ARTICLE IN PRESS
YGAME:2513

Games and Economic Behavior ••• (••••) •••–•••

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Does backwards induction imply subgame perfection?

Carlos Alós-Ferrer [a,*], Klaus Ritzberger [b]

[a] University of Cologne, Department of Economics, Albertus-Magnus Platz, D-50923 Cologne, Germany
[b] Vienna Graduate School of Finance and Institute for Advanced Studies, A-1080 Vienna, Austria

ARTICLE INFO

Article history:
Received 4 December 2014
Available online xxxx

JEL classification:
C72

Keywords:
Backwards induction
Subgame perfection
Large extensive form games
Perfect information
One-shot deviation principle

ABSTRACT

In finite games subgame perfect equilibria are precisely those that are obtained by a backwards induction procedure. In large extensive form games with perfect information this equivalence does not hold: Strategy combinations fulfilling the backwards induction criterion may not be subgame perfect in general. The full equivalence is restored only under additional (topological) assumptions. This equivalence is in the form of a one-shot deviation principle for large games, which requires lower semi-continuous preferences. As corollaries we obtain one-shot deviation principles for particular classes of games, when each player moves only finitely often or when preferences are representable by payoff functions that are continuous at infinity.

1. Introduction

Subgame perfect equilibrium is a natural solution concept for extensive form games of perfect information (Selten, 1965, 1975). In the finite case subgame perfect equilibria are precisely those delivered by the backwards induction algorithm due to Kuhn (1953). The equivalence is essentially a "one-shot deviation principle" (see, e.g., Perea, 2001, Chapter 3, or Osborne and Rubinstein, 1994, Lemma 98.2, for the finite case, and Blackwell, 1965, for the principle in dynamic programming). In order to check whether a strategy combination satisfies backwards induction the algorithm starts at the end of the tree and proceeds backwards, verifying at every decision point (node) that the player controlling it has no profitable deviation. Suppose, however, that a player i plays at least twice, once at a given node x and once further "down" the tree, i.e., at some other node y that comes after his or her first decision. Backwards induction will eventually fix the player's decision at y, and then proceed up until the first node, x, is reached. There, the algorithm will simply check whether player i can profitably deviate *at that node* (x), taking the behavior of all players in the future, including player i at y, as given.

In principle, such a strategy combination may satisfy backwards induction and still fail to constitute a subgame perfect equilibrium, indeed even a Nash equilibrium. For, a priori, even if there is no profitable deviation for i at x, taking i's behavior at y (and later) as given, there may still be *multiple* profitable deviations, where i changes his behavior both at x and at y, and possibly even at further nodes following y. This, of course, cannot happen in finite games, and in that sense the equivalence of backwards induction and subgame perfection amounts to the claim that checking for "one-shot deviations" is enough to establish subgame perfection.

We go beyond the class of finite games and consider general extensive form games with perfect information allowing for infinite horizon, infinite action spaces, and even infinitely many players. Moreover, we consider ordinal preferences rather than cardinal payoff functions. The purpose of this paper is to show that, in general, in infinite games backwards induction

* Corresponding author.
E-mail addresses: carlos.alos-ferrer@uni-koeln.de (C. Alós-Ferrer), ritzbe@ihs.ac.at (K. Ritzberger).

http://dx.doi.org/10.1016/j.geb.2016.02.005

may not be equivalent to subgame perfection, but that the equivalence (in the form of a one-shot deviation principle) holds under additional assumptions.

Somewhat surprisingly, those assumptions are topological in nature. The one-shot deviation principle holds as long as all the players' preferences are lower semi-continuous; in particular, full continuity is sufficient. Of course, this begs the question of which topology is used, and the even more primitive question of which space is endowed with a topology. Following the original formulation by von Neumann and Morgenstern (1944), we consider the space of ultimate outcomes (on which preferences are defined) as the appropriate primitive on which a topology is defined. This approach has the convenient feature that all elements of the tree, e.g. the nodes, become sets of outcomes rather than, say, abstract elements of a graph. With this approach, the one-shot deviation principle holds for *any* topology which provides the minimal conditions necessary for players to actually be able to solve optimization problems at every node, in particular that the topology is compact and every node is a closed set of outcomes—call such topologies *admissible*. Of course, these conditions are trivial in the finite case for the discrete topology (which makes every payoff function continuous), and one obtains the standard equivalence as a straightforward corollary.

Why are topological conditions needed? Our counterexample is an infinite-horizon game where players play infinitely often, as it is the case in e.g. infinitely repeated games or Rubinstein's (1982) alternating-offers bargaining game. The failure of equivalence arises from the fact that it is not possible to identify a "last decision" for a player. If this were possible, the one-shot deviation principle would hold without continuity assumptions. Indeed, we show that this is precisely the case for the class of games where it is always possible to identify the last decision of a player. This class includes all finite-horizon games independently of the cardinality of the action sets.

A second application of the main result concerns games with payoff functions that are continuous at infinity (Fudenberg and Levine, 1983). We show that the concept of continuity at infinity is equivalent to full continuity for a particular admissible topology. Since our result only requires lower semi-continuity (of preferences) with respect to some admissible topology, a one-shot deviation principle for games satisfying continuity at infinity follows directly.

The present results are conceptually related to the literature on dynamic programming, because a dynamic optimization problem is a one-player game. For this case the necessity of appropriate continuity assumptions is known. Subgame perfection corresponds to the concept of policy optimality in that literature, while backwards induction reduces to the statement that a policy cannot be improved through a single deviation; the latter property is sometimes referred to as "unimprovability". Blair (1984, Example 1) provides an example showing that without what amounts to a lower semi-continuity at infinity (his axiom A2′) a one-shot deviation principle may *not* hold. That is, backwards induction for the one-player case does not imply optimality. An example to that effect was also later provided by Streufert (1993, Section 5.2). Under A2′ (plus a weak axiom of monotonicity), however, the implication does hold (Blair, 1984, Theorem 4). Thus, for dynamic (single-player) optimization problems it is known that lower semi-continuity is needed for a one-shot deviation principle to hold. Our result extends this logic in two directions: First, to subgame perfection in multi-player games and, second, to arbitrary (admissible) topologies. The topologies that we consider allow greater flexibility, as will be illustrated below by an example (see Example 2). The result on continuity at infinity is also conceptually related to the work on consumer patience and myopia as initiated by Koopmans (1960) (see also Brown and Lewis, 1981).

Section 2 defines extensive form games with perfect information without finiteness assumptions and introduces the notation necessary for the analysis. Section 3 introduces the concept of backwards induction and shows by means of a counterexample that it does not imply subgame perfection in general. Section 4 introduces the topological framework and proves a one-shot deviation principle, which establishes that subgame perfection and backwards induction are equivalent if players' preferences are lower-semicontinuous. Section 5 contains the applications. Section 6 concludes.

2. Perfect information games

Many alternative definitions of extensive form games have been introduced in the literature. Since we will ultimately speak of topological properties of preferences, and preferences are defined on ultimate outcomes of the game, it is convenient to follow the original approach of von Neumann and Morgenstern (1944, Section 8). In that approach extensive form games are defined on trees, but the latter are not seen as graphs. Rather, trees are viewed as collections of (nonempty) subsets of an underlying set of outcomes (also called plays). Although Kuhn (1953) later popularized the "graphical approach" where trees are viewed as graphs on abstract nodes, in the original approach by von Neumann and Morgenstern (1944) a node is simply a collection of outcomes. The relation between both approaches is simple: a node should be seen as the set of outcomes which are still available when a player decides at that node. In that way, a node precedes another node if and only if the latter node is properly contained in the former. Intuitively, decisions discard possible outcomes and hence reduce the size of the nodes.

We have followed this approach in our previous work on large extensive form games, in particular Alós-Ferrer and Ritzberger (2005, henceforth AR1), Alós-Ferrer and Ritzberger (2008, henceforth AR2), and Alós-Ferrer and Ritzberger (2013, henceforth AR3). These papers develop the concept of a game tree, show that viewing nodes as sets of plays ordered by set inclusion is without loss of generality (AR1), characterize the class of game trees for which every pure strategy combination induces an outcome and does so uniquely (AR2; see also Alós-Ferrer et al., 2011), and characterize discrete game trees and the associated extensive forms (AR3). In this paper, however, we need a relatively small part of the formalism, because we will deal with games of perfect information only, and those are relatively simple.