). \tag{42}$$

Combining (39), (40), and (41) with (d.), for $z = z_1, ..., z_{K-1}$, for all $v \in N$,

$$(M, z) \models (\wedge\{B_i(Best_{\tau(v),Sub(v),s|_v}(v, s(v))) : i \in I\}). \tag{43}$$

For $v \in N$ such that $Sub(v) = \emptyset$, by (38),

$$(M, z_k) \models PC_v \wedge RC_v \Leftrightarrow$$
$$\{(D_{\tau(v)}(v, a) \Rightarrow Best_{\tau(v),\emptyset}(v, a)) \wedge (SBest_{\tau(v),\emptyset}(v, a) \Rightarrow D_{\tau(v)}(v, a)) : a \in A\}.$$

Since $V(D_{\tau(v)}(v, a), z) = \top$ if and only if $a = s(v)$ for $z = z_2, ..., z_K$ and for all $v \in N$ such that $Sub(v) = \emptyset$, by (34), it follows from (39) that, for $k = 2, ..., K$,

$$(M, z_k) \models \wedge\{PC_v \wedge RC_v : v \in V, Sub(v) = \emptyset\}. \tag{44}$$

Fix a $k \in \{2, ..., K\}$, and consider any $v \in N$ such that $Sub(v) \neq \emptyset$.

Case (1): $v \neq v_2, ..., v_K$. Since, by (34) and (35), $(M, z_k) \models D_{\tau(v')}(v', a)$ only if $a = s(v')$ for all $v \in N$, it follows that $(M, z_k) \models PD_{\tau(v')}(Sub(v'), t)$ only if $t = s|_{Sub(v')}$. Thus, by (43) and (38), $(M, z_k) \models RC_v$; by (34), and (39), $(M, z_k) \models PC_v \Leftrightarrow D_{\tau(v)}(v, s(v)) \vee \sim PD_{\tau(v)}(Sub(v), s|_v)$. Therefore, for such v, we have

$$(M, z_k) \models PC_v \wedge RC_v \Leftrightarrow (D_{\tau(v)}(v, s(v)) \vee \sim PD_{\tau(v)}(Sub(v), s|_v)). \tag{45}$$

By (38) and (34),

$$(M, z_k) \models (\wedge\{PC_v \wedge RC_v : v \neq v_2, ..., v_K\}). \tag{46}$$

Case (2): $v = v_2, ..., v_K$ (notice that $Sub(v_1) = \emptyset$). For each $k' = 3, ..., K$, by (d.), $R_i(z_{k'}) \subseteq \{z_{k'-1}, ..., z_K\}$ for all $i \in I$, and $R_i(z_{k'}) \subseteq \{z_{k'}, ..., z_K\}$ for all $i \in I$ other than $i_{k'}$. Moreover, $R_{i_2}(z_2) = \{z_2, w^*\}$. By (34) and (35), $(M, z_k) \models D_{\tau(v)}(v, s(v)) \wedge (\wedge\{\sim D_{\tau(v)}(v, a) : a \neq s(v), a \in A\})$ for $v = v_1, ..., v_{k-1}$, and $(M, z_k) \models \wedge\{\sim D_{\tau(v)}(v, a) : a \in A\}$ for $v = v_k, ..., v_K$. Thus, we have

$$(M, z_k) \models (\wedge\{\sim B_{i_k}(D_{\tau(v_{k-1})}(v_{k-1}, a)) : a \in A\}), \tag{47}$$

$$(M, z_k) \models (\wedge\{\sim B_i(D_{\tau(v)}(v, a)) : v = v_k, ..., v_K, i \in I, a \in A\}), \tag{48}$$

$$(M, z_k) \models PD_{i_k}(N - \{v_{k-1}, ..., v_K\}, s), \tag{49}$$

$$(M, z_k) \models \wedge\{PD_i(N - \{v_k, ..., v_K\}, s) : i \neq i_k\}. \tag{50}$$

By the same reason as in Case (1), we have $(M, z_k) \models RC_v$. It remains to show that $(M, z_k) \models PC_v$. Notice that by (39), for $v = v_2, .., v_{k-1}$, $(M, z_k) \models D_{\tau(v)}(v, s(v)) \Rightarrow PC_v$ and for $v = v_k, ..., v_K$, $(M, z_k) \models \sim PD_{\tau(v)}(Sub(v), s) \Rightarrow PC_v$.

For any $v = v_1, ..., v_{k-1}$, $(M, z_k) \models D_{\tau(v)}(v, s(v))$ and so $(M, z_k) \models PC_v \wedge RC_v$. For $v = v_k, ..., v_K$, by (47) and (48), it is easy to check that

$$(M, z_k) \models \sim PD_{\tau(v)}(Sub(v), s|_{Sub(v)})$$

and so, $(M, z_k) \models PC_v \wedge RC_v$. Finally, consider $v = v_k$. By (47),

$$(M, z_k) \models \sim PD_{i_k}(Sub(v_k), s|_{Sub(v_k)})$$

since $(M, z_k) \models (\wedge\{\sim B_{i_k}(v_{k-1}, a) : a \in A\})$. Thus, $(M, z_k) \models PC_{v_k} \wedge RC_{v_k}$.

We have shown that, for $k = 1, ..., K - 1$,

$$(M, z_k) \models \wedge\{PC_v \wedge RC_v : v \in N^0\}. \tag{51}$$

Combining (34), (37), and (51), we have for $k = 1, ..., K-1$,

$$(M, z_k) \models G \wedge DC. \tag{52}$$

We show by induction that, for $k = 2, ..., K$, and for $k' \geq k$,

$$(M, z'_k) \models \wedge\{B_i(B^{k-2}(DC \wedge G)) \wedge B^{k-2}(DC \wedge G) : i \neq i_k\}.$$

Consider $k = 2$. For all $k' \geq 3$, and for each $i \in I$, $R_i(z_{k'})$ is included in $\{z_2, ..., z_K\}$. Moreover, for all i other than i_2, and for all $k' \geq 2$, $R_i(z_{k'})$ is included in $\{z_2, ..., z_K\}$. Thus, for $k' \geq 2$,

$$(M, z_{k'}) \models (\wedge\{B_i(DC \wedge G) : i \in I, i \neq i_2\}) \wedge (DC \wedge G).$$

Suppose that the claim holds for all $l \leq k$, $k > 1$. Then we have for all $k' \geq k$,

$$(M, z_{k'}) \models (\wedge\{B_i(B^{k-2}(DC \wedge G)) : i \neq i_k\}) \wedge B^{k-2}(DC \wedge G).$$

Now, $k'' \geq k+2$ implies that $R_i(z_{k''}) \subseteq \{z_{k+1}, ..., z_K\}$, and so for such k'',

$$(M, z_{k''}) \models (\wedge\{B_i(B^{k-2}(DC \wedge G)) : i \in I\}).$$

This implies that

$$(M, z_{k''}) \models B^{k-1}(DC \wedge G).$$

Moreover, by (d.) and the assumption $i_k \neq i_{k+1}$, $R_{i_k}(z_{k+1}) \subseteq \{z_{k+1}, ..., z_K\}$, and so we have

$$(M, z_{k+1}) \models B_{i_k}(B^{k-2}(DC \wedge G)).$$

Combining this with the induction hypothesis, we get

$$(M, z_{k+1}) \models B^{k-1}(DC \wedge G).$$

Finally, $k'' \geq k+1$ implies that $R_i(z_{k''}) \subseteq \{z_{k+1}, ..., z_K\}$ for all $i \in I$ other than i_{k+1}, it follows that

$$(M, z_{k''}) \models (\wedge\{B_i(B^{k-1}(DC \wedge G)) : i \neq i_{k+1}\}).$$

This finishes the induction.

Thus, at z_K, we have

$$(M, z_K) \models (\wedge\{B_i(B^{K-2}(DC \wedge G)) : i \neq 1\}) \wedge B^{K-2}(DC \wedge G).$$

By (35),

$$(M, z_K) \models (\wedge\{\sim D_1(v^0, a) : a \in A\}.$$

Combining these results we get

$$(M, z_K) \models B^{K-2}(DC \wedge G) \wedge (\wedge\{\sim D_1(v^0, a) : a \in A\}).$$

Since we have $(M, z_K) \models B_i(\varphi) \Rightarrow \varphi$, it follows that

$$(M, z_K) \models (\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\}) \wedge (\wedge\{\sim D_1(v^0, a) : a \in A\}).$$

By theorem 3.1,

$$(\wedge\{B^k(DC \wedge G) : k = 0, ..., K-2\}) \wedge (\wedge\{\sim D_1(v^0, a) : a \in A\})$$

is a consistent formula. □

References

[1] Aumann, R.J. (1995), "Backward induction and common knowledge of rationality." In: Nobel Symposium on Game Theory, Bjökborn, 1993, *Games and Economic Behavior* vol. 8 (1), 6–19.

[2] Aumann, R. J., Brandenburger, A. (1995), "Epistemic conditions for Nash equilibrium." *Econometrica* vol. 63, 1161V1180.

[3] Balkenborg, D. and E. Winter (1997), "A necessary and sufficient epistemic condition for playing backward induction." *Journal of Mathematical Economics* 27, 325-345.

[4] Battigalli, P. and M. Siniscalchi (1999), "Hierarchies of conditional beliefs, and interactive epistemology in dynamic games." *Journal of Economic Theory* vol. 88, 188-230.

[5] Board, O. (2004), "Dynamic interactive epistemology." *Games and Economic Behavior,* vol. 49, pp.49-80.

[6] Clausing, T. (2003), "Doxastic conditions for backward induction." *Theory and Decision,* vol. 54, pp. 315-36.

[7] Giordano, L. and Schwind, C. (2004), " Conditional logic of actions and causation." *Artificial Intelligence*, vol. 157, 239-279.

[8] Halpern, J. Y. and Y. Moses (1992), "A guide to completeness and complexity for modal logics of knowledge and belief." *Artificial Intelligence* vol. 54(3), 319-79.

[9] Harsanyi, J.C. (1967), "Games with incomplete information played by Bayesian players, Part 1: the basic model." *Management Science* vol. 14, 159-182.

[10] Kaneko, M. (1999), "Epistemic considerations of decision making in games." *Mathematical Social Sciences,* vol. 38, 105-137.

[11] Kaneko, M. (2002), "Epistemic logics and their game theoretic applications: Introduction." *Economic Theory*, vol. 19, 7-62.

[12] Lewis, D. (1981), "Counterfactuals and Comparative Possibility." in: W.L. Harper, R. Stalnaker, G. Pearce (Eds.), *Ifs*, Reidel, Dordrecht.

[13] Pearce, D.G. (1984), "Rationalizable strategic behavior and the problem of perfection." *Econometrica* vol. 52, 1029-1050.

[14] Reny, P.J. (1993), "Common belief and the theory of games with perfect information." *Journal of Economic Theory*, vol. 59, pp. 257–74.

[15] Samet, D. (1996), "Hypothetical knowledge and games with perfect information." *Games and Economic Behavior* vol. 17, 230-251.