

On Backward and Forward Induction.

Roald Ramer, Department of Quantitative Economics,
Faculty of Economics and Econometrics, University of Amsterdam,
Roetersstraat 11, 1018 WB, Amsterdam, NL, e-mail: r.ramer@uva.nl

On von Neumann's account feasible actions and the information relevant for a conflict situation under scrutiny are modeled as an *extensive form* game. All possible strategic plans of individual players are listed in the corresponding *normal form*. Von Neumann and Morgenstern explicitly denied a possibility of any universally valid solution concept at that level. Considering the importance and ubiquity of bargaining and negotiations as conflict settling institutions they proposed to represent strategic positions of the players in terms of a *characteristic function* derived from any given normal form. The von Neumann-Morgenstern solution was intended to determine socially stable outcomes for games in characteristic form.

Without going into details we argue that all three constructions are missing a convincing behavioral or cognitive motivation. As purely mathematical models they are lacking any axiomatic foundation. On the other hand the whole project has had a remarkable internal consistency that has mostly escaped the attention of game theorists focused on the notion of a *strategic (Nash) equilibrium* as a solution concept. An interesting instance here is a general misconception concerning the axioms of the *expected utility* theory. It is invariably considered as one of the basic ingredients of the game theory, while von Neumann himself explicitly denied its usefulness for that purpose, and it can be even doubted if these axioms at all make sense if the states of nature are not exogenously given, but represent states of mind of the players.

The issue discussed in this paper is that of the identity of a *player*. From the point of view of *decision theory* a player has a set of actions at his disposal, and has some probabilistic information about exogenous factors, which could influence the outcomes of his/her actions. Whenever one physical person must make two or more choices at different points in time or given different information, it can be questioned if it should be represented as one player. I argue that the answer to this question must be negative. In general all so-called *agents* of a participant in a game should be viewed as separate players. Within von Neumann's framework they would have the same utility function and they would join the same coalition, hence they were certainly one player. Furthermore, within his context von Neumann allowed such agents to coordinate their actions. In fact he even insisted that his players should submit their full plans of action to an *umpire* who would actually play the game i.e. physically execute actions proposed by the players. In particular, in an extensive game lacking *perfect recall* von Neumann would allow players to employ mixed strategies which are not behavioral. Curiously, he even treated the partners in a game of bridge as one player, despite the fact that non-behavioral strategies are explicitly forbidden by the rules of the bridge.

In a strictly non-cooperative context, where commitments, whenever feasible, should be explicitly modeled as actions, all this makes little sense. For that matter it is not even clear which way the argument goes: should different extensive forms be seen as equivalent whenever they have the same (reduced) normal form, or rather, it could be said that nothing is lost in the normal form representation because the Thompson transformations which link different extensive forms are innocuous. Yet this equivalence lies at the heart of the fundamental notion of a *strategy* of a player and the representation of an extensive game in a normal form. It has been called the *classical point of view* by Kohlberg and Mertens. It provides the point of departure for their attempt to identify stable components of the sets of strategic equilibria of an arbitrary finite game.

On von Neumann's account subgame perfection, or more general backward induction type arguments do not bare any special significance. Yet their derivation from normal form arguments appears also highly unsatisfactory. One can argue that if perturbations of a given finite game are to play an important role in its strategic analysis, they should be incorporated explicitly in its specification. In fact, the notion of the strategic equilibrium itself is lacking any axiomatic foundation, hence its use in any specific instance should be motivate by a separate argument of one kind or another. Also results depending on the use of unrestricted inductive chains of reasoning of an arbitrary length should be viewed with suspicion. A widespread neglect of this point of view has led to various seemingly paradoxical conclusions.

Consider a game with perfect information played over a given finite number of rounds, like a repeated Prisoner's Dilemma or a Chain Store. Whenever the players are viewed as appropriately programmed automata, the unique subgame perfect equilibrium play will obtain. Therefore any argument to the contrary must depend on absence of the common knowledge of rationality on the part of the players. Curiously, even if the owner of the chain store were Selten himself and all competitors were his former students, one could not argue convincingly that accommodated entry should occur at every round of play. Suppose that Selten would fight and destroy the first two entrants at a high cost for himself. You are third in the line, would you try to challenge him? Thus even in this case there is scope for doubt which could be exploited by the long run player.

Even more revealing is the forward induction argument in the Battle of Sexes with an Outside Option. On the traditional account, after player 1 has rejected the option, player 2 is informed that the subgame, viewed as a simultaneous moves game, will be played. It has been argued that the first decision of player 1 commits her credibly to the strong strategy in the subgame. This claim can be disputed. In fact there are various possibilities here. If the backward induction argument holds, the players themselves know in advance which of the three equilibria of the subgame should be played. But then the forward induction argument has no bite. Player 1 will anticipate correctly and play accordingly. If on the other hand the players are not sure how they would behave in the same Battle of Sexes in the absence of an outside option, the forward induction could be quite convincing but the backward induction does not hold because an equilibrium play in the subgame can not be taken for granted. In fact, it appears that at least a part of an appeal of the forward induction argument stems from the fact that the game theorists as outside observers need not to know how the game will be played, even if the players do. This extensive game is strategically equivalent with a similar game in which player 1 makes her both moves at once but player 2 is only partially informed about player's 1 decision. There is no subgame here, and it could be argued that a rational player 1 for sure would not play her strongly dominated strategy at her single decision node. Therefore these two versions of the Battle of Sexes with an Outside Option are not equivalent.

These examples provide a strong argument for the view that every decision node of an extensive game should be treated as managed by a separate player. Considerations (of empirical, economic, experimental, evolutionary or epistemic nature) which could justify an expectation of an equilibrium play should also be instrumental in selecting one of them. If we can not tell which equilibrium will be played we should not take for granted any one of them need be played at all.