Games and Economic Behavior **39,** 206–214 (2002)
doi:10.1006/game.2001.0900, available online at http://www.idealibrary.com on IDEAL®

Backward Induction and the Game-Theoretic Analysis of Chess

Christian Ewerhart[1]

Department of Economics, University of Mannheim, Ludolfusstrasse 5, D-60487 Frankfurt a.M., Germany
E-mail: christian.ewerhart@planet-interkom.de

Received September 26, 2000; published online February 20, 2002

This paper scrutinizes various stylized facts related to the minmax theorem for chess. We first point out that, in contrast to the prevalent understanding, chess is actually an infinite game, so that backward induction does not apply in the strict sense. Second, we recall the original argument for the minmax theorem of chess—which is *forward* rather than backward looking. Then it is shown that, alternatively, the minmax theorem for the infinite version of chess can be reduced to the minmax theorem of the usually employed finite version. The paper concludes with a comment on Zermelo's (1913) nonrepetition theorem. *Journal of Economic Literature* Classification Number: C72. © 2002 Elsevier Science (USA)

Key Words: chess; minmax theorem; Zermelo's theorem.

1. INTRODUCTION

The classic paper of Zermelo (1913) has long been thought to contain a complete statement and proof of one of the first theorems of game theory, namely the minmax theorem for chess. This theorem asserts that either White has a winning strategy, or Black has a winning strategy, or both players can individually ensure a draw. However, a recent analysis by Schwalbe and Walker (2001) revealed that Zermelo's paper, while containing the base idea, does not make a formal statement of this theorem.

[1] I am grateful to Ehud Kalai, the editor in charge, and two anonymous referees. The paper benefited from discussions with Stephen Andrews, Dieter Balkenborg, Guy Haworth, Myron Lieberman, Martin Osborne, Jörn Rothe, Ulrich Schwalbe, and Bernhard von Stengel, and last but not least from the stimulating atmosphere of the First World Congress of the Game Theory Society in Bilbao.

0899-8256/02 $35.00

Instead, Kalmár (1928/29) seemingly was the first to state and prove this fundamental result.[2]

We believe that this discovery casts doubts on existing "folk knowledge" concerning the history of game theory and in particular concerning the game-theoretic analysis of chess. Motivated by the strong historical relevance of chess to the theory of games, but also by the conceptual importance of backward induction, this paper intends to scrutinize various stylized facts in connection with what became known as "Zermelo's Theorem."

To this end, we first point out that, in view of the official game laws, as provided by the World Chess Federation, chess is in fact an infinite game, so that backward induction does not deliver the minmax theorem in this game. Second, we recall the original argument of Kalmár (1928/29) for the minmax theorem of the infinite version of chess. Buried in a somehow technical paper, his simple method of proof, which is forward rather than backward looking, seems to be largely unknown. Then, it is shown that the minmax theorem for the infinite version of chess can alternatively be deduced from the minmax theorem of the usually employed finite version, which ends in a draw when a position appeared the third time. The paper concludes with a comment on Zermelo's (1913) nonrepetition theorem.

2. THE OFFICIAL RULES OF CHESS

Game theorists often assume that chess is finite. See, e.g., Binmore (1992), Osborne and Rubinstein (1994), or Mycielski (1992). However, strictly speaking, this is not the case. According to the official FIDE laws of chess, there is no rule or combination of rules that *generally* ends a game after a finite number of moves (cf. FIDE, 2000). Most importantly, it is not true that a game ends in a draw as soon as a position has appeared the third time. What is true is that a player may *claim* to end the game in a draw either if he is about to make a move to, or if he finds himself at a position that appeared the third time (not necessarily by repetition of moves).

Similarly, there is no official rule that generally terminates the game when the number of moves exceeds a given limit. Instead, there are two rules. The first says that a player may claim to end the game in a draw if he declares his intention to make a move that leads to, or finds himself at a position with the property that the last 50 (two-player) moves have been made without the movement of any pawn and without the capture of any

[2] Also, Kalmár was first (among Zermelo, König, and himself) to provide an explicit definition of a strategy notion ("Taktik").

piece. The second rule in fact ends the game effectively without the need for a player to make a claim, yet only under the condition that a position is reached from where a checkmate cannot occur by any possible series of legal moves, even with the most unskilled play (e.g., when solely the kings are left on the board).

Given these rules, it is easy to construct an infinite path in chess. For example, from the initial position, both players could alternatingly draw their respective right-wing horse out and back in—as many times as they wish. Other, less trivial examples can be constructed easily. Von Neumann and Morgenstern (1944, p. 60) mention the feasibility of indefinite "non-periodic" repetitions.[3] It will be noted that, however senseless any specific infinite play may seem, the possibility of infinite paths alone makes chess an infinite game.[4]

3. THE MINMAX THEOREM FOR POTENTIALLY INFINITE CHESS-LIKE GAMES

The minmax theorem for chess and similar games was very likely documented first by Kalmár (1928/29). The original paper's exposition, however, is probably not very accessible, in particular because the paper is written in German, because it does not use modern terminology, and because its scope requires the use of the somehow sophisticated method of transfinite induction. Moreover, according to Schwalbe and Walker (2001), there does not exist a usable English translation of Kalmár's paper. Schwalbe and Walker give a summary of Kalmár's paper, yet leave out the details of the argument that lead to the minmax theorem. Hence, at present, there seems to exist no simple, accessible exposition of the minmax theorem for standard infinite board games such as chess. To fill

[3] It seems that von Neumann and Morgenstern (1944) took an overall "liberal" position on the rules of chess, in the sense that they realized the possibility of paths of infinite length in the official version, but found it also appropriate to use additional rules which make the game finite.

[4] Of course, the rules of chess have changed over time. According to Myron Lieberman (personal communication), International Arbiter for FIDE and National Tournament Director for the US Chess Federation, by the end of the 19th century countries had significant differences in their rules. This period, however, ended with the first publication of the FIDE Laws of Chess in the year 1929. The draw rules of that time appear less accurately formulated, but otherwise different from the current ones only in details. Specifically, the fifty move count was apparently not reset by the move of a pawn, but only when a piece has been captured. Also, there was the explicit possibility of a claimed draw by perpetual check. Note that these modifications do not affect the potential infinity of chess. The rules remained unchanged until 1952, so that they appear to be the most relevant set of rules for von Neumann and Morgenstern's work on chess.

this gap, we will define a class of infinite games, one of which is the infinite version of chess, and recall Kalmár's argument that any game in this class has a value.

In this paper, all games are two-player strictly competitive (for definitions of the standard notions used from now on, see, e.g., Binmore, 1992). We will say that a node has height n if there are precisely n moves necessary to reach this node (throughout the paper, unless otherwise stated, we use the game-theoretic definition of a move as an action of just one player, and *not* the definition common in the chess literature, where a move consists of two consecutive actions by White and Black.) By a *potentially infinite chess-like game*, we mean a strictly competitive perfect-information game with at most three outcomes in which each node has a finite height and a finite number of immediate successors, and in which any infinite path yields a draw. Note that when infinite plays are declared to be draws (as was proposed by Zermelo, 1913), then the official version of chess is a potentially infinite chess-like game.

As will become clear below, the following theorem immediately implies the minmax assertion for chess.[5] According to Kalmár (1928/29), this result and its corollary were known already to John von Neumann.[6]

THEOREM 1 (Kalmár). *Consider any potentially infinite chess-like game G. Then, if player i cannot enforce a win, then player j can ensure at least a draw.*

We will say that a node x is a *nonwinning* position for i when there does not exist a winning strategy for player i in the subgame starting at x.

Proof. Note first that, when j is called upon to play at a position x, that is nonwinning for i, then there must exist an immediate successor node of x that is nonwinning for i. To see why, assume to the contrary, in any subgame starting at some immediate successor node of x, player i can choose a winning strategy. Then, in the subgame starting at x, a grand strategy composed of these strategies in the respective subgames will be a winning strategy for i. This, however, contradicts our assumption that x is nonwinning for i. Thus, whenever player j moves at a node x that is nonwinning for i, there is an immediate successor node that is also nonwinning for i.

[5] Even so, von Neumann and Morgenstern (1944) do not reference any of the earlier papers, except for König which they cite to prove the finiteness of the game tree when there is a stop rule.

[6] Referring to his Satz II, which is a statement of the minmax corollary, and which he derives from the subsequent Satz III, Kalmár (1928/29, Footnote 17) writes: "As Herr König kindliest communicated to me subsequently, this proposition was known also to Herr v. Neumann."

We can therefore define player j's action at any such x in the searched-for strategy by the requirement that it leads to an immediate successor node that is nonwinning for i. At all other nodes in G, choose any of the feasible actions for j. This defines a strategy s for player j. Now we claim that any path p generated by s and some strategy of i yields at least a draw for j. This is clear for any infinite path. Assume therefore that p is finite. We will show that the terminal node of p must be nonwinning for i, and therefore yield at least a draw for j. Note first that, by assumption, the initial node of G is nonwinning for i. Hence, it suffices to show that if x is nonwinning for i, the immediate successor node of x on p has the same property. This is clear for any x, at which j is called upon to play, because j's strategy was precisely defined that way. But this is also true for any node x at which it is i's turn, because, if i had available a winning strategy in a subgame starting at some immediate successor node w of x, this strategy could be complemented, in a way that i moves from x to w, to a strategy of i in the subgame starting at x. As w was arbitrary, this shows that any successor node of x is nonwinning for i. This proves the theorem. ■

Theorem 1 is essentially Kalmár's "Satz III".[7] It will be noted that the idea of the proof is forward rather than backward looking. The draw-ensuring strategy for player j requires j, whenever the current position is a nonwinning position for i, to make a move to some other nonwinning position for i (this is always feasible, as the proof shows). As the initial position is assumed to be nonwinning for i, this requirement in fact defines a strategy for j that ensures at least a draw.

Theorem 1 implies immediately that the infinite version of chess (in fact, any potentially infinite chess-like game) has a value, i.e., either White has a winning strategy, or Black has a winning strategy, or both can ensure themselves a draw, but not more. To see why, note first that, trivially, White and Black cannot both have a winning strategy. So either White has a winning strategy, or Black, or none has a winning strategy. But if no player possesses a winning strategy, then by Kalmár's theorem, both can individually ensure a draw. This proves the assertion.

[7] Our formulation differs from the original in two respects. Firstly, Kalmár (1928/29) considers more general games. Specifically, he allows nodes of infinite height (this is not more general than our setting as Kalmár also defines infinite paths as draws), and also an infinite number of immediate successors of a given node (this would require the use of the axiom of choice in our setting, which is unnecessary in Kalmár's paper because his notion of a "Taktik" does not require a specification of a *unique* action at each relevant node, as the modern notions of strategy, pure strategy and plan of action, do). The second difference is that, for a given game, Satz III is slightly stronger than Theorem 1. Specifically, Satz III says that if player i cannot enforce a win, then player j can ensure at least a draw using a "Taktik" *that depends only on the respective board position*, not on the whole history of the play.

In a companion paper in this journal (Ewerhart, 2000), I show that any *finite* chess-like game, i.e., any finite, strictly competitive perfect-information games with at most three outcomes can be reduced to a trivial game (i.e., all outcomes are equivalent) by two rounds of elimination of weakly dominated strategies. Since the minmax theorem is valid also for the infinite version of chess, and any infinite play yields a draw, this result and its proof extend straightforwardly to any potentially infinite chess-like game, and in particular to the infinite version of chess.

4. BACKWARD INDUCTION

It is natural to suggest that the infinite version of chess is in some sense strategically equivalent to some finite version, and that the value of chess is therefore deducible via backward induction. We find below that this intuition can be made precise.

We first give a more formal description of the common finite version of chess. Being a board game, infinite chess has a finite number L of positions (to be precise: *positions* are considered the same, if the same player has the move, pieces of the same kind and color occupy the same squares, and the possible moves of all the pieces of both players are the same. Positions are not the same if a pawn could have been captured *en passant* or if the right to castle immediately or in the future has been changed.) Hence, on any path p with more than $2L$ positions there is at least one position appearing the third time. Consider the first node x on the path p that corresponds to a third-time appearance of a position. Denote by X the set of all nodes in the infinite version of chess that can be obtained in that way. Let i be the player that is called upon to make a move at x. Clearly, x is not the initial node. Denote by y the node played by player j prior to some $x \in X$. Let Y be the set of all nodes y in the infinite version of chess that can be obtained in that way. By the official rules, j has the option to end the game in a draw at y, and i has the option to end the game in a draw at x. In the finite version of chess, the options at x and y are removed, and x is made into a terminal node that ends the game in a draw, for any $x \in X$, and any $y \in Y$.

THEOREM 2. *A player's value in the infinite version of chess is the same as his value in the finite version* (*that demands that the game ends in a draw as soon as a position appears the third time*).

The proof is essentially straightforward. Any strategy in the infinite version of chess that ensures a win (draw) will induce a strategy with the same property in the finite version. Conversely, any strategy ensuring a win (draw) in the finite version can be complemented and modified to become

a strategy with the same property in the infinite game, with the sole complication that in the case of a strategy that ensures a draw, the player in the infinite game will have to exploit all of his options to end the game.

Proof. Note first that the removal of the option at y is strategically irrelevant in the finite game, so that the finite game can be considered as a truncated game of the infinite version in the sense that all nodes in X are declared terminal. Then, any strategy in the infinite game induces a strategy in the finite game in a natural way. The proof proceeds in four steps.

(I) Assume that player i has a winning strategy in the infinite game. This strategy induces a winning strategy in the finite game. To see why, consider any strategy of j in the finite game. This strategy can be complemented to some strategy in the infinite game. Given that player i uses his winning strategy in the infinite game, player j's strategy does not reach any node $x \in X$, because otherwise j would have the option, either at x or at the node preceding x, to end the game in a draw. But this cannot be the case as i uses a winning strategy in the infinite game. Thus, the strategy for i in the finite game, that is induced by i's winning strategy in the infinite game, wins against any of j's strategies, and is therefore itself a winning strategy.

(II) Assume that i can ensure a draw in the infinite game by use of some strategy. We claim that this strategy induces a draw-ensuring strategy in the finite game. To see why, consider any strategy of j in the finite game. This strategy can be complemented to some strategy in the infinite game. There are two cases.

(a) Assume that the original strategy of i in the infinite game together with j's complemented strategy does not reach any node in X. Then it is clear that the induced strategy for i in the finite game yields at least a draw against j's original strategy.

(b) Assume now to the contrary that the draw-ensuring strategy of i together with the complemented strategy of j in the infinite game *does* reach some node in X. Then, by definition of the truncated game, the induced strategy for i yields a draw against j's original strategy.

Thus, if i can ensure a draw in the infinite game, he can also ensure a draw in the finite game.

(III) Assume now that i has a winning strategy in the finite game. Then this strategy can be complemented to a strategy in the infinite game in some way. We claim that the complemented strategy is a winning strategy in the infinite game. To see why, consider any strategy for j in the infinite game. This strategy induces a strategy for j in the finite game, against which player i's original strategy yields a win. But then clearly, as

the path that leads to i's win does not reach any node in X (otherwise, the path would end in a draw in the finite game), player i's complemented strategy yields a win against player j's strategy in the infinite game. This proves the assertion.

(IV) Finally, assume that i has a strategy that ensures a draw in the finite game. Complement this strategy to some strategy in the infinite game. Modify the resulting strategy in a way such that i chooses the option to end the game in a draw at all of his nodes in X or Y. Consider any strategy for j in the infinite game. Then clearly this strategy induces a strategy in the finite game, against which the original strategy of i ensures a draw. As above, there are two cases.

(a) The path in the finite game does not reach a node in X. Then clearly, i's modified strategy in the infinite game yields at least a draw for i.

(b) The path in the finite game reaches some node $x \in X$. Then, in the infinite game, player i exerts his option to end the game at either x or at the preceding node y.

Thus, if i can ensure a draw in the finite game, he can also ensure a draw in the infinite game.

Summing up, we have shown that i can ensure a win (draw) in the finite game if and only if i can ensure a win (draw) in the infinite game. This implies the theorem. ■

Using the minmax theorem in finite games (see, e.g., Binmore, 1992), Theorem 2 implies the existence of a value in the infinite version of chess. Moreover, the result shows that it is unambiguous to use the term "value of chess," and that this value can, at least theoretically, be determined via backward induction or iterated weak dominance in the finite version.

APPENDIX: A COMMENT ON ZERMELO'S NONREPETITION THEOREM

In Zermelo (1913) it is proved that if a player in chess has a winning strategy, then he can enforce to win in a number of moves that is smaller than the number of positions of chess. Later, König (1927) found and corrected an error in Zermelo's proof. In this section, we comment on the definition of a "position" in this literature.

There are only finitely many positions in chess, where a position is defined as in Section 4. To account for claims that may be made by a chess player, define the *z-position* of a node in chess as a quadruple consisting of the position corresponding to that node, the set of positions that appeared

precisely once before, the set of positions that appeared at least twice before, and the minimum of 100 and the number of moves (counting each player separately) that have been made without any movement of any pawn and without the capture of any piece. Then it is clear from the rules of chess that any two subgames starting at nodes with equal z-positions are isomorphic. Since there is only a finite number of positions, and the power set of a finite set is again finite, the total number of z-positions is finite. For Zermelo's argument in König's paper (1927), which is conveniently summarized in Schwalbe and Walker (2001), to be valid, the definition of a "position" should correctly be that of a z-position. Thus, in an adapted form, Zermelo's main theorem in his 1913 paper can be stated as follows.

THEOREM 3 (Zermelo). *If a player can enforce a win in chess, then he can do so in less than t moves, where t is the number of z-positions.*

Proof. See text above.

REFERENCES

Binmore, K. (1992). *Fun and games*. Lexington, Massachusetts: D.C. Heath.

Ewerhart, C. (2000). "Chess-like games are dominance-solvable in at most two steps," *Games Econom. Behav.*, **33**, 41–47.

FIDE (2000). *Fide handbook*. http://www.fide.com (official web-side of the World Chess Federation).

Kalmár, L. (1928/29). "Zur Theorie der abstrakten Spiele," *Acta Universitatis Szegediensis/Sectio Scientiarum Mathematicarum* **4**, 65–85.

König, D. (1927). "Über eine Schlussweise aus dem Endlichen ins Unendliche," *Acta Universitatis Szegediensis/Sectio Scientiarum Mathematicarum* **3**, 121–130.

Mycielski, J. (1992). "Games with perfect information," in *Handbook of Game Theory*, Vol. 1 (R. Aumann and S. Hart, Eds.). Amsterdam: Elsevier, pp. 41–70.

Osborne, M., and Rubinstein, A. (1994). *A Course in Game Theory*. Cambridge: MIT Press.

Schwalbe, U., and Walker P. (2001). Zermelo and the early history of game theory, *Games Econom. Behav.* **34**, 123–137.

Von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Zermelo, E. (1913). "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels," in *Proceedings of the Fifth Congress Mathematicians, Cambridge 1912*. Cambridge, UK: Cambridge University Press, pp. 501–504.