ECONOMIC RESEARCH REPORTS

BACKWARD INDUCTION IS NOT ROBUST: THE PARITY PROBLEM AND THE UNCERTAINTY PROBLEM

by **D. Marc Kilgour**
and
Steven J. Brams

RR # 96-21 **May 1996**

C.V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

Backward Induction Is Not Robust: The Parity Problem and the Uncertainty Problem

D. Marc Kilgour
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5
CANADA

Steven J. Brams
Department of Politics
New York University
New York, NY 10003

ABSTRACT

Backward Induction Is Not Robust: The Parity Problem and the Uncertainty Problem

A cornerstone of game theory is backward induction, whereby players reason backward from the end of a game in extensive form to the beginning in order to determine what choices are rational at each stage of play. Truels, or three-person duels, are used to illustrate how the outcome can depend on (1) the evenness/oddness of the number of rounds (*the parity problem*) and (2) uncertainty about the endpoint of the game (*the uncertainty problem*). Since there is no known endpoint in the latter case, an extension of the idea of backward induction is used to determine the possible outcomes.

The parity problem highlights the lack of robustness of backward induction, but it poses no conflict between two foundational principles and, hence, does not seem paradoxical. On the other hand, two conflicting views of the future underlie the uncertainty problem, depending on whether the number of rounds is bounded or unbounded. While in the bounded case the players invariably shoot from the start, in the unbounded case they may all cooperate and never shoot, despite the fact that the truel will end with near certainty—and therefore be *effectively* bounded—by the end of several rounds. Some real-life examples, in which destructive behavior sometimes occurred and sometimes did not, are used to illustrate these differences.

JEL Classification: C73. *Keywords:* Backward induction; bounded rationality; continuation probability; infinite horizon; uncertainty.

Backward Induction Is Not Robust: The Parity Problem and the Uncertainty Problem[1]

A cornerstone of game theory is *backward induction*, whereby players reason backward from the end of a game in extensive form to the beginning in order to determine what choices are rational at each stage of play.[2] Although backward induction seems, on occasion, to make heroic demands of players, it produces logically compelling, if not always plausible, results (e.g., "never cooperate" in finitely repeated Prisoners' Dilemma).

Seemingly more plausible results can be obtained not only for repeated Prisoners' Dilemma but also for the chain-store paradox (Selten, 1978)—a repeated game in which continued cooperation by the large player seems implausible—by introducing some chance that information about preferences is incomplete (Kreps and Wilson, 1982; Milgrom and Roberts, 1982), that the players will act irrationally (Kreps *et al.*, 1982), or that there is no common knowledge of rationality (Aumann, 1992; Bicchieri, 1993; Stuart, 1993). It turns out that these and other bounded-rationality assumptions can lead to an equilibrium outcome better for the large player in

---

[1]We gratefully acknowledge the valuable comments of Jordan Howard Sobel and Stephen J. Willson on an earlier version of this paper. Steven J. Brams is pleased to acknowledge the support of the C. V. Starr Center for Applied Economics at New York University, and D. Marc Kilgour the support of the Social Science and Humanities Research Council of Canada.

[2]Aumann (1995, p. 6) points out that backward induction is "the oldest idea in game theory" and that it has "maintained its centrality to this day." In perfect-information games of the kind we shall analyze in section 2, he shows that common knowledge of rationality implies backward induction. Although the uncertainty we introduce in section 3 about when a game ends creates problems for the straightforward application of backward induction, we can still use it to study what rational players, looking ahead, would do—and then trace these consequences back to the beginning of play.

the chain-store game, and for both players in Prisoners' Dilemma, than the backward-induction outcome in the unbounded-rationality case.[3]

In the games we shall analyze, play is *not* repeated, so there is no accumulation of payoffs from a constituent or stage game, played over and over again. True, several of our games may go many "rounds," but the completion of a round does not yield the players payoffs unless the game ends on that round.

The choices that backward induction prescribes in these games cast two kinds of doubt on the robustness of backward induction. The first doubt is caused by a "parity problem," whereby which one of two possible outcomes that can occur depends on whether the number of rounds is even or odd. Thus, if the number of rounds of play is, say, 64, we get a completely different outcome than if this number were either 63 or 65. Surprisingly, whatever the number, rational play never goes beyond either one or two rounds before the game ends.

The even-odd difference poses no conflict between any foundational principles of rational choice of which we are aware. Nor do we know of any alternative solution that is intuitively more reasonable. Thus, there do not seem grounds to label the even-odd fluctuations in the backward-induction solution paradoxical.

The second kind of doubt is caused by an "uncertainty problem," whereby uncertainty about when a game will end affects the outcome. This sensitivity also arises in repeated play of Prisoners' Dilemma when there is

---

[3]For more on bounded-rationality and related solutions to repeated games, see Radner (1986), Pettit and Sugden (1989), and Sobel (1994, pp. 345-365). Our analysis, by contrast, presumes that players are *fully* rational, though their choices may be constrained by uncertainty about when the game will end.

an "infinite horizon," wherein not knowing when play will cease may create incentives for the players to cooperate.

In repeated Prisoners' Dilemma, cooperation can occur if the number of *repetitions* of the stage game is uncertain, whereas in our games it can occur when the number of *rounds* is uncertain. (We shall say precisely what we mean by a "round" later.) In either case, the uncertainty means that the game has no clear endpoint, at which one can start the backward-induction process.

In the case of an uncertain number of rounds, we show that there is a natural way of extending the idea of backward induction. However, it may give completely different results from what is obtained by applying backward induction to each of the individual games that the uncertain case subsumes. In addition, we show that the *nature* of the uncertainty—in particular, the probability that the game will continue to a next round, and the boundedness or unboundedness of the number of rounds—also matters, which again tends to undermine the robustness of backward induction.

Underlying this lack of robustness are two conflicting views of the future: (1) every process must end by some definite point; and (2) the future is unpredictable, so the endpoint of a process cannot be predicted. While these views are not necessarily irreconcilable, they can give rise to backward-induction arguments that yield very different outcomes in the games we analyze. The second view, we suggest, offers a more sanguine outlook on the future than does the first, but both views are plausible and, therefore, render the uncertainty problem paradoxical.

We illustrate the parity and uncertainty problems with *truels*, or three-person duels, in which each of three players can fire or not fire at another

player. These attrition games are, admittedly, based on rather artificial rules of play that one would not expect ever to encounter.

Our purpose, however, is not to construct real-world models but to exhibit the frailties of backward induction. Insofar as these frailties carry over to more realistic settings, which we briefly discuss, the uncertainty problem is genuinely paradoxical. Hence, one should be circumspect about making backward induction a basis, let alone the defining characteristic, of rational play in games.

2. The Parity Problem

Assume that the three players in a truel are A, B, and C, and think of them as standing at the vertices of an equilateral triangle. Assume that they fire, one at a time, in a fixed, repeating sequence, such as A, B, C | A, B, C | . . ., where each A-B-C sequence—separated by a vertical line from the preceding one—is a *round*. Each player, hoping to survive itself, has the choice of shooting or not shooting one of its opponents, situated at one of the two other vertices.

More formally, an *outcome* of a truel is a subset of {A, B, C}. In the sequential truel we postulate, there is always at least one survivor (i.e., the subset is nonempty), because only one player fires at a time and the last one to do so will necessarily survive. The players, we assume, order all possible outcomes lexicographically, according to the following three goals:[4]

1. *Primary goal:* each player prefers an outcome at which it survives to one at which it does not survive.

---

[4]That is, a player prefers one outcome to another if, on the highest-ranked goal that distinguishes the two outcomes, the first outcome gives a better result than the second outcome.

2. *Secondary goal:* each player prefers an outcome at which fewer of its opponents survive.
3. *Tertiary goal:* each player, when exactly one of its opponents survives, prefers an outcome at which the surviving opponent is not its antagonist (whether the player survives or not).

Every player, we assume, dislikes one opponent, called its *antagonist*, more than the other. If the antagonist of A is B, we say Ant(A) = B.

As soon as Ant(A), Ant(B), and Ant(C) have been specified, the preference rankings of all three players can be strictly ordered. For example, if Ant(A) = B, then A's preference ranking in descending order is

{A}, {A, C}, (A, B}, {A, B, C}, Ø, {C}, {B}, {B, C},

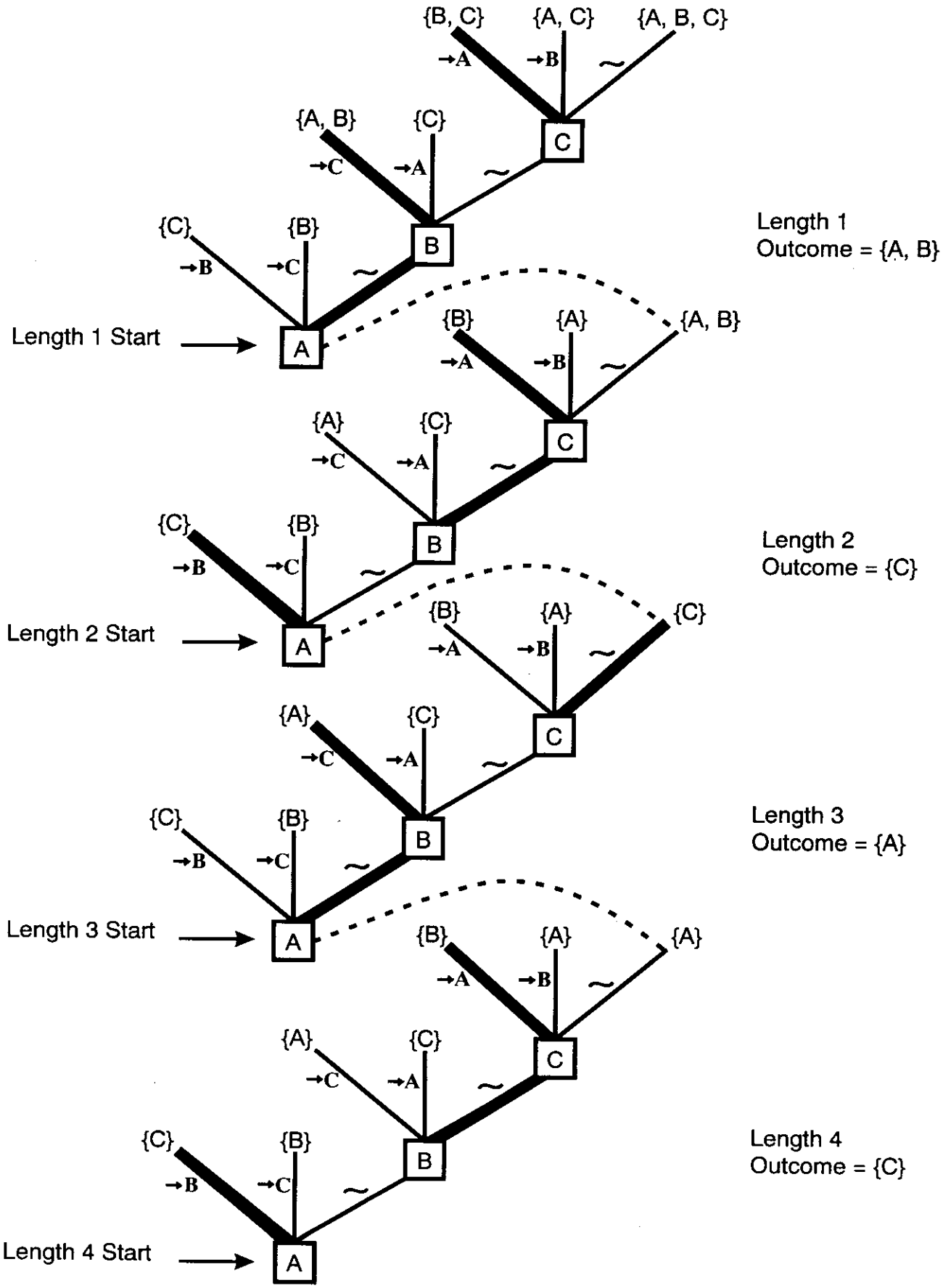
where Ø is the empty set (no survivors). As noted earlier, Ø cannot occur in a sequential truel, but it can occur in a truel in which all players fire simultaneously (Kilgour, 1972; Brams, 1994, pp. 8-10; Brams and Kilgour, 1996).

Suppose that each player is a perfect shot and has one bullet. Assume that the firing order is A, B, C | A, B, C | . . ., and each player can either fire at another player or not fire. Then truels of finite length can be analyzed using the game trees shown in Figure 1.

Figure 1 about here

The game tree at the top of Figure 1 describes the set of choices available to each player when its turn comes up in the first round (i.e., in a game of length 1). Thus, A begins by choosing among three options: shoot B (→B), shoot C (→C), or not fire (~). Whenever one player shoots

Figure 1: A Sequential Truel with Different Numbers of Rounds



another, the outcome of the truel is determined, based on our earlier ranking assumptions, and is shown in Figure 1. For instance, if A shoots B, C will then shoot A, making {C} the outcome, as shown in Figure 1. Likewise, if A shoots C, the outcome will be {B}.

The game goes to a second stage in round 1 if A does not fire, giving B the three options shown. Thus, if B shoots C, then the survivors will be A and B, making the outcome {A, B}, as shown in Figure 1.

Finally, the game goes to a third stage if B does not fire after A does not, leaving C the final choice. C can either shoot one of its two opponents, leaving two survivors, or not fire, which would enable all three players to survive. All the choices of A, B, and C are shown in Figure 1, beginning at "length 1 start" and proceeding upward.

Suppose that the antagonists of the three players are the following: Ant(A) = B, Ant(B) = C, and Ant(C) = A, or, for short, Ant(A, B, C) = (B, C, A). We next apply backward induction to games of length 1 (one round), and then games of greater length (more than one round), as shown in Figure 1.

One Round. To determine rational choices in a length 1 game, we work backward from C's final choices in the third stage. If C should survive until this stage, it would prefer to shoot its antagonist, A, yielding the outcome {B, C}, rather than choose one of its other two options, so we thicken the branch →A to indicate that this branch would be chosen.

Working backward to the second stage, B can anticipate that its choice of branch ~ would result in {B, C}, which it compares with the outcomes of choosing branch →C (i.e., {A, B}) and of choosing branch →A (i.e., {C},

because C would then shoot B). Preferring the outcome {A, B} to either {B, C} or {C}, it would choose →C, given that B survives the first stage.

Working backward to the first stage, A can anticipate that its choice of branch ~ would result in outcome {A, B}, which it compares with the other two outcomes it can effect, {C} and {B}. Preferring {A, B} to these, it would choose branch ~, and B would in turn choose branch →C, making the *backward-induction outcome* of the game {A, B}.

Having worked backward from the top of the game tree (third decision point) to the bottom (first decision point) to determine the players' rational choices (darkened branches) at each stage, we reverse this process to determine what choices the players would actually make. Starting from the bottom of the tree and following the darkened branches upward, we see that play will never reach the third decision point, when it is C's turn to choose, because after A chooses ~, B will choose →C, eliminating C from play and yielding the outcome {A, B} in the one-round game.

We next describe the general solution to the truel for games of any finite length greater than 1. It turns out that there is always only one survivor, not the two (A and B) we just found in a length 1 game.

More than one round. What we have just done for a length 1 game we can do for games of length 2, 3, 4, . . ., which add successive rounds of play to the game of length 1. (A *round* here comprises three decisions—one for each player.) Their analysis simply takes the rational outcome of a length k game and substitutes it as the outcome of the first round of a length $k + 1$ game when nobody shoots anybody in the first round (these substitutions in Figure 1 are indicated by the dashed lines).

For example, we know from the foregoing analysis of a length 1 game that if nobody shoots in the first round of the length 2 game, the rational outcome will be {A, B} in the second round, because the remaining game is a length 1 game. Consequently, in the first round of the length 2 game, the outcome of branch ~ for C in the third stage is {A, B} (the outcome of the length 1 game) rather than {A, B, C} (the outcome, if nobody shoots, of the length 1 game). Backward induction in the length 2 game shows that the rational outcome is {C}, as indicated in Figure 1.

Substituting {C} as the outcome of branch ~ in the third stage of the length 3 game yields {A} as the outcome of this game. Substituting {A} as the outcome of branch ~ in the third stage of the length 4 game yields {C} as the outcome of this game. In summary, games of length 1, 2, 3, and 4 have as outcomes {A, B}, {C}, {A}, {C}, respectively; the {C} - {A} alternation continues to repeat for longer-length games, with {C} as the outcome of all even-length games and {A} as the outcome of all odd-length games longer than one round.

Not only does this truel have three different outcomes ({A, B}, {C}, and {A}), depending on how many rounds are played, but it also does not "settle down," as play continues indefinitely, because of the even-odd alternation.[5] Technically, this truel has no outcome, in the limit, as the number of rounds approaches infinity.

By contrast, for all other possible antagonisms of the three players (each of the three players can have one of two antagonists, so there are $2^3 = 8$ possible antagonisms, as shown in Table 1), the outcome of the truel does

---

[5]The sensitivity of outcomes to parity considerations was first noted, as far as we know, by Kilgour (1984), who showed this sensitivity to occur in two-person games.

Table 1 about here

not depend on its length, once past the first round. Observe that for the six antagonisms in which there is one player who is nobody's antagonist, that person is the invariably the only long-run survivor, underscoring the value of not having enemies.[6]

The sequential truel for Ant(A, B, C) = (B, C, A) has another curious feature besides its even-odd alternation: whatever its length, all shooting occurs in either the first round, or in the first two rounds. If the truel is of length 1, A does not fire and then B shoots C. If the truel is of length 2 (or any other even length), A shoots B and then C shoots A in the first round. If the truel is of length 3 (or any other odd length except length 1), A does not fire and B shoots C in the first round; then A shoots B in the second round.

While shooting never happens in rounds 3, 4, . . ., it is the *anticipation* of these subsequent rounds—in particular, whether the total number of rounds to be played is even or odd—that completely determines whether the outcome is {C} (even) or {A} (odd). Since every finite truel in which there is more than one round of play must end in an even or an odd number of rounds, the rational outcome of *every* multiple-round truel must be either {C} or {A}.

[6]George Bush and Bill Clinton were each other's antagonists in the 1992 U.S. presidential election, putting Ross Perot in the role of the nonantagonist. This is perhaps a partial explanation—another being Perot's massive campaign spending—of why Perot received a larger percentage of the popular vote (19%) than any third-party candidate since Theodore Roosevelt, who received 27% in 1912 (Roosevelt had previously been president).

TABLE 1: Dependence of Outcomes on Length and Antagonisms in Sequential Truels

Ant(A, B, C) =	(B, C, A)	(C, A, B)	(B, A, A)	(C, A, A)	(B, C, B)	(B, A, B)	(C, C, A)	(C, C, B)
Length = 1	{A, B}	{A, B}	{C}	{B}	{A, B}	{A, B}	{A, B}	{A, B}
Length = 2	{C}	{B}	{C}	{B}	{A}	{C}	{B}	{A}
Length = 3	{A}	{B}	{C}	{B}	{A}	{C}	{B}	{A}
Length = 4	{C}	{B}	{C}	{B}	{A}	{C}	{B}	{A}
Length = 5	{A}	{B}	{C}	{B}	{A}	{C}	{B}	{A}

3. The Uncertainty Problem

We begin with an informal argument of why no player would want to shoot another in a sequential truel of uncertain length. At the start of play, and looking ahead to future rounds, each player knows that if it shoots either of its opponents, the other opponent will shoot it at the first opportunity—either in the same round or the next round (assuming there is one)—so shooting in the first round will not satisfy its primary goal. Because this logic carries through to all subsequent rounds, nobody will shoot an opponent. Consequently, the players will all survive the truel.

We now proceed to describe a procedure for solving sequential truels of finite but uncertain length in order to ascertain when the preceding informal argument holds rigorously. In our formal model, we postulate that a truel will continue to the next round with a known probability, called a *continuation probability*.[7] Also, we assume that each player has von Neumann-Morgenstern utilities for the outcomes, consistent with its goals (as given in section 2), and that these utilities are common knowledge. Thus, the players can make expected-utility calculations and choose strategies that maximize their expected utilities.[8]

The game tree shown in Figure 2*a* duplicates the length 1 tree in Figure

---

[7]Such probabilities are used in recent Prisoners' Dilemma repeated-game models (Jones, 1995a, 1995b), but in these models the continuation is to a new stage game rather than a new round of play.

[8]However, if there is genuine uncertainty—in the sense of Knight (Dow and Werlang, 1994)—about which round a game will end on, no expected-utility calculations, based on a risk model, are possible because the probabilities are not known.

Figure 2 about here

1. Taking x_1 to be the outcome of this one-round game, we can substitute $x = x_1$ into the game shown in Figure 2*b* to find the outcome, $x_2 = f(x_1)$, of the two-round game. In general, the outcome of the n-round game is x_n, defined recursively by $x_n = f(x_{n-1})$. The tree in Figure 2*b* presumes that the game will continue at least one more round.

Assume the antagonisms are Ant(A, B, C) = (C, A, B), which yields $x_1 = \{A, B\}$ and $x_2 = x_3 = \ldots = \{B\}$ (see Table 1).[9] Unlike the antagonisms we assumed in section 2, there is no even-odd alternation and hence no parity problem.

To model a sequential truel of indefinite length, suppose that, at the end of round i, a random event occurs that determines whether the truel continues at least one more round (with probability p_i), or whether it ends immediately (with probability $1 - p_i$). Thus, the probability that a truel ends after exactly k rounds is $p_1 p_2 \cdots p_{k-1}(1 - p_k)$. The truel is *bounded* iff $p_i = 0$ for some i.

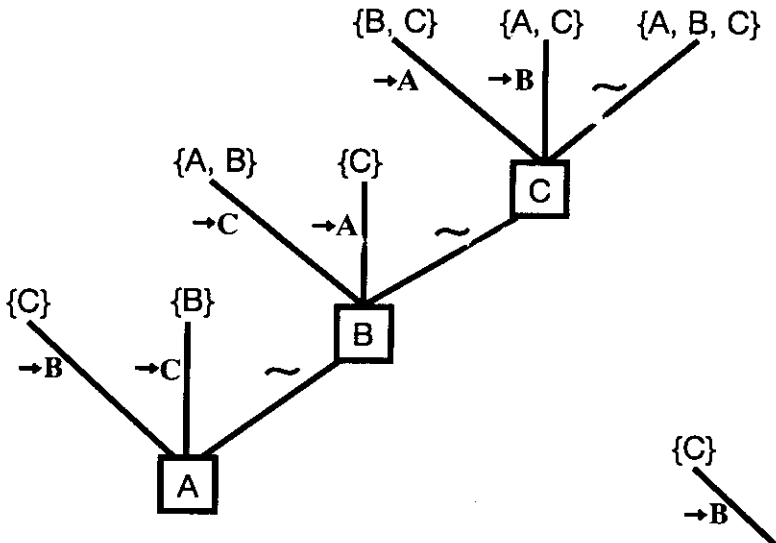
It is possible for a truel to have both a positive probability of terminating and a positive probability of going on forever. For instance, take $p_i = 2^{-(2^{-i})}$. In this case, the probability that the truel does not terminate in any of the first k rounds is

$$2^{-(2^{-1})} \cdots 2^{-(2^{-k})} = 2^{-(2^{-1} + \ldots + 2^{-k})} = 2^{-(1 - 2^{-k})},$$

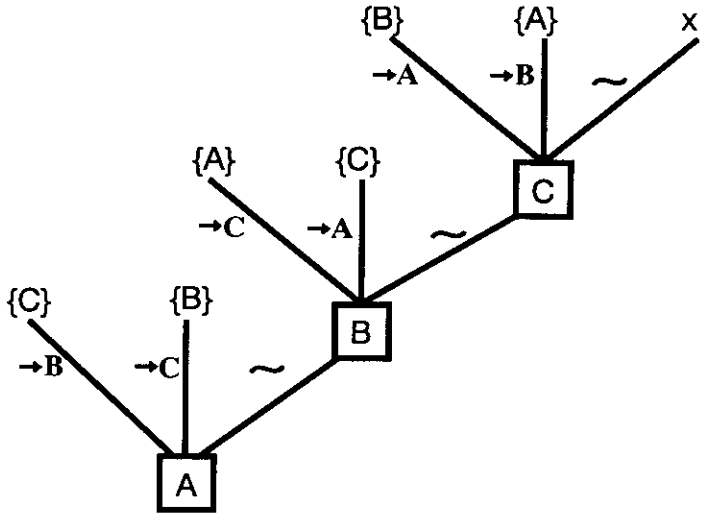
which is a decreasing function of k that approaches $2^{-1} = 1/2$ as $k \to \infty$. Thus, there is a positive probability that this truel ends after each round, but,

[9] In terms of our notation given in the previous paragraph, $f(\{A, B\}) = f(\{B\}) = \{B\}$.

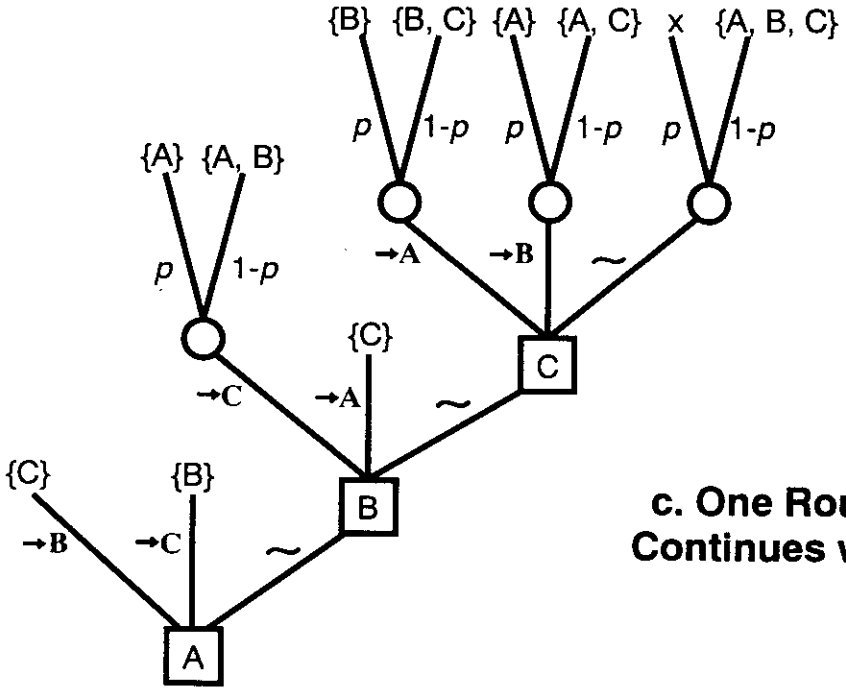
Figure 2: Sequential Truels that Continue



a. Final Round of Truel



b. One Round of Truel that Continues



c. One Round of Truel that Continues with Probability *p*

for any finite k, the probability that it ends at or before the end of round k is strictly less than 1/2.

To analyze a sequential truel of indefinite length, we use the game tree in Figure 2*c*, which provides a snapshot of the game in one round, before which nobody has fired. In this tree, x is the outcome at the upper right that occurs if A, B, and C do not fire and, with continuation probability p, the game continues to the next round.

C's Decision. For our postulated truel in which Ant(A, B, C) = (C, A, B), C's preference order is

$$\{C\}, \{A, C\}, \{B, C\}, \{A, B, C\}, \varnothing, \{A\}, \{B\}, \{A, B\}.$$

Suppose that, in a round with continuation probability p, C fires at A. Then the outcome is a lottery, which we write as

$$p\{B\} + (1 - p)\{B, C\},$$

because C will survive only if the game ends immediately with probability $1 - p$. On the other hand, if C fires at B, then the outcome is

$$p\{A\} + (1 - p)\{A, C\}.$$

Because C prefers {A} to {B} and {A, C} to {B, C}, C prefers the latter lottery.

Consider a truel that cannot exceed two rounds in length (i.e., one in which $p_2 = 0$). Then if round 2 is played, it matches the game in Figure 2*a*, and its outcome is {A, B}. Set x = {A, B} in the game in Figure 2*c*, and note that if C chooses branch ~, the result is a lottery,

$$p\{A, B\} + (1 - p)\{A, B, C\},$$

where $p = p_1$. Because C prefers {A} to {A, B} and {A, C} to {A, B, C}, C will fire at B and receive the lottery

$$p\{A\} + (1 - p)\{A, C\}.$$

With this information, we next consider B's prior decision.

B's Decision. B's preference order is

$$\{B\}, \{B, C\}, \{A, B\}, \{A, B, C\}, \varnothing, \{C\}, \{A\}, \{A, C\}.$$

If B chooses branch ~, the lottery

$$p\{A\} + (1 - p)\{A, C\}$$

results. If B fires at A, the result is {C} for certain. Because B prefers {C} both to {A} and to {A, C}, B will fire at A rather than choose ~.

Finally, if B fires at C, the lottery

$$p\{A\} + (1 - p)\{A, B\}$$

results. Because B prefers this lottery to

$$p\{A\} + (1 - p)\{A, C\},$$

B will fire at C rather than choose ~.

In summary, B's best choice may be either to fire at A, yielding {C}, or to fire at C, yielding the lottery

$$p\{A\} + (1 - p)\{A, B\},$$

because {A, B} is better than {C}, but {C} is better than {A}. Clearly, {C} is preferable when p is large, and the lottery is preferable when p is small.

Denoting player j's utility for outcome E by $u_j(E)$—so that, for example, B's utility for {A, B} is $u_B(A, B)$—it is easy to verify that the threshold probability q at which B would be indifferent between {C}, and the lottery comprising {A} and {A, B}, is

$$q = \frac{u_B(A, B) - u_B(C)}{u_B(A, B) - u_B(A)}.$$

A's Decision. To complete the analysis, notice that A's preference order is

$$\{A\}, \{A, B\}, \{A, C\}, \{A, B, C\}, \varnothing, \{B\}, \{C\}, \{B, C\}.$$

The possible outcomes are {C} if A fires at B and {B} if A fires at C. If A chooses ~, the outcome will be the lottery,

$$p\{A\} + (1 - p)\{A, B\},$$

if p is small and {C} if p is large, based on the preceding analysis of B's decision. It is not difficult to verify that if p exceeds q, A's best choice is to fire at C, yielding {B}, whereas if p falls below q, A will not fire, resulting in the aforementioned lottery.

General Case: Bounded Truel. Now consider a truel that is certain to end in some finite number of rounds, k, because $p_k = 0$. What we have just analyzed is the players' behavior in the preceding round, k - 1, with the possibility of continuation to round k, assuming that no shots were fired prior to round k - 1. To analyze what will happen in round k - 2, consider the two possible results, x = {B} and

$$x = p_{k-1}\{A\} + (1 - p_{k-1})\{A, B\},$$

that can occur if nobody has fired until round k - 1 in Figure 2*c*.

It is now possible to proceed by backward induction. Because C prefers

$$p_{k-2}\{A\} + (1 - p_{k-2})\{A, C\}$$

to either {B} or

$$p_{k-1}\{A\} + (1 - p_{k-1})\{A, B\},$$

it follows that C will always fire at B.[10] Now all the previous analysis (i.e., beginning with C's decision and going back to B's and A's decisions) applies, and the game repeats all the way back to round 1.

In summary, if the truel is bounded, then the outcome is determined by the value of p_1. If p_1 exceeds q, the outcome is {B}; if p_1 falls below q, the outcome is

$$p_1\{A\} + (1 - p_1)\{A, B\}.$$

Thus, if the continuation probability of going to a second round in a bounded truel is "high," A will shoot C, and B in turn will shoot A, in round 1, yielding {B} as the outcome. If the continuation probability is "low," A will not fire and B will shoot C (who is *not* B's antagonist; if B shot its antagonist, A, C in turn would shoot B), giving {A, B} as the outcome. But in the likely event that the game continues to round 2, A would shoot B, making {A} the outcome.

---

[10]For the latter preference to be strict, the assumption $p_{k-1} < 1$ is required. But this is innocuous, because p_{k-1} must fall below the previous threshold for the lottery, $p_{k-1}\{A\} + (1 - p_{k-1})\{A, B\}$, to be available.

General Case: Unbounded Truel. Now consider a truel of unbounded length—that is, for which $p_i > 0$ for all i. Each player's strategy is now an infinite sequence of actions that, because of the assumption that players are perfect shots, may be a function of the history of the game—in particular, of the opponent's previous choices (e.g., a player cannot fire if he or she was previously eliminated).

Suppose that, in every round after round k - 2, the players plan to act just as they would have if the truel were of bounded length. Then if no player fired in round k - 2 (or earlier), the outcome will be either

$$x = \{B\} \text{ or } x = p_{k-1}\{A\} + (1 - p_{k-1})\{A, B\},$$

as we showed earlier in the case of the bounded truel. In fact, all choices in round k - 2 follow from backward induction in Figure 2c, just as in the bounded case. To wit,

- A fires at C if $p_{k-2} > q$ and does not fire if $p_{k-2} < q$;
- if A does not fire, B fires at C;
- if both A and B do not fire, C fires at B.

If no player fired before round k - 2, the outcome, therefore, will be either {B} (if A fires at C) or $p_{k-2}\{A\} + (1 - p_{k-2})\{A, B\}$ (if A does not fire), depending on whether p_{k-2} does or does not exceed q.

This argument applies to any round; as in the bounded truel, we can carry backward induction to round 1. It follows that, in the unbounded truel, there always exists what we call a *bounded equilibrium*, in which the players act exactly as they would in the bounded truel. At this equilibrium, the outcome is either {B} or $p_1\{A\} + (1 - p_1)\{A, B\}$, according to whether the round 1 continuation probability, p_1, does or does not exceed q.

However, an entirely different equilibrium can emerge in the unbounded truel, coexisting with the bounded equilibrium. A *cooperative equilibrium* occurs when no player ever fires in any round i, yielding {A, B, C} as the outcome. C prefers this outcome to firing at B, and receiving the lottery

$$p_i\{A\} + (1 - p_i)\{A,C\}$$

if, in every round i,

$$u_C(A, B, C) \geq p_i u_C(A) + (1 - p_i)u_C(A, C),$$

which is true if

$$p_i \geq \frac{u_B(A, B) - u_B(A, B, C)}{u_B(A, C) - u_B(A)} \equiv r_C.$$

By a similar analysis, if B anticipates that not firing will result in {A, B, C}, then B will not fire in round i provided that

$$u_B(A, B, C) \geq p_i u_B(A) + (1 - p_i)u_B(A, B),$$

which is true if

$$p_i \geq \frac{u_B(A, B) - u_B(A, B, C)}{u_B(A, B) - u_B(A)} \equiv r_B.$$

For A, there is no similar condition: A can achieve {C} by firing at B and can achieve {B} by firing at C. Because A prefers {A, B, C} to either of these outcomes, A will unconditionally accept the cooperative equilibrium and not fire.

In summary, the cooperative equilibrium can occur whenever every continuation probability p_i exceeds both thresholds, r_B and r_C, making it rational for the players always to continue to the next round.[11] In symbols,

$$p_i \geq \max\{r_B, r_C\} \equiv r.$$

Numerical Example. To get some feel for the threshold q in the bounded equilibrium, and the threshold r in the cooperative equilibrium when player is unbounded, suppose that B and C both attach utilities of 7, 6, . . ., 1, 0 to the eight possible outcomes, giving their most preferred outcomes utilities of 7 and their least preferred outcomes utilities of 0. At the bounded equilibrium, $q = 3/4$, so the probability that play goes into the second round must be relatively high ($p_1 > 3/4$) to induce

- A to shoot C, and B in turn to shoot A, making {B} the outcome.

By comparison, if $p_1 < 3/4$,

- A will refrain from firing, B will shoot C, and there will be a positive probability (p_1) that {A, B} will be the outcome (if there is no second round) and a complementary positive probability ($1 - p_1$) that {A} will be the outcome (if there is a second round, it will begin with A's shooting B).

The cooperative equilibrium threshold is $r = 1/2$. Hence, if the continuation probability, p_i, on every round is at least 1/2, the cooperative equilibrium, in which nobody ever shoots, can be sustained. Indeed, A will

---

[11]Note that because we do not know the last round when the truel is unbounded, this calculation of the expected utility of continuing replaces that of using backward induction when the last round is known.

prefer it if $p_1 > 3/4$, because otherwise A will shoot C and B will shoot A in the first round, producing {B}, which is less preferred by A than {A, B, C}.

But the reasonableness of {A, B, C} as the outcome is very sensitive to one's thinking about the continuation of the game. For example, if $p_i = .51$ for every i, the truel will almost certainly (i.e., with probability .9999986) not continue past the twentieth round. Now if this near certainty were seen to be a certainty, we would have only the bounded equilibrium, yielding {A} as the outcome with probability .51 and {A, B} as the outcome with probability .49. This equilibrium is a drastically different from the cooperative equilibrium {A, B, C}; it is caused by only a .0000014 change (slightly more than one in a million) in the probability that the game ends by round 20.[12]

4. Is There a Paradox?

We have shown that backward induction may yield radically different outcomes when the number of rounds that a sequential truel is played switches from being even to being odd, or when the endpoint changes from being certain to being uncertain. But while the parity and the uncertainty

---

[12]Recall that a truel is bounded iff $p_i = 0$ for some round i; because $p_i = .51$ for all i, this truel is, technically, unbounded. But, realistically speaking, is it? In the unlikely event that it continues beyond the twentieth round, the prospective situation is exactly at it was before: the expected length of the truel at the beginning of any round is 2.04, and the probability that it continues 20 more rounds is again 0.0000014. For all practical purposes, the truel will have ended by round 20 (or round 40, if it should reach round 20, etc.), making it *appear* bounded, which would seem to reinforce the bounded equilibrium. On the other hand, because in our example it is not rational for A to fire initially in either the bounded or the cooperative equilibrium, it is B who "selects" the equilibrium by firing or not firing. Since $p_i = .51 > r = .50$, the condition for the cooperative equilibrium is met, so neither B nor C should fire if the truel is viewed as unbounded. Hence, we have something of a paradox in this numerical example, depending on whether the players think of the truel as bounded or not. Beyond this example, we shall later explore two conflicting views of the future that give rise to the different equilibria.

problems afflict the robustness of backward induction, is there a more fundamental problem—something paradoxical—with backward induction?

We have mixed views on this question. With respect to the parity problem, we consider such fluctuating behavior not genuinely paradoxical in the sense of there being a clash of different logical principles. Neither is it paradoxical that the choices of the players reverberate to the beginning of the game, when all shots are fired. On the other hand, we would be hard pressed to say that there is a significant difference, in any conceivable game that we think real players might play (a truel or anything else), between its running 63 or 64 rounds. Thus, this sensitivity of backward induction to the number of rounds is highly unlikely to model what real people would think and do.

The uncertainty problem seems to us of a different nature. First, recall that the truel we analyzed in section 3, when the number of rounds was certain, gave {A, B} as an outcome in one round—the same as in our first truel analyzed in section 2—but a steady stream of {B}'s in all finite games that run more than one round. Not only is there is no parity problem, but there seems to us nothing inconsistent with what happens when the endpoint is made uncertain, as long as the number of rounds has an upper bound: {A, B} is the outcome when the truel is "likely" to terminate after one round (i.e., when the continuation probability in the first round is less than some threshold probability, q)—except in the "unlikely" case that it does not terminate (in which case {A} becomes the outcome)—whereas {B} is the outcome when the truel is "likely" to continue.

The former mixed result is little more than a probabilistic refinement of the certain case that always yielded {A, B} in a one-round truel. Now, however, it is qualified by the uncertainty surrounding whether the truel continues past the first round (in which case A shoots B, making {A} the

outcome). By contrast, the constant {B} result in the uncertain case simply duplicates the {B} result in the certain case.

The story is strikingly different when the number of rounds is finite but unbounded. If the continuation probability is sufficiently high on each round *i* (i.e., greater than some threshold probability, *r*, which in our numerical example equals .50), then nobody will ever fire and the outcome will be {A, B, C}. This is true despite the fact that the truel *must* end in a finite number of rounds, which in the certain case means that the outcome is either {A, B} or {B}.

In the boundedly uncertain case, we pick up {A} as an additional possible outcome. But none of these three possible outcomes ({A}, {B}, or {A, B}) occurs in the unboundedly uncertain case when the continuation probability is sufficiently high for the cooperative equilibrium and its associated outcome, {A, B, C}. Thus, a qualitatively different result can occur when the truel might continue indefinitely (but is still finite).[13]

In the finite case, of course, the truel *must* end; moreover, it will do so with a higher and higher probability as the number of rounds increases. But because this looming endpoint, and its associated boundedness equilibrium, are inconsistent with unbounded play and the cooperative equilibrium, we have an apparent paradox.

5. A Conflict of Two Futures

In section 3 we highlighted this difference with an example in which, if the probability of termination increased minutely from .9999986 to 1 after

---

[13]We do not consider here the possibility that truels might be infinite, or sometimes finite and sometimes infinite, as illustrated in section 3. Our main purpose is to draw comparisons among finite truels in three cases—one in which the number of rounds is certain, and two in which this number is uncertain (with and without an upper bound).

twenty rounds of play, {A, B, C} would be undermined, because now uncertain play would be bounded. Clearly, there is an obvious lack of robustness between the boundedly and the unboundedly uncertain cases.

Beyond this robustness problem, there may be a conflict between two possible futures that are, seemingly, logically inconsistent:

1. Every process must end by some definite point (e.g., every person's life now seems to have an upper bound of about 120 years);
2. The precise future is unpredictable, so the exact endpoint of a process cannot be predicted (it may be highly unlikely that a person will ever live to be 120, but it is not impossible).

Future 1 suggests that it is proper to assume that all games are bounded, whereas future 2 suggests that unboundedness is a more appropriate assumption.

In fact, future 2 has been argued to be essential in sustaining cooperation in games like finitely repeated Prisoners' Dilemma. If the endpoint is known for certain, then backward induction can be applied, resulting in noncooperative behavior. But both intuition and experimental results in repeated Prisoners' Dilemma—as well as games more akin to the truel (e.g., the Centipede Game, which can go several rounds)—demonstrate that even knowing their maximum length, cooperation occurs frequently in these games (Binmore, 1990, discusses reasons for this). Furthermore, cooperation may be rational even in one-shot Prisoners' Dilemma and other games, such as Chicken, if the rules of play allow for "farsighted thinking."[14]

---

[14]See Brams (1994) and references cited therein. More recently, Ecchia and Mariotti (1995) and Willson (1996) also allow for farsighted thinking under different rules.

In our truel, the optimistic {A, B, C} outcome, in which nobody fires, is consistent with future 2, whereas the pessimistic {A}, {B}, or {A, B} outcomes are consistent with either future 1 or future 2. It seems that some real-world players have adhered more to the thinking reflected by the cooperative equilibrium of future 2, such as the United States, Russia, and China: although each possesses nuclear weapons, all have refrained from using them against each other in anything resembling a truel.

The same self-restraint manifested itself with the nonuse of poison gas in World War II, partly in response to revulsion with its use in World War I and partly in fear of reprisal. By contrast, Bosnian Serbs, Bosnian Muslims, and Croats engaged in a very destructive truel in the former Yugoslavia in the early and mid-1990s, mirroring the boundedness thinking of future 1 and the bounded equilibrium of future 2.

Truels in recent films give diametrically opposed results. The climactic scenes in Quentin Tarantino's two films, *Reservoir Dogs* (1992) and *Pulp Fiction* (1995), are truels, but the outcomes are very different in each. Arguably, the truelists in *Reservoir Dogs*, in which several people die, were more bounded in their thinking than those in *Pulp Fiction*, in which nobody is killed in the truel.

Everybody would be better off, we believe, if players did *not* think they were so clever as to be able to reason backward, from some endpoint, in plotting each other's destruction. Indeed, our results suggest that players would be less aggressive if the future were seen as somewhat murky, which would render predictions about how many rounds a game will go, or even an upper bound on this number, hazardous.

The absence of a fixed order of play in most real-world three-person conflicts—as opposed to the sequentiality we postulated in our examples—

probably tends to discourage shooting. After all, if any of A, B, or C contemplates shooting first, even in a one-round nonsequential truel, then it would ensure its own death when the remaining survivor takes aim.

6. Conclusions

The main argument of this paper is that backward induction may be extremely sensitive to seemingly innocuous changes in the rules, such as the number of rounds a game is played or the nature of the uncertainty about the endpoint of a game. In the case of the latter, two possible views of the future seem to underlie bounded and unbounded play.

The unbounded view is probably more hopeful—if not always more realistic—in truel-like games. It is important to recognize, however, that the bounded view is certainly justified in certain situations, like elections, in which campaigns end on election day. Of late, elections have suffered from a good deal of negative campaigning, perhaps because, like truels, there is a cascade effect: one player's "shot" sets off others.

Yet many of the most important decisions we make in life, especially of an existential nature, are not substantially constrained by law, custom, or time.[15] To the degree that the future seems to stretch out indefinitely, people probably act more responsibly toward each other, knowing that tomorrow they may pay the price for their untoward behavior today.[16] Not only do individuals try to develop reputations that will sustain them in the long run,

---

[15]In the absence of law, in particular, people often are able to work out their differences, suggesting that amicable settlements may be facilitated when deadlines are not fixed and procedures are somewhat inchoate (Ellickson, 1991).

[16]Of course, if one believes that one's ultimate reward comes in some afterlife, then a violent act like a suicide bombing, which propels one immediately into that afterlife, can be justified. Fortunately, most people do not prize martyrdom of this kind.

but some, by acting morally, seek an inner peace, which Frank (1988) persuasively argues can be eminently rational.

At a theoretical level, characterizing multiple-round games in which unbounded play can lead to a cooperative equilibrium but bounded play does not will probably not be easy. Insofar as unbounded play is *effectively* bounded (as in our numerical example), this equilibrium may not be unique or particularly sturdy. This lack of robustness is especially likely in games in which the probability of continuation past a few rounds becomes vanishingly small, rendering backward-induction calculations more sensible.

Unfortunately, in such games players may well be able to rationalize shooting from the start. An important intellectual task, we believe, is to try to devise institutions that render such behavior unprofitable. But how one makes the future seem to run on smoothly, and instill confidence that the social fabric will not suddenly unravel, is not so clear.

Perhaps the best antidote to people's fears of the future is a past record of institutions' responding well to potentially disruptive events. Democratic institutions usually get high marks in this regard, primarily because they provide an escape valve that tends to prevent explosions. The mechanism by which they do this, however, we leave for future work.

References

Aumann, Robert J. (1992). "Irrationality in Game Theory." In Partha Dasgupta *et al.*, *Economic Analysis of Markets and Games* (*Essays in Honor of Frank Hahn*). Cambridge, MA: MIT Press, pp. 214-227.

Aumann, Robert J. (1995). "Backward Induction and Common Knowledge of Rationality." *Games and Economic Behavior* 8, no. 1 (January): 6-19.

Bicchieri, Cristina (1993). *Rationality and Coordination.* Cambridge, UK: Cambridge University Press.

Binmore, Ken (1990). *Essays on the Foundations of Game Theory.* Cambridge, MA: Basil Blackwell.

Brams, Steven J. (1994). *Theory of Moves.* Cambridge, UK: Cambridge University Press.

Brams, Steven J., and D. Marc Kilgour (1996). "The Truel." Preprint, Department of Politics, New York University.

Dow, James, and Sérgio Ribeiro de Costa Werlang (1994). "Nash Equilibrium under Knightian Uncertainty: Breaking Down Backward Induction." *Journal of Economic Theory* 64, no. 2 (December): 305-324.

Ellickson, Robert C. (1991). *Order without Law: How Neighbors Settle Disputes.* Cambridge, MA: Harvard University Press.

Ecchia, Giulio, and Marco Mariotti (1995). "International Environmental Coalitions: Some Theoretical Observations." Preprint, Economics Department, University of Manchester, UK.

Frank, Robert H. (1988). *Passions within Reason: The Strategic Role of the Emotions.* New York: W. H. Norton.

Jones, Michael A. (1995a). "The Classification of Continuation Probabilities." Preprint, Department of Mathematics, U.S. Military Academy (West Point, NY).

Jones, Michael A. (1995b). "Cones of Cooperation for Indefinitely Repeated, Generalized Prisoner's Dilemma Games." Preprint, Department of Mathematics, U.S. Military Academy (West Point, NY).

Kilgour, D. Marc (1972). "The Simultaneous Truel." *International Journal of Game Theory* 1, no. 4: 229-242.

Kilgour, D. Marc (1984). "Equilibria for Far-sighted Players." *Theory and Decision* 16, no. 2 (March): 135-157.

Kreps, David M., and Robert Wilson (1982). "Reputation and Imperfect Information." *Journal of Economic Theory* 27, no. 2 (August): 253-279.

Milgrom, Paul, and John Roberts (1982). "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27, no. 2 (April): 280-312.

Pettit, Philip, and Robert Sugden (1989). "The Backward Induction Paradox." *Journal of Philosophy* 86: 169-82.

Radner, Roy (1986). "Can Bounded Rationality Resolve the Prisoner's Dilemma?" In Andreu Mas-Colell and Werner Hildenbrand (eds.), *Essays in Honor of Gerard Debreu.* Amsterdam: North-Holland.

Selten, Reinhard (1978). "The Chain-Store Paradox." *Theory and Decision* 9, no. 1 (April): 127-159.

Sobel, Jordan Howard (1994). *Taking Chances: Essays on Rational Choice.* Cambridge, UK: Cambridge University Press.

Stuart, Harborne W., Jr. (1993). "The Finitely Repeated Prisoner's Dilemma: An Interactive, Decision-Theoretic Approach." Preprint, Harvard Business School.

Willson, Stephen J. (1996). "Long-Term Behavior in the Theory of Moves." Preprint, Department of Mathematics, Iowa State University.