

# Are Admissibility and Backward Induction Consistent?\*

Adam Brandenburger<sup>†</sup>

Amanda Friedenberg<sup>‡</sup>

Version: February 16, 2014

## Abstract

We revisit a fundamental question in the axiomatic approach to non-cooperative games—viz., the consistency of admissibility (A) and backward induction (BI). The literature has concluded that A and BI are consistent. However, we argue that, to reach this conclusion, the literature has implicitly assumed that BI satisfies a monotonicity property: If a solution concept satisfies BI, then a refinement of the solution concept also satisfies BI. We provide a formalization of BI in terms of fundamentals of the game and, from this, conclude that BI is a non-monotonic property. In fact, we show that A and BI are inconsistent on the domain of all games. It appears to be an open question whether they are consistent on a subdomain of games on which there is a well-defined BI outcome.

## 1 Introduction

Admissibility (i.e., the avoidance of weakly dominated strategies) and backward induction are basic properties to demand of a solution concept. They are viewed as fundamental requirements in the axiomatic approach to non-cooperative game theory. (See [Kohlberg and Mertens \(1986\)](#) and, for a recent survey, [Govindan and Wilson \(2008\)](#).) As such, a basic question arises: Are admissibility (A) and backward induction (BI) consistent?

The question of the consistency of A and BI is old. Indeed, the common view is that the question is closed—that A and BI are, in fact, consistent. Two results point to this conclusion:

- [Van Damme \(1984, Proposition 1, p.9\)](#) observed that any quasi-perfect equilibrium is a sequential equilibrium.
- [Kohlberg and Mertens \(1986, Proposition 0, p.1009\)](#) showed that any proper equilibrium induces a sequential equilibrium.

Recall, both quasi-perfect and proper equilibrium satisfy A. Moreover, the argument goes, both satisfy BI since both induce a sequential equilibrium. The consistency of A and BI follows.

---

\*We have benefited greatly from conversations with Hari Govindan and Bill Zame. We thank the editor, two referees, Selvin Akkus, Heski Bar-Isaac, Pierpaolo Battigalli, Willemien Kets, Priscilla Man, Andy McLennan, John Nachbar, Alex Peysakhovich, Carlos Pimienta, Ariel Rpeck, Marciano Siniscalchi, Harborne Stuart, and seminar participants at UCLA and the World Congress of the Econometric Society for valuable input. Support from the NYU Stern School of Business, the W.P. Carey School of Business at Arizona State University, and the Department of Economics at UCLA is gratefully acknowledged.

<sup>†</sup>Address: Stern School of Business and Center for Data Science, New York University, New York, NY 10012, adam.brandenburger@stern.nyu.edu, www.stern.nyu.edu/~abranden

<sup>‡</sup>Address: W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287, amanda.friedenberg@asu.edu, www.public.asu.edu/~afrieden

But, do quasi-perfect equilibrium and proper equilibrium satisfy BI? Let us review the argument for “yes.” Recall, in perfect-information games (satisfying a no-ties condition), any quasi-perfect equilibrium—and, therefore, any proper equilibrium—yields the BI outcome. In general games (i.e., both perfect- and imperfect-information games), sequential equilibrium is commonly thought to embody BI. As refinements of sequential equilibrium, quasi-perfection, and properness are presumed to inherit this property and, therefore, to satisfy BI in general.

Notice two key steps in this argument. The first is that there is a ‘direct’ definition of BI, and sequential equilibrium satisfies the definition. The second is that, if a solution concept satisfies BI, then a refinement of the concept also satisfies BI. Does each of these steps hold up to scrutiny?

Take the first step. What do we mean by a direct definition of BI? We mean a test on a solution concept so that a solution concept satisfies BI if and only if it passes the test. Our position—which is admittedly controversial—is that it is precisely such a test that has been missing from the literature.

The early literature attempted to provide such a test, but fell short of providing a conclusive definition. (Section 3 reviews earlier proposals. The conclusion discusses one expert’s view on the provision of such a test.) Instead, the modern literature turned to, what we would call, an ‘indirect’ definition. By this we mean a test for whether a given solution concept satisfies BI in terms of other solution concepts (often sequential equilibrium, but also other solution concepts), as opposed to a test based on the solution concept itself. For example, a solution concept is often said to satisfy BI if each component of the solution contains a sequential-equilibrium outcome. (See, e.g., the survey by [Govindan and Wilson \(2008\)](#).) From the point of view of axiomatics, such an indirect definition seems to us to be rather unconventional, precisely because it refers to other solution concepts. More commonly, in axiomatics, when we ask whether a particular member of some family of objects satisfies a certain property, we do not look at other objects in the family.<sup>1</sup>

What, then, is our BI test? It is a requirement that the solution of the whole game be induced by the solution on each of the subgames—a property we will call difference (D). We will argue that D captures the essential idea underlying BI. We show that sequential equilibrium does satisfy D and so, in this sense, it satisfies BI.

Now, the second step: if a solution concept satisfies BI, then a refinement of the concept also satisfies BI. Given our definition of BI, this is false. We show that, while sequential equilibrium satisfies BI, quasi-perfection and properness do not. There is a basic non-monotonicity in whether or not a solution concept satisfies BI. (This is the case even if we restrict attention to a class of games satisfying a ‘no-ties’ condition.)

In light of this non-monotonicity, the question of the consistency of A and BI appears to be open. We investigate the extent to which the question does or does not, in fact, remain open.

The paper proceeds as follows. Section 2 gives the formalism. Section 3 defines the property of BI. It explores the behavior of this property on both perfect and imperfect-information games. It goes on to point to solution concepts that satisfy the property of BI. Section 4 shows the basic non-monotonicity in whether or not a solution concept satisfies BI. Section 5 investigates the extent to which A and BI are or are not consistent. Section 6 concludes by discussing the approach at a broader level.

---

<sup>1</sup>An analogy may help: To test whether a map  $f$  is continuous, we need look no further than  $f$  itself (and the relevant topologies). In particular, we do not need to look at other functions  $g$  in whatever family of functions we have in mind.

## 2 Formulation

We fix the following notation throughout. Given sets  $X_1, \dots, X_I$ , write  $X = \times_{i=1}^I X_i$  and  $X_{-i} = \times_{j \neq i} X_j$ . Likewise, given maps  $f_i : X_i \rightarrow Y_i$ ,  $i = 1, \dots, I$ , write  $f : X \rightarrow Y$  for the product map, i.e.,  $f(x_1, \dots, x_I) = (f_1(x_1), \dots, f_I(x_I))$ . Define product maps  $f_{-i} : X_{-i} \rightarrow Y_{-i}$  analogously. If  $X$  is either a finite or a closed subset of  $\mathbb{R}^n$ , let  $\mathcal{M}(X)$  be the set of Borel probability measures on  $X$ . Write  $\text{Supp } \mu$  for the support of  $\mu$ .

**Extensive-Form Game** We consider finite extensive forms of perfect recall, without chance moves. Write  $\Gamma$  for such a game. We take the definition of  $\Gamma$ , as in [Kuhn \(1950\)](#)-[Kuhn \(1953\)](#), with the exception that we allow a non-terminal node to have only one outgoing branch (rather than two).

Let  $N$  (resp.  $Z$ ) be the set of non-terminal (resp. terminal) nodes of such an extensive-form. The players are labelled  $i = 1, \dots, I$ . Write  $H_i$  for the family of information sets for player  $i$  and  $H = \bigcup_{i=1}^I H_i$  for the family of all information sets. (Recall, under the Kuhn definition of a game, an information set is a subset of  $N$ .) Write  $M_i[h]$  for the set of moves  $m$  available to  $i$  at  $h \in H_i$ . (Recall, under the Kuhn definition of a game, a move is a subset of  $N$ .) A pure strategy  $s_i$  for player  $i$  maps each  $h \in H_i$  to some  $m_i \in M_i[h]$ . Let  $\Pi_i : Z \rightarrow \mathbb{R}$  be the payoff function for player  $i$ . The **outcome map**  $\Pi : Z \rightarrow \mathbb{R}^I$  is given by  $\Pi(z) = (\Pi_1(z), \dots, \Pi_I(z))$ .

The extensive-form game induces a strategic form: Write  $S_i$  (resp.  $\Sigma_i$ ) for the set of pure strategies (resp. mixed strategies) for player  $i$ . The map  $\zeta : S \rightarrow Z$  takes each pure-strategy profile into the terminal node it reaches. Write  $\pi_i : S \rightarrow \mathbb{R}$  for player  $i$ 's strategic-form payoff function, i.e.,  $\pi_i = \Pi_i \circ \zeta$ . Extend  $\pi_i$  to  $\Sigma_i \times \Sigma_{-i}$  in the usual way.

Say a pure strategy  $s_i \in S_i$  **allows** an information set  $h$  if there exists some  $s_{-i} \in S_{-i}$  so that the path induced by  $(s_i, s_{-i})$  passes through  $h$ . Say  $\sigma_i \in \Sigma_i$  (resp.  $\sigma_{-i} \in \Sigma_{-i}$ ) **allows** an information set  $h$  if there is some  $s_i$  with  $\sigma_i(s_i) > 0$  (resp.  $s_{-i}$  with  $\sigma_{-i}(s_{-i}) > 0$ ) such that  $s_i$  (resp.  $s_{-i}$ ) allows  $h$ . Say  $\sigma_i \in \Sigma_i$  (resp.  $\sigma_{-i} \in \Sigma_{-i}$ ) **fully allows** an information set  $h$  if, for each  $s_i$  with  $\sigma_i(s_i) > 0$  (resp.  $s_{-i}$  with  $\sigma_{-i}(s_{-i}) > 0$ ),  $s_i$  (resp.  $s_{-i}$ ) allows  $h$ . Write  $\Sigma_i(h)$  (resp.  $\Sigma_{-i}(h)$ ) for the set of strategies  $\sigma_i$  (resp.  $\sigma_{-i}$ ) that fully allow  $h$ . (Note carefully that we abuse notation here, since  $\Sigma_{-i}(h)$  need not be a product set.)

Say a strategy profile  $(\sigma_i, \sigma_{-i})$  **allows a move**  $m \in M_i[h]$  if there exists some strategy profile  $s$  with  $\sigma(s) > 0$ , so that the path induced by  $s$  passes through  $h$  and  $m$  is played with strictly positive probability under  $\sigma_i$ . Given a subset of strategy profiles  $Q \subseteq \Sigma$ , say  $Q$  **allows a move**  $m$  if there is some  $\sigma \in \Sigma$  that allows  $m$ .

**Outcome Equivalence** Terminal nodes  $z, \tilde{z} \in Z$  are **outcome equivalent** if  $\Pi(z) = \Pi(\tilde{z})$ . At times, we will consider payoff functions that satisfy a no-ties condition:

**Definition 2.1** *A game  $\Gamma$  satisfies the **Single-Payoff Condition (SPC)** if, for all  $z, \tilde{z} \in Z$ , the following holds: If  $i$  moves at the last common predecessor of  $z$  and  $\tilde{z}$ , then  $\Pi_i(z) = \Pi_i(\tilde{z})$  implies  $\Pi(z) = \Pi(\tilde{z})$ .*

In words, a game satisfies SPC if, whenever player  $i$  is indifferent between two terminal nodes over which he is decisive, those two terminal nodes are outcome equivalent. It is clear that in a perfect-information (PI) game satisfying SPC, there is a unique BI outcome. Moreover, SPC appears to be a minimal requirement for this purpose.

We will also be interested in outcome equivalence for strategy profiles: A strategy profile  $\sigma \in \Sigma$  induces a distribution over outcomes, viz., the measure in  $\mathcal{M}(\mathbb{R}^I)$  given by the image measure of  $\sigma$  under  $\Pi \circ \zeta$ . In

particular, the probability of outcome  $x \in \mathbb{R}^I$  is  $\sigma((\Pi \circ \zeta)^{-1}(x))$ . Call strategy profiles  $\sigma$  and  $\tilde{\sigma}$  **outcome equivalent** if they induce the same distribution on outcomes. Note, we can (and do) define this notion of outcome equivalence, even when  $\sigma$  and  $\tilde{\sigma}$  are strategy profiles in two (possibly different)  $I$ -player games.

Given subsets of strategy profiles  $Q \subseteq \Sigma$  and  $\tilde{Q} \subseteq \Sigma$  (of two, possibly different,  $I$ -player games), say that  $Q$  **induces the outcomes in**  $\tilde{Q}$  if, for each  $\tilde{\sigma} \in \tilde{Q}$ , there is some  $\sigma \in Q$  such that  $\sigma$  and  $\tilde{\sigma}$  are outcome equivalent. Call  $Q$  and  $\tilde{Q}$  **outcome equivalent** if  $Q$  induces the outcomes in  $\tilde{Q}$ , and  $\tilde{Q}$  induces the outcomes in  $Q$ .

**Solution Concept** A **solution concept**  $\mathcal{S}$  associates with each game  $\Gamma$  a family of subsets of strategy profiles for  $\Gamma$ . Formally, a solution concept  $\mathcal{S}$  maps each game to a family of subsets of strategy profiles for the game, i.e.  $\mathcal{S}(\Gamma) \subseteq \prod_{i=1}^I 2^{\Sigma_i}$ . The family  $\mathcal{S}(\Gamma)$  is called the **solution** of  $\Gamma$ . Each element of  $\mathcal{S}(\Gamma)$ , i.e., each subset of mixed-strategy profiles  $Q \in \mathcal{S}(\Gamma)$ , is called a **component** of the solution. A solution concept  $\mathcal{R}$  is a **refinement** of  $\mathcal{S}$  if, for each game  $\Gamma$  and every  $R \in \mathcal{R}(\Gamma)$ , there is a  $Q \in \mathcal{S}(\Gamma)$  so that  $Q$  induces the outcomes in  $R$ .

Two examples of how a solution concept is modeled: First, we can take the solution concept of Nash equilibrium to map each game to multiple components, where each component is a singleton and consists of a particular Nash equilibrium. Second, we can take the solution to consist of a single component—e.g., all the iterated undominated profiles.

**Remark 2.1** The tradition—following [Kohlberg and Mertens \(1986\)](#) and their successors—is to model the Nash equilibrium as a solution concept that has finitely many components, each of which consists of a connected set of Nash equilibria. Thus, they instead model an equilibrium-based solution concept,  $\mathcal{S}$ , as mapping each game to finitely many non-product subsets of strategy profiles.

We take an alternate route, modeling an equilibrium-based solution concept as consisting of singleton (product) components. Let us preview why we take this alternate route: For the purpose of defining backward induction, we will use the concept of a component to specify a player’s predictions about the strategies her co-players’ choose. To fix ideas, consider a game with two distinct Nash equilibria, viz.  $\sigma = (\sigma_1, \dots, \sigma_I)$  and  $\rho = (\rho_1, \dots, \rho_I)$ . When we focus on  $\sigma$ , we implicitly assume that each player  $i$  thinks others play according to  $\sigma_{-i}$ . The strategies  $\rho_{-i}$  are irrelevant for the purpose of specifying player  $i$ ’s predictions about the strategies of her co-players, given the focus on the equilibrium  $\sigma$ .

For the specific case of an equilibrium-based solution concept, there is a natural alternate choice of a player’s prediction: We could collect all equilibrium strategy profiles into a single component (or finitely many connected components) and define a local concept of a prediction, i.e., if  $(\sigma_i, \sigma_{-i})$  is an equilibrium, then  $\sigma_{-i}$  is player  $i$ ’s prediction about the strategies her co-players’ choose. When using this concept of prediction (in our definition of backward induction to come), all subsequent results follow through.<sup>2</sup> While this alternate approach works for equilibrium-based concepts, it would not work for non-equilibrium based concepts. Our position is that the idea of backward induction is not fundamentally tied to equilibrium.<sup>3</sup>

Of course, our approach does not mitigate the fact that, for other purposes, it may be quite important to focus on an object that, in the context of our paper, is a collection of unions over components. Our

<sup>2</sup>Of course, here we would amend the definition of a solution concept to map a game into a set of non=product subsets of strategies.

<sup>3</sup>There is a large literature that investigates the relationship between backward induction and non-equilibrium solution concepts. See, e.g., [Nachbar \(1992\)](#), [Battigalli \(1997\)](#), [Cressman and Schlag \(1998\)](#), [Samuelson \(1998\)](#), [Hart \(2002\)](#), [Perea \(2010\)](#), [Chen and Micali \(2012\)](#), among many others.

approach does not subtract from such an endeavor. Different modeling choices can be made for different purposes.  $\square$

**Remark 2.2** We have defined solution concepts in terms of mixed strategies. Of course, some solution concepts are defined in terms of pure strategies (e.g., extensive-form rationalizability) and some solution concepts are defined using behavioral strategies (e.g., sequential equilibrium). When needed, we will understand all definitions in terms of pure or behavioral strategies. We will use the notation  $\beta_i$  for a behavioral strategy for player  $i$ .  $\square$

### 3 Backward Induction

The intuitive idea behind BI is clear: Fix a game  $\Gamma$  and a subgame  $\Delta$  of  $\Gamma$ . Now discard  $\Delta$ , leaving behind only the solution on this subgame—leaving behind the ‘ghost’ of the subgame, if you like. Then, we do not change our original analysis.

While the idea is intuitively clear, it is less clear how to formalize it. There have been a number of attempts at such a definition, especially in the earlier literature. (See [Kohlberg and Mertens \(1986, pp.1012-1013\)](#) and [Hillas and Kohlberg \(2002\)](#).) Here, we argue that the correct definition is given by, what we call, the Difference property. We also argue that alternative definitions proposed by the literature do not succeed in capturing BI. Moreover, the Difference definition does not suffer from the same drawbacks as these alternative definitions.

#### Defining Backward Induction

To formalize the idea of BI, we first need to specify what it means to delete a subgame, leaving behind only the solution on the subgame. The relevant concept goes back to [Kuhn \(1953, p.208\)](#); we will call it a **difference game**. A difference game is defined relative to a solution concept  $\mathcal{S}$ . Begin with a game  $\Gamma$  and a subgame  $\Delta$  of  $\Gamma$ . (Note,  $\Delta$  need not be a proper subgame, i.e., it may be  $\Gamma$  itself.) Fix a nonempty component of  $\mathcal{S}(\Delta)$ , which we will denote  $Q^\Delta$ . The  $(\mathcal{S}, Q^\Delta)$ -difference game is obtained by deleting from the original game  $\Gamma$  any move not allowed by  $Q^\Delta$ . It is readily verified that each  $(\mathcal{S}, Q^\Delta)$ -difference game is a well-defined game. (This uses the fact that we required  $Q^\Delta$  to be nonempty.) Write  $\Gamma_{\mathcal{S}, Q^\Delta}$  for the  $(\mathcal{S}, Q^\Delta)$ -difference game. Note, the difference game depends on the solution concept, subgame, and particular component of the solution on the subgame.

We can now state our key property, which we refer to as the Difference Property (D):

- (D) A solution concept  $\mathcal{S}$  satisfies **Difference** (on  $\mathcal{G}$ ) if, for each game  $\Gamma$  (in  $\mathcal{G}$ ), and each subgame  $\Delta$  of  $\Gamma$  the following holds: If  $Q \in \mathcal{S}(\Gamma)$ , there is a nonempty component  $Q^\Delta \in \mathcal{S}(\Delta)$  and a component  $\tilde{Q} \in \mathcal{S}(\Gamma_{\mathcal{S}, Q^\Delta})$ , so that  $\tilde{Q}$  induces outcomes in  $Q$ .

Difference bears a similarity to property (BI3) in [Kohlberg and Mertens \(1986, p.1012\)](#). (The Appendix discusses the relationship further.) Loosely, it says that the solution on the whole game should be included in the solution on what is left after replacing a subgame with what the solution allows on the subgame.

The Difference property works in much the same way as the backward induction algorithm. The algorithm works by using what it prescribes on future subgames to pin down behavior on the current subgame. The Difference property applies this same principle to general extensive-form games. Solutions on subgames yield difference games, which are used to pin down the solution on the overall game. Formally,

each (distribution on) outcome(s) allowed by the solution on the overall game must also be allowed by the solution on some difference game.

**Remark 3.1** A technical remark: Note that we say that a solution  $\mathcal{S}$  may satisfy D on  $\mathcal{G}$ , even if there is some  $\Gamma \in \mathcal{G}$  and some subgame  $\Delta$  of  $\Gamma$ , so that  $\Delta \notin \mathcal{G}$ . No confusion should result.  $\square$

Note, by itself, D is missing two basic ingredients, which are built into the backward induction algorithm: existence and optimization. Thus, we will impose D in the presence of two background axioms. First is existence:

(E) A solution concept  $\mathcal{S}$  satisfies **Existence** (on  $\mathcal{G}$ ) if, for each game  $\Gamma$  (in  $\mathcal{G}$ ), there is a nonempty component of  $\mathcal{S}(\Gamma)$ .

Now turn to the idea of optimization. The basic idea will be that a strategy  $\sigma_i$  is rational if it maximizes player  $i$ 's subjective expected payoffs at each information set that the strategy allows. That is, at each information set  $h$  allowed by  $\sigma_i$ , there is some assessment  $\rho_{i,h}$  about the how others play the game—concentrated on the event that information set  $h$  is reached—so that the conditional distribution  $\sigma_i(\cdot|S_i(h))$  maximizes player  $i$ 's subjective expected payoffs under  $\rho_{i,h}$ . Since it is assumed that all players choose mixed strategies, each  $\rho_{i,h}$  is (by definition) a probability measure on  $\Sigma_{-i}(h)$ .<sup>4</sup>

Extend  $\pi_i$  to  $\Sigma_i \times \mathcal{M}(\Sigma_{-i}) \rightarrow \mathbb{R}$  in the usual way. Say a strategy  $\sigma_i^*$  is **optimal** under  $\rho_i \in \mathcal{M}(\Sigma_{-i})$  among strategies in  $Q_i \subseteq \Sigma_i$ , if  $\sigma_i^* \in Q_i$  and  $\pi_i(\sigma_i^*, \rho_i) \geq \pi_i(\sigma_i, \rho_i)$  for each  $\sigma_i \in Q_i$ . Fix some  $Q_{-i} \subseteq \Sigma_{-i}$  Borel. Say a strategy  $\sigma_i^*$  is  **$Q_{-i}$ -rational** if, for each information set  $h \in H_i$  allowed by  $\sigma_i^*$ , there is some  $\rho_{i,h} \in \mathcal{M}(\Sigma_{-i}(h))$  satisfying the following:

$\sigma_i^*(\cdot|S_i(h))$  is optimal under  $\rho_{i,h}$  among strategies in  $\Sigma_i(h)$ , and

if  $Q_{-i} \cap \Sigma_{-i}(h) \neq \emptyset$ , then  $\rho_{i,h}(Q_{-i}) = 1$ . If  $\sigma_i$  is  $\Sigma_{-i}$ -rational, we will simply say that  $\sigma_i$  is **rational**. Note, if  $\sigma_i$  is  $Q_{-i}$ -rational for some  $Q_{-i}$ , then  $\sigma_i$  is rational.

(CR) A solution concept  $\mathcal{S}$  satisfies **Component-Wise Rationality** (on  $\mathcal{G}$ ) if, for each game  $\Gamma$ , (in  $\mathcal{G}$ ) and each component  $Q = \prod_{i=1}^I Q_i \in \mathcal{S}(\Gamma)$ , the following criterion is satisfied: If  $\sigma_i \in Q_i$ , then  $\sigma_i$  is  $Q_{-i}$ -rational.

Component-Wise Rationality asks that strategies in a component  $Q$  of a solution are rational with respect to the other players' strategies in the component. Note that, imposing CR does not confine the solution concept to be an equilibrium refinement. In particular, CR does not require that players have a 'correct assessment' about the strategies played. Consider the solution concept of extensive-form rationalizability, i.e., the solution concept  $\mathcal{S}_{\text{EFR}}$  that maps each game  $\Gamma$  to a single component corresponding to the set of (pure strategy) extensive-form rationalizable strategies of  $\Gamma$ . Then,  $\mathcal{S}_{\text{EFR}}$  satisfies CR. But, there exists a game  $\Gamma$ , so that the following holds:  $\mathcal{S}_{\text{EFR}}(\Gamma) = \{\prod_{i=1}^I Q_i\} \subseteq \prod_{i=1}^I 2^{S_i}$  has some  $s_j \in Q_j$  so that  $s_j$  is not optimal under any  $s_{-j} \in Q_{-j}$ .

**Remark 3.2** In defining CR, we implicitly allow players to hold a correlated assessment about how others play the game. Particular solution concepts satisfy a requirement of CR, relative to an independent assessment requirement. (See, e.g., Battigalli (1996), Battigalli and Veronesi (1996), Kohlberg and Reny (1997), and Swinkels (1994) on defining independence in extensive-form games.)

<sup>4</sup>Note, carefully, the subscript  $i$  refers to an object of player  $i$ . So,  $\sigma_i$  is a strategy of player  $i$  and  $\rho_{i,h}$  is an assessment of player  $i$  (about players  $j \neq i$ ).

The question of correlation vs. independence has a long history in game theory. While the question is important, our viewpoint is that it is orthogonal to the question of backward induction. Thus, we do not preclude players from holding correlated assessments about how others play the game.  $\square$

**Definition 3.1** *Say that a solution concept satisfies BI if it satisfies D, E, and CR.*

## Background Check: Perfect-Information Games

We begin with a basic background requirement: A solution concept satisfies Difference, Existence, and Component-Wise Rationality on the domain of PI games satisfying SPC if and only if it is outcome equivalent to the unique backward induction outcome on this set of games. Write  $\mathcal{G}_{PI-SPC}$  for the class of PI games satisfying SPC.

**Proposition 3.2** *Fix a solution concept  $\mathcal{S}$ .*

- (i) *If  $\mathcal{S}$  satisfies D, E, and CR on  $\mathcal{G}_{PI-SPC}$  then, for each  $\Gamma \in \mathcal{G}_{PI-SPC}$ , each component of  $\mathcal{S}(\Gamma)$  is outcome equivalent to the backward induction algorithm on  $\Gamma$ .*
- (ii) *If, for each  $\Gamma \in \mathcal{G}_{PI-SPC}$ , each component of  $\mathcal{S}(\Gamma)$  is outcome equivalent to the backward induction algorithm on  $\Gamma$ , then  $\mathcal{S}$  satisfies D, E, and CR on  $\mathcal{G}_{PI-SPC}$ .*

Proposition 3.2 says that D, E, and CR characterize BI on the domain of games satisfying SPC. Part (ii) is standard. Part (i) is key: In part (ii), we could replace Difference with alternative axioms put forward in the literature. But, we will soon point out that we cannot do the same for part (i).

**Proof of Proposition 3.2.** Begin with Part (i). The proof is by induction on the length of the game. For a game of length 1, the result is immediate from E, CR, and the fact that the game satisfies SPC.

Assume the statement holds for any game of length  $l$  or less. Fix a game of length  $l + 1$ , where  $i$  moves first and write  $\Delta^k$ ,  $k = 1, \dots, K$ , for the immediate subgames. For each such subgame  $\Delta^k$ , fix a component  $Q^k$  of  $\mathcal{S}(\Delta^k)$ . Using the induction hypothesis,  $Q^k \neq \emptyset$  and any  $(\sigma_1^k, \dots, \sigma_I^k) \in Q^k$  gives the unique backward induction outcome on that subgame.

Consider the game obtained by deleting from each immediate subgame  $\Delta^k$  any move not allowed by  $Q^k$ . Call this game  $\Gamma_{\mathcal{S}}$ . Then, in  $\Gamma_{\mathcal{S}}$ , each of  $i$ 's choices  $k = 1, \dots, K$  leads to a unique outcome in the associated subgame. Of course, these outcomes do not depend on the particular choices of  $Q^k$ .

Using SPC, all rational strategies for  $i$  (in  $\Gamma_{\mathcal{S}}$ ) are outcome equivalent. By E, there is a nonempty component of  $\mathcal{S}(\Gamma_{\mathcal{S}})$ . By CR, any nonempty component of  $\mathcal{S}(\Gamma_{\mathcal{S}})$  must be outcome equivalent to the backward induction algorithm applied to  $\Gamma_{\mathcal{S}}$ . It follows that any nonempty component of  $\mathcal{S}(\Gamma_{\mathcal{S}})$  must be outcome equivalent to the backward induction algorithm applied to  $\Gamma$ .

Now, successively apply D to each subgame, so that any outcome allowed by any component of the solution on the overall game must be allowed by the component on  $\Gamma_{\mathcal{S}}$ . (This uses the fact that there is a unique outcome in this difference game and this outcome does not depend on the initial choice of solutions  $Q^k$ .) It follows that any outcome allowed by any component of the solution on the overall game must be the backward induction outcome on that game. By E, the solution must have some nonempty component, establishing part (i).

Turn to Part (ii). Fix a solution concept  $\mathcal{S}$ , as in the premise. It is immediate that  $\mathcal{S}$  satisfies E and CR. We show D. Fix a PI game  $\Gamma$  satisfying SPC, so that there is a unique backward induction outcome.



Let  $Q \in \mathcal{S}(\Gamma)$  and note that, by assumption, each strategy profile in  $Q$  induces the backward induction outcome in  $\Gamma$ . Fix a subgame  $\Delta$ , a component  $Q^\Delta \in \mathcal{S}(\Delta)$ , and a component  $\tilde{Q} \in \Gamma_{\mathcal{S}, Q^\Delta}$ . Since the difference game, viz.  $\Gamma_{\mathcal{S}, Q^\Delta}$ , is also a PI game satisfying SPC, it has a unique backward induction outcome and any profile in  $\tilde{Q}$  must be outcome equivalent to this backward induction outcome. Now D follows from the fact that the backward induction outcome on  $\Gamma$  coincides with the backward induction outcome of  $\Gamma_{\mathcal{S}, Q^\Delta}$ . ■

To better appreciate Proposition 3.2, consider three alternate proposals:

- A solution concept  $\mathcal{S}$  satisfies **Projection** (on  $\mathcal{G}$ ) if for each game  $\Gamma$  (in  $\mathcal{G}$ ) the following holds: For each subgame  $\Delta$  and component  $Q \in \mathcal{S}(\Gamma)$ , there is a component  $Q^\Delta \in \mathcal{S}(\Delta)$  such that for each  $\sigma \in Q$ , the restriction of  $\sigma$  to the subgame  $\Delta$  is contained in  $Q^\Delta$ .
- A solution concept  $\mathcal{S}$  satisfies **Reverse Projection** (on  $\mathcal{G}$ ) if for each game  $\Gamma$  (in  $\mathcal{G}$ ) the following holds: For each subgame  $\Delta$  and component  $Q \in \mathcal{S}(\Gamma)$ , there is a component  $Q^\Delta \in \mathcal{S}(\Delta)$  such that, for each  $\sigma^\Delta \in Q^\Delta$ , there is some  $\sigma \in Q$ , so that  $\sigma^\Delta$  is the restriction of  $\sigma$  to  $\Delta$ .
- A solution concept  $\mathcal{S}$  satisfies **Reverse Difference** (on  $\mathcal{G}$ ) if for each game  $\Gamma$  (in  $\mathcal{G}$ ) and each subgame  $\Delta$  of  $\Gamma$  the following holds: If  $Q \in \mathcal{S}(\Gamma)$ , there is a nonempty component  $Q^\Delta \in \mathcal{S}(\Delta)$  and a component  $\tilde{Q} \in \mathcal{S}(\Gamma_{\mathcal{S}, Q^\Delta})$ , such that  $Q$  induces the outcomes in  $\tilde{Q}$ .

Each of these properties are aimed at relating the solution on the whole game to the solution on parts of the game: Projection asks that the solution on the whole game be pinned down by the solution on the subgame, i.e., as opposed to the solution on the difference game. (Projection is property (BI1) in Kohlberg and Mertens (1986, p.1012).) Reverse Projection asks that the solution on the subgame be pinned down by the solution of the whole game. (Reverse Projection is property (BI2) in Kohlberg and Mertens (1986, p.1012).) Reverse Difference asks that the solution on the difference game be pinned down by the solution on the whole game.

**Example 3.1** Let  $\mathcal{G} = \{\Gamma, \Delta, \tilde{\Gamma}\}$ , as in Figures 3.1a-3.1c. These PI trees satisfy SPC. Let  $\mathcal{S}$  be a solution concept with  $\mathcal{S}(\Gamma) = \{\{Out, In\} \times \{L\}\}$ ,  $\mathcal{S}(\Delta) = \{\{L\}\}$ , and  $\mathcal{S}(\tilde{\Gamma}) = \{\{(Out, L)\}\}$ . Notice that  $\Delta$  is a subgame of  $\Gamma$  and  $\tilde{\Gamma}$  is the  $(\mathcal{S}, \{L\})$ -difference tree. So,  $\mathcal{S}$  satisfies Projection, Reverse Projection, Reverse Difference, Component-Wise Rationality, and Existence on  $\mathcal{G}$ . Of course,  $\mathcal{S}$  fails Difference on  $\mathcal{G}$ .

The problem here is clear: The idea of BI is that solutions on parts of the game should be used to pin down the solution on the whole game. If the notion of a “part” is a difference game (as in Reverse Difference), but we instead require the reverse inclusion to D, then the solutions on difference games do not pin down the solution on the whole game. Likewise, if the notion of a “part” is a subgame (as in Projection), then, again, the solutions on the subgames do not pin down the solution on the whole game. To see this, return to the example. While the solution  $\mathcal{S}(\Delta)$  is used to pin down Bob’s behavior in  $\Gamma$ , it cannot be used to pin down Ann’s behavior in  $\Gamma$ , because she has no move in  $\Delta$ . Under D, even if Ann has no move in  $\Delta$ , we can use a component  $Q^\Delta$  of the solution on  $\Delta$  to pin down Ann’s behavior in  $\Gamma$ , since Ann does have a move in the associated difference game.

Proposition 3.2 shows that Difference does not suffer from these same flaws. That said, D, E, and CR only give the BI outcomes and not the BI strategies, as the next example illustrates.



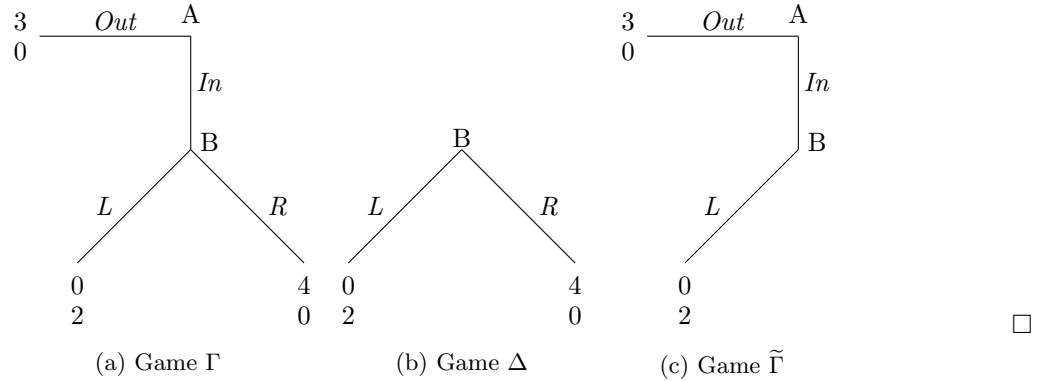


Figure 3.1

**Example 3.2** Consider the game  $\Gamma$  in Figure 3.2, which is Figure 3 in Reny (1992, p.637). There, the backward induction algorithm gives the strategies  $(Out-Down, Out-Down)$ . Consider the solution concept of extensive-form rationalizability, viz.  $\mathcal{S}_{EFR}$ . Proposition 3.4, to come, will show that this concept satisfies D, E, and CR. But,  $\mathcal{S}_{EFR}(\Gamma) = \{\{Out-Down, Out-Across\} \times \{In-Down\}\}$ . So, the set of extensive-form strategies are disjoint from the backward induction strategies, even though they both imply the same path of play.

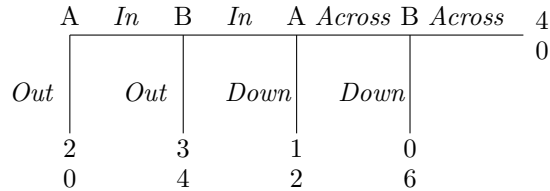


Figure 3.2: Backward Induction Outcomes vs. Strategies

Example 3.2 shows that, on PI games satisfying SPC, properties D, E, and CR are not sufficient to give the backward induction strategies. One might conjecture that this is an artifact of defining D as a requirement outcomes and not strategies—i.e., by requiring that *outcomes* of the solution on the whole tree are pinned down by *outcomes* of the solution on the difference tree and not imposing an analogous requirement on *strategies*. (Such a strategy-wise definition would be closer to Kohlberg and Mertens’s (1986) (BI3).) But, the Appendix argues that this is not the case. In particular, replacing D with a strategy-wise concept does not allow us to repeat the proof of Proposition 3.2 and come to a stronger conclusion, i.e., that the backward induction strategies are played.

Finally, we note that, in a PI game that violates SPC, we may not get an outcome consistent with the backward induction algorithm. The reason is that the algorithm implicitly imposes an equilibrium requirement, which is not imposed by our definition.

**Example 3.3** Consider the game  $\Gamma$  in Figure 3.3, which violates SPC. The backward induction algorithm gives either the strategy profile  $(Out, L)$  or the strategy profile  $(In, R)$ . Consider the solution concept of extensive-form rationalizability. We pointed out that  $\mathcal{S}_{EFR}$  satisfies D, E, and CR. But,  $\mathcal{S}_{EFR}(\Gamma) = \{\{Out, In\} \times \{L, R\}\}$ .

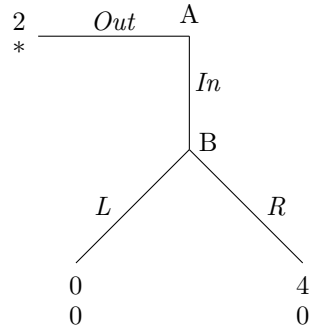


Figure 3.3: Violation of SPC

The key is that EFR does not require that players have a correct assessment about play. Thus, in  $\Gamma$ , Ann may play *In* expecting that Bob play the rational strategy *R*. But, her assessment need not be correct. This does not take away from the idea of backward induction: Ann does predict that Bob will play ‘rationally’ (or ‘according to the solution’) in the future. This is the essence of the idea of backward induction. The key is that the algorithm adds an additional equilibrium requirement above and beyond the idea of backward induction.  $\square$

While D, E, and CR need not give an outcome consistent with the backward induction algorithm on PI games that violate SPC, there is structure on the outcomes obtained. For instance, there are PI games that violate SPC, for which a backward induction outcome would obtain.

**Example 3.4** Consider the game  $\Gamma$  in Figure 3.4. If  $\mathcal{S}$  satisfies D, E, and CR on  $\mathcal{G}$  and  $\Gamma \in \mathcal{G}$ , then for each component  $Q$  of  $\mathcal{S}(\Gamma)$ ,  $Q$  is a non-empty subset of  $\{L, R\} \times \{(l-r')\}$ . Of course, the backward induction algorithm gives strategy profiles  $(L, l-r')$  and  $(R, l-r')$ .  $\square$

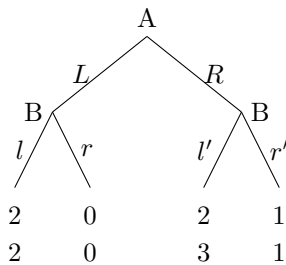


Figure 3.4: Another Violation of SPC

## Beyond Perfect-Information Games: Examples

We turn to two examples, which will illustrate how our definition of BI works in games of imperfect information.

**Example 3.5** Let  $\Gamma$  be Battle of the Sexes with an Outside Option, as in Figure 3.5. Here backward induction is typically thought to lead to one of two outcomes—either A plays *Out* immediately or A

plays  $In-U$  and B plays  $L$ .<sup>5</sup> Indeed, if a pure-strategy solution concept  $\mathcal{S}$  satisfies D, E, and CR, then  $\mathcal{S}(\Gamma) \subseteq \{(Out, R)\}, \{(Out, L), (Out, R)\}, \{(In-U, L)\}$ .<sup>6</sup>

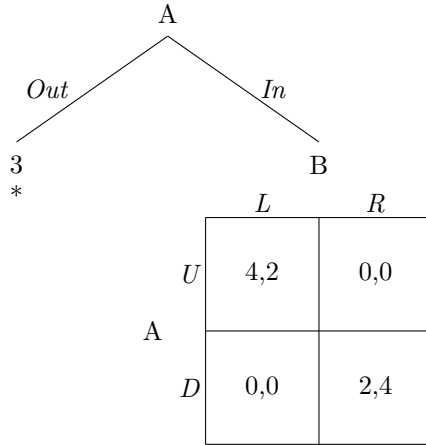


Figure 3.5: Battle of the Sexes with an Outside Option

Fix some  $Q_a \times Q_b \in \mathcal{S}(\Gamma)$ . By repeated application of CR and E,

$$Q_a \times Q_b \in \{(Out, R)\}, \{(Out, L), (Out, R)\}, \{(In-U, L)\}.$$

To see this claim, note that, by CR,  $Q_a$  does not contain  $In-D$ . First suppose that  $In-U \in Q_a$ . By CR and E,  $Q_b = \{L\}$ . Then, again using CR,  $Q_a = In-U$ . Next, suppose that  $In-U \notin Q_a$ , i.e.,  $Q_a = \{Out\}$ . By CR,  $Q_b \neq \{L\}$ .

Notice, the analysis did not make use of D. In fact, D need not impose further restrictions. To understand why, write  $\Delta$  for the Battle of the Sexes subgame. Consider a pure-strategy solution concept with  $\mathcal{S}(\Gamma) = \{(Out, R)\}, \{(Out, L), (Out, R)\}, \{(In-U, L)\}$  and  $\mathcal{S}(\Delta) = \{(U, L)\}, \{(D, R)\}, \{U, R\} \times \{L, R\}$ . This solution concept satisfies D, E, and CR on the domain  $\mathcal{G} = \{\Gamma, \Delta\}$ .  $\square$

**Example 3.6** Consider the game  $\Gamma$  in Figure 3.6.<sup>7</sup> This is a game with no proper subgames. But, still we will see that, in the presence of E and CR, D has bite. In particular, if  $\mathcal{S}$  satisfies D, E, and CR, then  $\mathcal{S}(\Gamma) = \{(L-x, l)\}$ .

To see the claim, fix some  $Q = Q_a \times Q_b \in \mathcal{S}(\Gamma)$ . Now apply D: We must be able to find some  $Q^\Gamma = Q_a^\Gamma \times Q_b^\Gamma$  and some  $\tilde{Q} = \tilde{Q}_a \times \tilde{Q}_b \in \mathcal{S}(\Gamma_{Q^\Gamma})$  so that  $\tilde{Q}$  induces outcomes in  $Q$ . By CR and E applied to the game  $\Gamma$ ,  $Q_a^\Gamma$  is a non-empty subset of  $\{L-x, R-x\}$ . It follows from CR and E applied to the  $(\mathcal{S}, Q^\Gamma)$ -difference game that  $\tilde{Q}_b = \{l\}$ . Thus, applying D and E to  $\Gamma$ ,  $Q_b = \{l\}$ . Now by CR applied to the game  $\Gamma$ ,  $Q_a = \{L-x\}$ , as claimed.  $\square$

Example 3.6 illustrates that our definition of BI has bite, even if a game has no proper subgame. It also illustrates the importance of taking CR as our definition of optimization. To see this last claim, compare CR with the following weaker property:

<sup>5</sup>Note, it is *forward induction* that is typically thought to rule out the case that Ann plays *Out*.

<sup>6</sup>The focus on pure strategies is only for notational convenience. An analogous argument can be made for a solution concept based on mixed or behavioral strategies.

<sup>7</sup>We thank a referee for this example and for pointing out a flaw in our earlier definition of BI, which only required R (below).

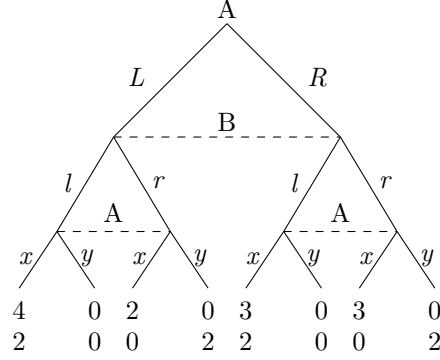


Figure 3.6: A Game with No Proper Subgames

(R) A solution concept  $\mathcal{S}$  satisfies **Rationality** (on  $\mathcal{G}$ ) if, for each game  $\Gamma$ , (in  $\mathcal{G}$ ) and each component  $\prod_{i=1}^I Q_i \in \mathcal{S}(\Gamma)$ , the following criterion is satisfied: If  $\sigma_i \in Q_i$ , then  $\sigma_i$  is rational.

Fix a solution concept  $\mathcal{S}$  that satisfies D, E, and R—but not necessarily CR. When we apply the solution concept to the game  $\Gamma$  in Figure 3.6, we get that  $\mathcal{S}(\Gamma) \subseteq \{(L-x, l), (R-x, l)\}$ . This fact can be seen by repeating the argument in Example 3.6 up to (but excluding) the last sentence. The last sentence does not hold under R. In fact, the solution may very well specify  $\mathcal{S}(\Gamma) = \{(R-x, l)\}$ : Focus on D. We can construct a difference game by taking  $Q^\Gamma = \{(R-x, l)\}$ . In this difference game, no player has a non-trivial choice and, so, R is trivially satisfied. Then, we can take then  $\tilde{Q} = Q^\Gamma$  and there is no conflict with D.

Notice, in the above,  $R-x$  is rational in the whole game, i.e., if Ann assigns probability 1 to  $r$  at each history. But, if Ann understands that the solution on the difference game pins down the solution on the whole game, she would best respond by playing  $L-x$ . Property R does not require that a player has an assessment consistent with the idea that “the solution on the difference game pins down behavior on the whole game.” Property CR imposes this additional requirement.

**Remark 3.3** Property R is an analogue of requirement (BI0) in Kohlberg and Mertens (1986). (See the Appendix.) A solution concept  $\mathcal{S}$  satisfies D, E, and CR on  $\mathcal{G}_{\text{PI-SPC}}$  if and only if it satisfies D, E, and R on  $\mathcal{G}_{\text{PI-SPC}}$ . The “only if” is immediate. For the “if,” suppose  $\mathcal{S}$  satisfies D, E, and R on  $\mathcal{G}_{\text{PI-SPC}}$ . Repeating the argument in the proof of Proposition 3.2(i), for each  $\Gamma \in \mathcal{G}_{\text{PI-SPC}}$ , each component of  $\mathcal{S}(\Gamma)$  is outcome equivalent to the BI algorithm on  $\Gamma$ . Then the claim follows from Proposition 3.2(ii).  $\square$

## Beyond Perfect-Information Games: Solution Concepts

We now point to two solution concepts, which both satisfy D, E, and CR. For the connection to D, we need some more notation.

Fix a game  $\Gamma$  and a subgame  $\Delta$  of  $\Gamma$ . Note that the set of information sets and moves in  $\Delta$  are subsets of the information sets and moves in  $\Gamma$ . Write  $S_i$  (i.e., without qualification) for the set of  $i$ 's strategies in  $\Gamma$  and  $S_i^{[\Delta]}$  for the set of  $i$ 's strategies in the subgame  $\Delta$ . Given some  $s_i^\Delta \in S_i^\Delta$ , write  $[s_i^\Delta]$  for the set of strategies  $s_i \in S_i(\Delta)$  (i.e., the set of strategies in  $\Gamma$  that allow subgame  $\Delta$ ) which coincide with  $s_i^\Delta$ .

Next, fix a solution concept  $\mathcal{S}$  and a difference game  $\tilde{\Gamma} = \Gamma_{\mathcal{S}, Q^\Delta}$ . Write  $\tilde{H}_i$  for the family of  $i$ 's information sets in the difference game  $\tilde{\Gamma} = \Gamma_{\mathcal{S}, Q^\Delta}$ . Note that there is an injective mapping  $\eta : \tilde{H} \rightarrow H$  with  $\tilde{h} \subseteq \eta(\tilde{h})$ . Write  $\tilde{M}_i[\tilde{h}_i]$  for the moves available to  $i$  at  $\tilde{h}_i$  in the difference game, and note that, for

each  $\tilde{h}_i$ , there is an injective mapping  $\xi[\tilde{h}_i] : \tilde{M}_i[\tilde{h}_i] \rightarrow M_i[\eta(\tilde{h}_i)]$  so that  $\tilde{m}_i \subseteq \xi[\tilde{h}_i](\tilde{m}_i)$ . If  $\tilde{s}_i$  is a pure strategy for  $i$  in the difference game, we write  $[\tilde{s}_i]$  for the set of pure strategies for  $i$  in  $\Gamma$  which coincide with  $\tilde{s}_i$  in the difference game.

**Sequential Equilibrium** Sequential equilibrium is typically viewed as embodying backward induction. We now show that, indeed, sequential equilibrium satisfies D, E, and CR.

Recall the definition of sequential equilibrium (Kreps and Wilson, 1982): Fix a pair  $(\beta, \mu)$ , where  $\beta$  is a profile of behavioral strategies and  $\mu$  is a system of beliefs. (That is:  $\mu : H \rightarrow \mathcal{M}(N)$  with each  $\mu(h)(h) = 1$ .) Say  $(\beta, \mu)$  is **consistent** if there is a sequence  $(\beta^k, \mu^k) \rightarrow (\beta, \mu)$  where each  $\beta^k$  is a profile of completely mixed behavioral strategies (i.e., for each  $i$  and  $h_i \in H_i$ ,  $\text{Supp } \beta_i^k(h_i) = M_i[h_i]$ ) and each  $\mu^k$  is derived from  $\beta^k$  by conditioning. The pair  $(\beta, \mu)$  is a **sequential equilibrium** if it is consistent and, for each  $i$ , every  $\beta_i(h_i)$  is optimal under  $\mu$  (among strategies in  $\Sigma_i(h_i)$ ). We define the sequential equilibrium solution concept  $\mathcal{S}_{SE}$  by

$$\mathcal{S}_{SE}(\Gamma) = \{ \{\beta\} : \text{there is a system of beliefs } \mu \text{ s.t. } (\beta, \mu) \text{ is a sequential equilibrium of } \Gamma \}.$$

**Proposition 3.3** *The solution concept  $\mathcal{S}_{SE}$  satisfies D, E, and CR.*

**Proof.** The fact that  $\mathcal{S}_{SE}$  satisfies E and CR is standard. We focus on D.

Fix a game  $\Gamma$  and some  $\beta = (\beta_1, \dots, \beta_I)$  with  $\{\beta\} \in \mathcal{S}_{SE}(\Gamma)$ . Then, there exists some system of beliefs  $\mu : H \rightarrow \mathcal{M}(N)$  such that  $(\beta, \mu)$  is a sequential equilibrium. Fix a subgame  $\Delta$ . For each information set  $h_i$  of  $\Delta$ , set  $\beta_i^\Delta(h_i) = \beta_i(h_i)$  and  $\mu^\Delta(h_i) = \mu(h_i)$ . It is immediate that  $(\beta^\Delta, \mu^\Delta)$  is a sequential equilibrium of  $\Delta$ , i.e.,  $\{\beta^\Delta\} \in \mathcal{S}_{SE}(\Delta)$ .

Construct the difference game  $\Gamma_{SE, \{\beta^\Delta\}}$  by deleting from  $\Gamma$  any path (in  $\Delta$ ) that is played with zero probability under  $\beta^\Delta$ . This amounts to deleting from  $\Gamma$  any path which is in  $\Delta$  and which is played with zero probability under  $\beta$ . So, certainly, each  $\beta_i(\eta(\tilde{h}_i))(\xi[\tilde{h}_i](\tilde{M}_i[\tilde{h}_i])) = 1$ . Moreover, if  $\eta(\tilde{h})$  is in  $\Delta$ , then  $\eta(\tilde{h})$  is reached with strictly positive probability under  $\beta^\Delta$ . So, in this case,  $\mu(\eta(\tilde{h}))(\tilde{h}) = \mu^\Delta(\eta(\tilde{h}))(\tilde{h}) = 1$ . Indeed, this is true more generally, i.e., for each  $\eta(\tilde{h})$  (whether or not it is in  $\Delta$ )  $\mu(\eta(\tilde{h}))(\tilde{h}) = 1$ . We use these facts repeatedly below.

Now, we define a pair  $(\tilde{\beta}, \tilde{\mu})$  of the difference game  $\Gamma_{SE, \{\beta^\Delta\}}$ . Choose  $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_I)$  so that each  $\tilde{\beta}_i(\tilde{h}_i)$  satisfies  $\tilde{\beta}_i(\tilde{h}_i)(\tilde{m}_i) = \beta_i(\eta_i(\tilde{h}_i))(\xi[\tilde{h}_i](\tilde{m}_i))$ , for all  $\tilde{m}_i \in \tilde{M}_i[\tilde{h}_i]$ . (Recall that each  $\beta_i(\eta(\tilde{h}_i))(\xi[\tilde{h}_i](\tilde{M}_i[\tilde{h}_i])) = 1$ , so this is well defined.) Likewise, choose  $\tilde{\mu}$  so that each  $\tilde{\mu}(\tilde{h})(n) = \mu(\eta(\tilde{h}))(\tilde{h})(n)$ , for each node in  $\tilde{h}$ . (Recall that each  $\mu(\eta(\tilde{h}))(\tilde{h}) = 1$ , so this is well-defined.) We will show that  $(\tilde{\beta}, \tilde{\mu})$  is a sequential equilibrium of the difference game, so that  $\{\tilde{\beta}\} \in \mathcal{S}(\Gamma_{SE, \{\beta^\Delta\}})$ . Since, by construction, any outcome allowed by  $\beta$  is allowed by  $\tilde{\beta}$ , this will establish the result.

It is immediate from the construction that each  $\tilde{\beta}_i(\tilde{h}_i)$  is a best reply under  $\tilde{\mu}$ . So, it suffices to show that  $(\tilde{\beta}, \tilde{\mu})$  is consistent.

Since  $(\beta, \mu)$  is consistent, there is some  $(\beta^k, \mu^k) \rightarrow (\beta, \mu)$  where each  $\beta^k$  is completely mixed and each  $\mu^k$  is derived from  $\beta^k$  by conditioning. As such,  $\beta_i^k(\eta(\tilde{h}_i))(\xi[\tilde{h}_i](\tilde{M}_i[\tilde{h}_i])) > 0$  and  $\mu^k(\eta(\tilde{h}_i))(\tilde{h}_i) > 0$  for all  $\tilde{h}_i$ . Define  $(\tilde{\beta}^k, \tilde{\mu}^k)$  as follows: For each  $\tilde{h}_i$  and each  $\tilde{m}_i \in \tilde{M}_i[\tilde{h}_i]$ , set

$$\tilde{\beta}_i^k(\tilde{h}_i)(\tilde{m}_i) = \beta_i^k(\eta(\tilde{h}_i))(\xi[\tilde{h}_i](\tilde{m}_i) | \xi[\tilde{h}_i](\tilde{M}_i[\tilde{h}_i])).$$

Likewise, for each  $\tilde{h}_i$  and  $n \in \tilde{h}_i$ , set  $\tilde{\mu}^k(\tilde{h}_i)(n) = \mu^k(\eta(\tilde{h}_i))(n | \tilde{h}_i)$ . Note, by construction  $\tilde{\beta}^k$  is completely

mixed and  $\tilde{\mu}^k$  is derived from  $\tilde{\beta}^k$  by conditioning. Moreover, using the fact that each  $\beta_i^k(\eta(\tilde{h}))(\xi[\tilde{h}_i](\tilde{M}_i[\tilde{h}_i])) \rightarrow 1$ ,  $\mu^k(\eta(\tilde{h}_i))(\tilde{h}_i) \rightarrow 1$ , it follows that  $(\tilde{\beta}_i^k, \tilde{\mu}_i^k) \rightarrow (\tilde{\beta}_i, \tilde{\mu}_i)$  as required. ■

**Extensive-Form Rationalizability** Extensive-form rationalizability (EFR) can be seen as embodying forward induction reasoning. (See [Battigalli and Siniscalchi, 2002](#) for a formal statement.) Here we argue that EFR also satisfies BI.

It will be convenient to amend the definition of  $Q_{-i}$ -rationality to the case of pure strategies. Write  $S_i(h)$  (resp.  $S_{-i}(h)$ ) for the set of  $s_i \in S_i$  (resp.  $s_{-i} \in S_{-i}$ ) that allow  $h$ . Fix some  $Q_{-i} \subseteq S_{-i}$ . Say a strategy  $s_i^*$  is  **$Q_{-i}$ -pure strategy rational** if, for each information set  $h \in H_i$  allowed by  $s_i^*$ , there is some  $\rho_{i,h} \in \mathcal{M}(S_{-i}(h))$  satisfying the following:

$\pi_i(s_i^*, \rho_{i,h}) \geq \pi_i(s_i, \rho_{i,h})$  for all  $s_i \in S_i(h)$ , and

if  $Q_{-i} \cap S_{-i}(h) \neq \emptyset$ , then  $\rho_{i,h}(Q_{-i}) = 1$ .

Now we can state the definition of EFR ([Pearce, 1984](#)): Fix a game  $\Gamma$  and the associated strategy sets  $S_1, \dots, S_I$ . Write  $S_i^0[\Gamma] = S_i$ . Assume each  $S_i^m[\Gamma]$  is defined and let  $S_i^{m+1}[\Gamma]$  be the set of  $s_i \in S_i^m[\Gamma]$  that are  $S_{-i}^m[\Gamma]$ -rational.<sup>8</sup> Then  $S^m[\Gamma]$  is the set of **m-extensive form rationalizable (m-EFR)** strategies and  $\bigcap_{m \geq 0} S^m[\Gamma]$  is the set of **extensive form rationalizable (EFR)** strategies. As above, take  $\mathcal{S}_{\text{EFR}}$  so that

$$\mathcal{S}_{\text{EFR}}(\Gamma) = \left\{ \bigcap_m S^m[\Gamma] \right\}.$$

**Proposition 3.4** *The solution concept  $\mathcal{S}_{\text{EFR}}$  satisfies D, E, and CR.*

**Proof.** Properties E and CR are standard. We focus on D.

Fix a game  $\Gamma$  and a subgame  $\Delta$ , thereof. Let  $\tilde{\Gamma}$  be the  $(\mathcal{S}_{\text{EFR}}, \bigcap_m S^m[\Delta])$ -difference game. We will construct sets  $X_i^0, X_i^1, X_i^2, \dots$  satisfying the following criterion:

- (i) If  $s_i \in X_i^m \setminus X_i^{m+1}$ , then  $s_i$  is not  $X_{-i}^m$ -rational.
- (ii) There exists  $M$  so that, each  $s_i \in X_i^M$  is  $X_{-i}^M$ -rational.
- (iii) There exists  $N \leq M$  so that  $\Pi(\zeta(X^N)) = \Pi(\zeta(\bigcap_m S^m[\tilde{\Gamma}]))$ .

Requirements (i)-(ii) say that the sets  $X^0, X^1, X^2, \dots$  form some order of elimination of iteratively irrational strategies (in  $\Gamma$ ). By [Chen and Micali \(2012, Theorem 1\)](#), this order of elimination is outcome equivalent to EFR (in  $\Gamma$ ), i.e.,  $\Pi(\zeta(X^M)) = \Pi(\zeta(\bigcap_m S^m[\Gamma]))$ . Now use the fact that  $X^M \subseteq X^N$  and so  $\Pi(\zeta(\bigcap_m S^m[\Gamma])) \subseteq \Pi(\zeta(X^N))$ . By (iii), the set of outcomes allowed by EFR on  $\Gamma$  is contained in the set of outcomes allowed by EFR on the difference tree, i.e.,  $\Pi(\zeta(\bigcap_m S^m[\Gamma])) \subseteq \Pi(\zeta(\bigcap_m S^m[\tilde{\Gamma}]))$ . Thus, D holds.

To show (i)-(iii), begin by consider the sets of  $m$ -EFR strategies in the subgame  $\Delta$ , i.e.,  $S^0[\Delta], S^1[\Delta], \dots$ , and the sets of  $m$ -EFR strategies in the difference game  $\tilde{\Gamma}$ , i.e.,  $S^1[\tilde{\Gamma}], S^2[\tilde{\Gamma}], \dots$ . There exists  $K$  and  $J$  so that  $S^K[\Delta] = \bigcap_m S^m[\Delta]$  and  $S^J[\tilde{\Gamma}] = \bigcap_m S^m[\tilde{\Gamma}]$ .

Inductively define the sets  $X^0, X^1, \dots$ , as follows:

- $X_i^0 = S_i^0[\Gamma]$ .
- For each  $m = 0, \dots, K-1$ ,  $X_i^m \setminus X_i^{m+1}$  is the set of all  $[s_i^\Delta] \cap X_i^m$  for  $s_i^\Delta \in S_i^m[\Delta] \setminus S_i^{m+1}[\Delta]$ .

<sup>8</sup>This is not [Pearce's \(1984\)](#) original definition, since it does not explicitly require that we choose assessments to satisfy the rules of conditional probability. However, a standard argument shows that it is equivalent to the original definition. See, e.g., [Battigalli et al. \(2012\)](#).

- For each  $m = K, \dots, K + J - 1$ ,  $X_i^m \setminus X_i^{m+1}$  is the set of all  $[\tilde{s}_i] \cap X_i^m$  for  $\tilde{s}_i \in S_i^{m-K}[\tilde{\Gamma}] \setminus S_i^{m-K+1}[\tilde{\Gamma}]$ .
- For each  $m \geq K + J$ ,  $X_i^m \setminus X_i^{m+1}$  is the set of all  $s_i \in X_i^m$  that are not  $X_i^m$ -rational.

Take  $N = K + J$  and note that, there exists  $M \geq N = K + J$  so that  $\bigcap_{n \geq 0} X^n = X^M$ . By construction, requirements (ii)-(iii) hold. The remainder of the argument focuses on requirement (i).

For  $m = 0, \dots, K - 1$  or  $m = K + J, \dots$ , it is immediate that requirement (i) is satisfied. For  $m = K, \dots, K + J - 1$ : Not that, for each  $i$ , there exists some  $f_i : X_i^K \rightarrow S_i^0[\tilde{\Gamma}]$  so that

$f_i(s_i)$  is the restriction of  $s_i$  to the difference game  $\tilde{\Gamma}$ , and

$f_i$  is onto. It follows that  $s_i \in X_i^K \setminus X_i^{K+1}$  if and only if  $f_i(s_i) \in S_i^0[\tilde{\Gamma}] \setminus S_i^1[\tilde{\Gamma}]$ . Then, by induction, for each  $j = 2, \dots, J$ ,  $s_i \in X_i^{K+j-1} \setminus X_i^{K+j}$  if and only if  $f_i(s_i) \in S_i^{j-1}[\tilde{\Gamma}] \setminus S_i^j[\tilde{\Gamma}]$ . ■

As a corollary of Propositions 3.2 and 3.4, we have the following:

**Corollary 3.5** *Fix a PI game satisfying SPC. Then any extensive-form rationalizable strategy profile is outcome equivalent to the backward induction algorithm.*

Theorem 4 in Battigalli (1997) shows that, in a PI game satisfying “no relevant ties,” any extensive-form rationalizable strategy profile is outcome equivalent to the backward induction algorithm. (See Chen and Micali (2012) and Heifetz and Perea (2013) for alternate proofs of the same result.) A game that satisfies “no relevant ties” satisfies SPC, but the converse does not hold. In particular, many auction and voting games satisfy SPC but fail “no relevant ties.”

## 4 Non-Monotonicity of Backward Induction

Now that we have put forward a formal definition of BI, we can establish a basic non-monotonicity. Importantly, we establish that this non-monotonicity holds, even on the class of games that satisfy SPC, written  $\mathcal{G}_{\text{SPC}}$ .

**Theorem 4.1** *There exists a solution concept  $\mathcal{S}$  and a refinement  $\mathcal{R}$  of  $\mathcal{S}$ , so that [label=()]*

$\mathcal{S}$  satisfies D, E, and CR,

$\mathcal{R}$  satisfies E and CR, and

$\mathcal{R}$  fails D on  $\mathcal{G}_{\text{SPC}}$ .

In the proof of the theorem, we will take  $\mathcal{S}$  to be sequential equilibrium. We will then show that we can take  $\mathcal{R}$  to be proper equilibrium or quasi-perfect equilibrium. That is, neither proper equilibrium nor quasi-perfect equilibrium satisfies BI. This implies that the question of the consistency of A and BI is not addressed by the results in Van Damme (1984, Proposition 1, p.9) and Kohlberg and Mertens (1986, Proposition 0, p.1009).

Begin with proper equilibrium and quasi-perfect equilibrium. Recall the definitions (Myerson, 1978; Van Damme, 1984). A profile of completely mixed strategies  $\sigma^\varepsilon = (\sigma_1^\varepsilon, \dots, \sigma_I^\varepsilon)$  is an  $\varepsilon$ -proper equilibrium of  $\Gamma$  if, for each  $i$ ,  $\pi_i(s_i, \sigma_{-i}^\varepsilon) < \pi_i(r_i, \sigma_{-i}^\varepsilon)$  implies  $\sigma_i^\varepsilon(s_i) \leq \varepsilon \sigma_i^\varepsilon(r_i)$ . A profile  $\sigma$  is a proper equilibrium of



$\Gamma$  if there is a sequence of  $\varepsilon$ -proper equilibria  $\sigma^\varepsilon$  of  $\Gamma$  with  $\lim_{\varepsilon \rightarrow 0} \sigma^\varepsilon = \sigma$ . We define the proper equilibrium solution concept  $\mathcal{S}_{PE}$  by

$$\mathcal{S}_{PE}(\Gamma) = \{\{\sigma\} : \sigma \text{ is a proper equilibrium of } \Gamma\}.$$

A profile of behavioral strategies  $\beta = (\beta_1, \dots, \beta_I)$  is a **quasi-perfect equilibrium** if there exists a sequence of completely mixed behavioral strategy profiles, viz.  $\beta^k = (\beta_1^k, \dots, \beta_I^k)$ , so that  $\beta^k \rightarrow \beta$  and, for each  $i$  and each  $h \in H_i$ ,  $\beta_i$  maximizes  $i$ 's conditional expected payoffs under each  $(\beta_j^k)_{j \neq i}$  (among strategies in  $\Sigma_i(h_i)$ ). We define the quasi-perfect equilibrium solution concept  $\mathcal{S}_{QPE}$  by

$$\mathcal{S}_{QPE}(\Gamma) = \{\{\beta\} : \beta \text{ is a quasi-perfect equilibrium of } \Gamma\}.$$

**Proposition 4.2** *The solution concepts  $\mathcal{S}_{PE}$  and  $\mathcal{S}_{QPE}$  fail D.*

**Proof.** Consider the game  $\Gamma$  given in Figure 4.1. We first show that there is a proper equilibrium where Ann plays  $L$  (at the initial node) with probability 1. To see this, note that there is an  $\varepsilon$ -proper equilibrium where Ann uses (unnormalized) weights  $(1:L, \frac{2}{3}\varepsilon:R-l, \frac{1}{3}\varepsilon:R-r)$  and Bob uses (unnormalized) weights  $(\varepsilon:L-Out, \varepsilon^3:R-Out, \frac{3}{5}:L-l, \frac{3}{5}\varepsilon^2:R-l, \frac{2}{5}:L-r, \frac{2}{5}\varepsilon^2:R-r)$ . So, the outcome  $(1, 1)$  is allowed under properness. It follows that the outcome  $(1, 1)$  is also allowed under quasi-perfection. (See [Van Damme, 1984](#), Theorem 1, p.9.)

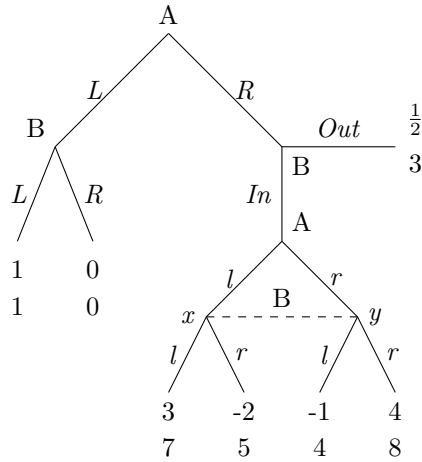


Figure 4.1

Now take  $\Delta$  to be the subgame beginning at the node where Bob can choose  $Out$ . Writing Ann's (resp. Bob's) strategies for  $\Delta$  in the order  $(l, r)$  (resp.  $(Out, l, r)$ ), there are three subgame-perfect equilibria of the subgame:  $((1, 0), (0, 1, 0))$ ,  $((0, 1), (0, 0, 1))$ , and  $((\frac{2}{3}, \frac{1}{3}), (0, \frac{3}{5}, \frac{2}{5}))$ . (Note that  $O-l$  and  $O-r$  are strongly dominated in the subgame, and so cannot be part of a subgame-perfect equilibrium.) Each of these subgame-perfect equilibria are both proper and quasi-perfect.

Figures 4.2a-4.2c give three possible difference games (associated with  $\Delta$ ) for both properness and quasi-perfection. In each of these games, Ann's strategies  $L-l$  and  $L-r$  are weakly dominated. (In Figure 4.2c, they are weakly dominated by a  $\frac{1}{2}:\frac{1}{2}$  mixture of  $R-l:R-r$ .) Therefore, the outcome  $(1, 1)$  cannot arise in a proper or quasi-perfect equilibrium of  $\Delta$ . This contradicts D. ■

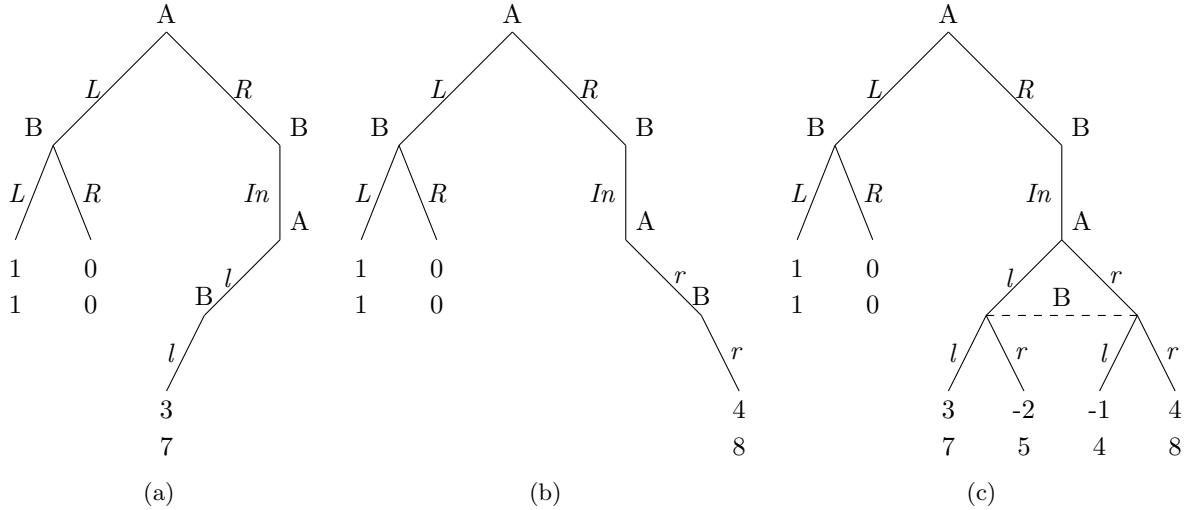


Figure 4.2

It is instructive to compare the behavior of sequential equilibrium with that of proper (or quasi-perfect) equilibrium in the game of Figure 4.1. Much as with proper (or quasi-perfect) equilibrium, there is a sequential equilibrium where: (i) Ann puts weight 1 on  $L$ ; and (ii) Bob puts weight 1 on  $L$ , weight 1 on  $In$ , and weights  $\frac{3}{5}:\frac{2}{5}$  on  $l$  vs.  $r$ . This is supported by a system of beliefs that puts weights  $\frac{2}{3}:\frac{1}{3}$  on node  $x$  vs. node  $y$ . Likewise, the three proper (or quasi-perfect) equilibria of the subgame  $\Delta$ , there are three sequential equilibria  $\Delta$ . In particular, Figure 4.2c is again a difference game under sequential equilibrium. The distinction is that there is a sequential equilibrium of this third difference game in which Ann plays  $L$ . (The details are the same as for the sequential equilibrium of the original game.) So, this time, D is satisfied (as required by Theorem 4.1).

Under properness or quasi-perfection, the situation is different. The strategies  $L-l$  and  $L-r$  for Ann are admissible in the original game of Figure 4.1. In fact,  $L$  is played in a proper (and, therefore, quasi-perfect) equilibrium. It is supported by a mixed strategy for Bob that puts  $\varepsilon$ -times less weight on  $R-I$  vs.  $L-O$ . But, in the difference game,  $O$  is eliminated for Bob, and so there cannot be a mixed strategy for Bob that puts  $\varepsilon$ -times less weight on  $R-I$  vs.  $L-O$ . As a result,  $L$  is weakly dominated in each of the difference games of Figures 4.2a-4.2c and so cannot be part of a proper (or quasi-perfect) equilibrium of these games.

Ex post, the non-monotonicity in BI which we identify in this paper is, perhaps, not that surprising. At least, it may not be that surprising once one has a direct definition of BI, as we provide. Here is the essence of the argument:

- Start with a solution concept that satisfies BI. (In our case, this is sequential equilibrium.)
- Next consider a stronger solution concept. (In our case, this is proper equilibrium.)
- The stronger solution concept may prune more moves in forming a particular difference game. (In our case, this is the move *Out* for Bob.)
- From elementary game theory, we know that when we prune a move for one player in a game, we can change previously good strategies for other players into bad strategies. (In our case, these are the strategies  $L-l$  and  $L-r$  for Ann.)

- Suppose such a previously good strategy is played under the stronger solution concept on the overall game. Then, this solution concept will fail D—and, therefore, BI. (In our case, Ann’s playing  $L$  is indeed part of a proper equilibrium of the overall game.)

The argument is actually very elementary. Propositions 4.2 and 3.3 serve simply to convert the in-principle argument into a specific instance of interest.

We note in passing that there is another (potential) source of non-monotonicity. Figure 4.2c was a difference game for both the solution concept  $\mathcal{S}$  and the refinement  $\mathcal{R}$ . However, a refinement might rule out a difference game altogether. This, too, could lead to a failure of D.

## 5 Are Admissibility and Difference Consistent?

Now, return to the main question: Are admissibility and BI consistent? Recall, a strategy  $\sigma_i \in \mathcal{M}(S_i)$  is **admissible** if there is no strategy  $\rho_i \in \mathcal{M}(S_i)$  which weakly dominates it. We define:

(A) *A solution concept satisfies **Admissibility** if it contains only admissible strategies.*

The question of whether A and BI are consistent amounts to: Does there exist a solution concept that satisfies A, D, E, and CR? The answer depends on the family of extensive-form games to which we apply the solution concept. Let us begin with two extreme cases.

Q1 Does there exist a solution concept that satisfies A, D, E, and CR on the domain of all finite (perfect-recall) games?

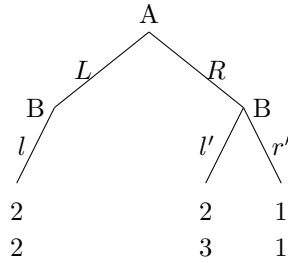
Q2 Does there exist a solution concept that satisfies A, D, E, and CR on the domain of all “generic” finite (perfect-recall) games?

Answers to these questions can already be found in the literature. Let us review.

Begin with Q1. Here the answer is no. This negative conclusion can already be seen from a modification of Figure 5 in Kohlberg and Mertens (1986, p.1013). (The modification allows us to talk about outcomes rather than strategies.) Kohlberg and Mertens did not, themselves, conclude that there is an inconsistency between A and BI, precisely because they did not have a direct definition of BI. (In fact, their paper suggests the opposite—that A and BI are consistent.) We now review the modified example.

**Example 5.1** Refer to Figure 3.4 and consider the subgame following Ann’s play of  $L$ . By E and A, the solution on the subgame requires Bob to play  $L$ . Now, refer to the difference game in Figure 5.1 and note that, by E and A, Ann must play  $L$  in this game. So, by E and D, Ann must play  $L$  in the original game. This yields the (2, 2) outcome. But a similar argument applies to the subgame following Ann’s play of  $R$ . This yields the (2, 3) outcome—a contradiction.

Next turn to Q2. Here the answer is yes. This positive conclusion can be obtained from a result by Pimienta and Shen (2010). Fix an extensive form, written  $G$ , and let  $\text{PF}(G)$  be the set of payoff functions for the extensive-form  $G$ . Say  $X \subseteq \text{PF}(G)$  is **generic relative to  $G$**  if its complement is a lower-dimensional semi-algebraic set. (See Blume and Zame, 1994 and the references there.) Theorem 1 in Pimienta and Shen (2010) shows that, for a given extensive form  $G$ , the set of associated payoffs so that sequential equilibrium and quasi-perfect equilibrium coincide is generic relative to  $G$ . Recall, sequential



□

Figure 5.1

equilibrium satisfies D and E (Proposition 3.3) and quasi-perfect equilibrium satisfies A. So, this implies that, for a given extensive form  $G$ , the set of payoffs so that the associated games satisfy A, D, E, and CR is generic relative to  $G$ .

But, arguably, both Q1 and Q2 are too extreme. Q1 asks the question on a domain that includes certain ‘pathological’ extensive-form games, such as that in Figure 3.4. Indeed, we can already see the non-monotonicity of D (across all games) by looking at Figures 3.4 and 5.1.<sup>9</sup> (There is a sequential and proper equilibrium of Figure 3.4 where Ann plays  $R$  and Bob plays  $l-r'$  with probability 1. This induces a sequential equilibrium of the game in Figure 5.1, but not a proper equilibrium of that game.) The reason for the non-monotonicity is that Ann is indifferent between two terminal nodes over which she is decisive, despite the fact that Bob is not. Put differently, these two terminal nodes correspond to distinct outcomes, despite the fact that one player is indifferent between them.

On the other hand, Q2 asks the question on a domain that excludes many ‘non-pathological’ extensive-form games. It is asked on a strict subdomain of the domain of games where any two terminal nodes must correspond to different outcomes. (This domain excludes Figure 4.2c, which has no ties.) The domain excludes many extensive-form games of applied interest. In particular, it excludes extensive-form games where distinct terminal nodes are associated with the same outcome (or consequence), e.g., voting games, auctions, Bertrand competition, etc. (See the discussions in Mertens, 1989 and Marx and Swinkels, 1997.) In the particular case of voting and auction games, both backward induction and admissibility are often important aspects of the analysis.<sup>10</sup> This leads to the question:

Q3 Does there exist a solution concept that satisfies A, D, E, and CR on the domain of all finite (perfect-recall) games satisfying SPC?

We view this as an important open question.

## 6 Conclusion

In this paper, we have taken the classical view that A and BI are desirable properties for a solution concept to satisfy. We provided a direct definition of BI and questioned the consistency of A and BI. We suggest that there is an important open question: the consistency of A and BI on the domain of all finite (perfect-recall) games satisfying SPC.

<sup>9</sup>We thank Priscilla Man and Andy McLennan for this point.

<sup>10</sup>Examples include Besley and Coate, 1997, Caillaud and Mezzetti, 2004, Gerardi and Yariv, 2007, and Hörner and Jamison, 2008, among many others.

We departed from the classical view in our insistency on giving a direct definition of BI, i.e., a test on a solution concept that does not refer to other solution concepts. In our view, it is important to give such a direct definition of BI—the indirect definitions in the literature cannot be viewed as primitive. Of course, this is not to say that an indirect definition is necessarily problematic. Perhaps, there is a theorem that says that a particular indirect approach is equivalent to a direct approach. But, it would seem that, for this, we would still need a direct definition of BI. If we take D, E and CR as the direct definition of BI, then the non-monotonicity result shows that the indirect approaches advocated to date do not work this way.

Experts have suggested a need for such a direct definition. For instance, in a recent paper, [Govindan and Wilson \(2012, p.1671\)](#) say of their (indirect) definition of BI:

“[it] will not be appropriate for a genuine axiomatic development because it invokes one refinement (quasi-perfection) to drive another (stability).”

But, at the same time, it is important to be pragmatic—certain tools are simply more expedient for making progress in a research area. Indeed, Govindan wrote to us on the direct approach:

“I’m afraid if we go down that path, we are constraining ourselves severely and an axiomatic approach may never be possible.”<sup>11</sup>

Arguably, the desire for pragmatics is a key factor that led the literature to shy away from a direct definition of BI.

We agree with both viewpoints. Certainly, a pragmatic approach is often helpful and desirable in pushing a research area forward. Our position is that it is also important to push the development of a purist approach. First, as we said above, we do not know whether or not a consistency problem arises on the family of games that seems to us to be of most interest. Second, suppose inconsistency is, in fact, found. We see no a priori reason to think this would be a dead end. It is easy to call to mind inconsistency results in various fields (e.g., the inconsistency of naive set theory) that have spurred important subsequent developments.

## A Appendix

In Section 3, we argued that, taken together, D, E, and CR capture BI. We showed that some other proposals do not capture BI. We now return to investigate alternative formulations of BI, based on either amending D or amending CR. The first two are suggested by property (BI3) in [Kohlberg and Mertens \(1986, p.1012\)](#). The latter is suggested by property (BI0) in [Kohlberg and Mertens \(1986, p.1012\)](#).

**a. Strategies vs. Outcomes** We formulated our property D in terms of outcomes not strategies. In accordance with this, the statement of Proposition 3.2 also involves outcomes: It says that D, E, and CR give the BI outcome—not the BI strategies. We saw that this cannot be improved to deliver strategies.

Perhaps we should restate D, so that it is a requirement on strategies and not outcomes. Specifically:

(SD) A solution concept  $\mathcal{S}$  satisfies **Strategy-Wise Difference** (on  $\mathcal{G}$ ) if for each game  $\Gamma$  (in  $\mathcal{G}$ ) and each subgame  $\Delta$  of  $\Gamma$  the following holds. Let  $Q \in \mathcal{S}(\Gamma)$ . Then there exists a  $Q^\Delta \in \mathcal{S}(\Delta)$  a  $P^\mathcal{S} \in \mathcal{S}(\Gamma_{\mathcal{S}, Q^\Delta})$  such that for each  $\sigma \in Q$ , the restriction of  $\sigma$  to  $\Gamma_{\mathcal{S}, Q^\Delta}$  is contained in  $P^\mathcal{S}$ .

---

<sup>11</sup>Hari Govindan: Personal communication.

The proof of Proposition 3.3 shows that sequential equilibrium satisfies SD. Moreover, since quasi-perfect and proper equilibrium fail D, they also fail SD. Therefore, our main results still hold if we replace D with SD.

What if we take the definition of BI to be SD, E, and CR? One might think that, in the proof of Proposition 3.2, we can replace D line-by-line with the stronger requirement of SD and reach a stronger conclusion, viz., that we get BI strategy-wise and not just outcome-wise. But this is false.

Return to Figure 3.2, where the BI strategies were  $(In-Down, Out-Down)$ . The proof of Proposition 3.2 3.3(i) requires the following analysis: Consider the subgame in Figure A.1a. Per the new induction hypothesis, suppose that the solution on this subgame gives the BI strategies. Now consider the associated difference game in Figure A.1b. By E and CR, Ann must choose  $In$ . From this, E and SD say that, in the original game, Ann must choose some strategy and this strategy must be consistent with  $In$ . But this strategy need not be  $In-Down$ ; it could be  $In-Across$ . Certainly, then, if replace D with SD, our proof will not yield the stronger conclusion. We conjecture that a solution concept can satisfy SD, E, and CR, even though it fails to give the BI strategies. (Of course, it must give the BI outcome.)

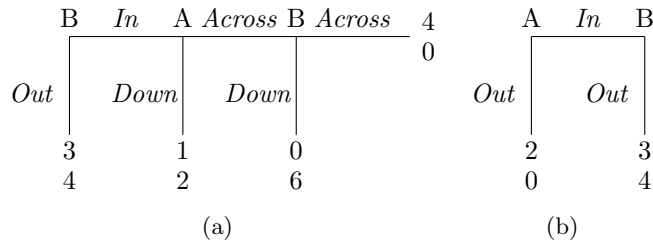


Figure A.1

**b. Expected Payoffs vs. Pruning** Fix a solution concept  $\mathcal{S}$  where each nonempty component is a singleton. In this case, we could formulate D in terms of expected payoffs rather than outcomes: Given a component  $Q \in \mathcal{S}(\Gamma)$ , we could ask that there is a nonempty (singleton) component  $Q^\Delta \in \mathcal{S}(\Delta)$  such that, when we replace  $\Delta$  with a terminal node whose payoffs are the expected payoffs under  $Q^\Delta$ , there is a component  $\tilde{Q}$  of the solution on the new game that induces the outcomes in  $Q$ . We can mimic the proofs of Propositions 4.2 and 3.3 to show that sequential equilibrium will satisfy this expected-payoff version of D, but proper (and quasi-perfect) equilibrium will not.

This said, there is no clear way to extend this version of D to the common cases of solution concepts with multi-valued components. So, this version would, say, rule out the solution concept of extensive-form rationalizability as satisfying D *because* the solution concept can have a multi-valued component.

**c. Rationality** Our definition of BI requires CR. In Section 3, we pointed to the fact that this is stronger than R.

Kohlberg and Mertens's (1986) Property (BI0) strengthens R on a second dimension by requiring that a strategy  $\sigma_i$  is optimal at every information set  $h \in H_i$ , i.e., even at information sets that are not allowed by  $\sigma_i$ . They impose this stronger requirement, as they focus on equilibrium-based solution concepts, where player  $i$ 's strategy serves as other players' assessment about how  $i$  plays the game. We don't insist on this requirement, because we don't insist on equilibrium-based solution concepts.

We could amend the definitions of R and CR to incorporate this stronger requirement. To see this, note the following fact: If a pure strategy  $s_i$  is  $Q_{-i}$ -rational, then every pure strategy  $r_i$  that induces the same plan of action as  $s_i$  is also  $Q_{-i}$ -rational. Thus, if  $s_i$  is  $Q_{-i}$ -rational, there will be some strategy  $r_i$  that induces the same plan of action as  $s_i$ , is  $Q_{-i}$ -rational, and satisfies Kohlberg and Mertens's (1986) (BI0).

## References

- Battigalli, P. 1996. "Strategic Independence and Perfect Bayesian Equilibria." *Journal of Economic Theory* 70(1):201–234.
- Battigalli, P. 1997. "On Rationalizability in Extensive Games." *Journal of Economic Theory* 74(1):40–61.
- Battigalli, P., A. Brandenburger, A. Friedenberg and M. Siniscalchi. 2012. *Strategic Uncertainty: An Epistemic Approach to Game Theory*. (Working Title).
- Battigalli, P. and M. Siniscalchi. 2002. "Strong Belief and Forward Induction Reasoning." *Journal of Economic Theory* 106(2):356–391.
- Battigalli, P. and P. Veronesi. 1996. "A Note on Stochastic Independence without Savage-Null Events." *Journal of Economic Theory* 70(1):235–248.
- Besley, T. and S. Coate. 1997. "An Economic Model of Representative Democracy." *The Quarterly Journal of Economics* 112(1):85–114.
- Blume, L. and W. Zame. 1994. "The Algebraic Geometry of Perfect and Sequential Equilibrium." *Econometrica* pp. 783–794.
- Caillaud, B. and C. Mezzetti. 2004. "Equilibrium Reserve Prices in Sequential Ascending Auctions." *Journal of Economic Theory* 117(1):78–95.
- Chen, J. and S. Micali. 2012. "The Order Independence of Iterated Dominance in Extensive Games." *Theoretical Economics* .
- Cressman, R. and K. Schlag. 1998. "The Dynamic (In) Stability of Backwards Induction." *Journal of Economic Theory* 83(2):260–285.
- Gerardi, D. and L. Yariv. 2007. "Deliberative voting." *Journal of Economic Theory* 134(1):317–338.
- Govindan, S. and R. Wilson. 2008. "Refinements of Nash equilibrium." *The New Palgrave Dictionary of Economics* 2.
- Govindan, S. and R. Wilson. 2012. "Axiomatic Equilibrium Selection for Generic Two-Player Games." *Econometrica* 80(4):1639–1699.
- Hart, S. 2002. "Evolutionary Dynamics and Backward Induction." *Games and Economic Behavior* 41(2):227–264.



- Heifetz, A. and A. Perea. 2013. "On the Outcome Equivalence of Backward Induction and Extensive Form Rationalizability."
- Hillas, J. and E. Kohlberg. 2002. "Foundations of Strategic Equilibrium." *Handbook of Game Theory with Economic Applications* 3:1597–1663.
- Hörner, J. and J. Jamison. 2008. "Sequential Common-value Auctions with Asymmetrically Informed Bidders." *The Review of Economic Studies* 75(2):475–498.
- Kohlberg, E. and J.F. Mertens. 1986. "On the Strategic Stability of Equilibria." *Econometrica* 54(5):1003–1037.
- Kohlberg, E. and P. Reny. 1997. "Independence on Relative Probability Spaces and Consistent Assessments in Game Trees." *Journal of Economic Theory* 75(2):280–313.
- Kreps, D.M. and R. Wilson. 1982. "Sequential Equilibria." *Econometrica* 50(4):863–894.
- Kuhn, H.W. 1950. "Extensive Games." *Proceedings of the National Academy of Sciences of the United States of America* 36(10):570.
- Kuhn, H.W. 1953. "Extensive Games and the Problem of Information." *Contributions to the Theory of Games* 2(28):193–216.
- Marx, L.M. and J.M. Swinkels. 1997. "Order Independence for Iterated Weak Dominance." *Games and Economic Behavior* 18(2):219–245.
- Mertens, J.F. 1989. "Stable Equilibria: A Reformulation Part I. Definition and Basic Properties." *Mathematics of Operations Research* pp. 575–625.
- Myerson, R. 1978. "Refinements of the Nash Equilibrium Concept." *International Journal of Game Theory* 7(2):73–80.
- Nachbar, J. 1992. "Evolution in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Behavior & Organization* 19(3):307–326.
- Pearce, D.G. 1984. "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52(4):1029–1050.
- Perea, A. 2010. "Belief in the Opponents Future Rationality."
- Pimienta, C. and J. Shen. 2010. "On the Equivalence Between (Quasi-) Perfect and Sequential Equilibria." *International Journal of Game Theory* pp. 1–8.
- Reny, P. 1992. "Backward Induction, Normal Form Perfection, and Explicable Equilibria." *Econometrica* 60(3):627–649.
- Samuelson, L. 1998. *Evolutionary Games and Equilibrium Selection*. Vol. 1 The MIT Press.
- Swinkels, J. 1994. "Independence for Conditional Probability Systems." *Center for Mathematical Studies in Economics and Management Science, Northwestern University* .
- Van Damme, E. 1984. "A Relation Between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games." *International Journal of Game Theory* 13(1):1–13.