Backward Induction

Adam Brandenburger*

Version 01/04/07

In this note we'll consider game trees with perfect information, and explore the backward-induction algorithm, which is designed for such trees. In the next note we'll look at trees with imperfect information.

1 The Algorithm

The conventional method of analyzing perfect-information trees is the *backward-induction algorithm*. This involves going to the end of the tree and working back towards the beginning. The first step in the algorithm is to assign to the last player to move, the choice that maximizes that player's payoff. The second step is then to turn to the second-to-last player and, taking the last player's choice as determined in the first step, to assign to the second-to-last player the choice that maximizes her payoff. And so on.[1]

As a trivial application of the backward-induction algorithm, let us apply it to the game depicted in Figure 1 below, reproduced from the note "Game Trees."

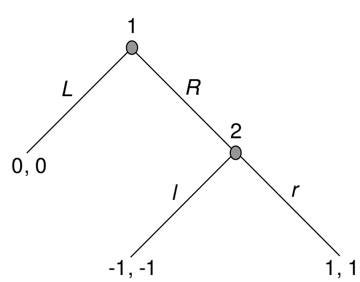

Figure 1

*With the assistance of Amanda Friedenberg and Konrad Grabiszewski. backward-01-04-07

[1] As stated, the algorithm isn't quite well-defined. First, there may be ties among the payoffs to a particular player. Second, the notion of "the last player" isn't quite precise. Do you see how to tidy up these issues?

Assign to the last player, player 2, the choice r, since that yields a higher payoff to player 2 than does the choice l. This first step of the algorithm yields the truncated game of Figure 1a. Now turn to player 1. Assign to player 1 the choice R, since that yields a higher payoff (namely 1) to player 1 than does the choice L.

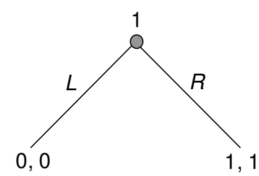

Figure 1a

In the interpretation of this game, player 1 safely trusts player 2, anticipating that it is in player 2's interest to honor that trust.

Example 1 *There are n lions in a clearing in the jungle, along with one dead lamb, and the lions are ranked from L_1 (highest) to L_n (lowest). The lions move sequentially, in order of rank, and they can choose to eat or not to eat. They are hungry, and therefore prefer to eat, but they are also cautious; they will not eat if eating will lead to their death. The lions have reason to be fearful, because they are narcoleptic, cannibalistic, and cowardly: if they eat, they fall asleep immediately, at which time they will be prey to the next lion in the sequence, who will eat only sleeping lions (or a dead lamb). Finally, the lions are finicky, so they will eat only recently dead, or newly asleep, meat–in other words, they will not eat meat that has been passed over by others.*[2]

What do you think will happen if there are six lions in the pride? What if there are seven lions in the pride?

2 Common Belief of Rationality Revisited

What conditions on the players' rationality and beliefs would lead them to play along the backward-induction path? The obvious guess is that what we need is that each player is rational and that there is common belief of rationality. The idea can be seen in the game of Figure 1. Take player 2. The rational choice for player 2 is r. Thus, if player 1 believes that player 2 is rational, player 1 will believe that player 2 will choose r. From this it follows that if player 1

[2] Quoted from "Backward Induction is Not Robust: The Parity Problem and the Uncertainty Problem," by S. Brams and M. Kilgour, *Theory and Decision*, 45, 1998, 263-289. As Brams and Kilgour note, the story appears to be part of the folklore of game theory. They thank Jeffrey Lax for the particular formulation of the story.

is rational, and believes that player 2 is rational, player 1 will choose R. Just these assumptions–and so certainly rationality and common belief thereof–yield the backward-induction path. It looks like a similar argument should work in any perfect-information tree.

If this is right, then we would have the analog for game trees–at least for trees of perfect information–of Theorem 1 in "Dominance and Iterated Dominance." Recall that there we were looking at matrices and not trees. In that context, we said (Theorem 1) that rationality and common belief of rationality implies that the players choose iteratively undominated strategies.

Is backward induction the right analog in the tree? It turns out that matters are a lot more complicated in the tree. There is a way to make rationality and common belief thereof imply the backward-induction path. But the story of getting this result is a long–and interesting!–one. One reference is my survey "The Power of Paradox: Some Recent Developments in Interactive Epistemology" (*International Journal of Game Theory*, 35, 2007, 465-492).